



Open Library of Humanities

Perception of ATR contrasts by Akan speakers: a case of perceptual near-merger

Sharon Rose*, Department of Linguistics, University of California San Diego, USA, sxrose@ucsd.edu

Michael Obiri-Yeboah, Department of Linguistics, Georgetown University, USA, michael.obiriyeboah@georgetown.edu

Sarah C. Creel, Department of Cognitive Science, University of California San Diego, USA, screel@ucsd.edu

*Corresponding author.

Despite many acoustic, articulatory, and phonological studies of Advanced Tongue Root (ATR) vowel contrasts and vowel harmony, studies of the perception of ATR contrasts by speakers of languages with ATR vowel distinctions are lacking. This paper explores how vowels that differ for ATR or height, or both, are distinguished by speakers of Akan, a Kwa language of Ghana. We examine whether the phonologically contrastive status of the vowels impacts perception or whether perception is driven by acoustic similarity. Results from two experiments reveal that vowels that differ only for ATR are well distinguished, even those that are in an allophonic relationship. Yet, vowels that are contrastive and differ by both ATR and height features, but are acoustically similar, are poorly perceived. We suggest that these vowel contrasts constitute a case of perceptual near-merger.



1 Introduction

Many languages in Africa are reported to have vowels that contrast for the feature [Advanced Tongue Root] (ATR). This contrast is realized articulatorily as a distinction between an advanced position of the tongue root, along with pharyngeal expansion, in opposition to a neutral (Hudu, 2014) or retracted position (Lindau, 1978).¹ It is also commonly accompanied or identified by the presence of vowel harmony for this feature. Indeed, ATR vowel harmony is especially common within the Nilo-Saharan and Niger Congo phyla (Casali, 2008; Clements, 2007). Rolle, Lionnet, and Faytak (2020) report that out of 681 languages in their Areal Linguistic Features of Africa database, 358, or 53%, exhibit ATR vowel harmony. ATR vowel harmony has long contributed to the development of phonological theoretical accounts of vowel harmony (Stewart, 1967, 1971, 1983; Schacter, 1969; Clements, 1981, 1984, 1985; Bakovic, 2000; Casali, 2008). Phoneticians have also studied the articulation of ATR contrasts in African languages using a variety of methods, including X-rays and ultrasound. Acoustic properties of the vowel contrasts have also been extensively studied. Yet, despite such a large focus on ATR systems, and anecdotal reports of difficulty perceiving distinctions between certain vowels, surprisingly little attention has been paid to the perception of ATR contrasts, especially by speakers of languages with ATR harmony.

1.1 Articulation

Ladefoged (1964), investigating Igbo (Benue-Congo, Nigeria), was one of the first studies of the articulation of ATR vowel contrasts using X-rays. Although he describes the distinction in the cineradiographic tracings as involving the tongue body, Stewart (1967) interpreted these tracings, combined with his own observations of chin lowering in Akan (Kwa, Ghana), as involving the tongue root, with tongue body height changes as a natural consequence of advancing the tongue root. He proposed the feature [Advanced Tongue Root]. Indeed, one can clearly see the retracted and advanced positions of the tongue root in Ladefoged's images. This corroborates Pike (1947), who speculated that pharyngeal modification leading to a deeper resonance in vowels could be accomplished through forward position of the tongue body and root, larynx lowering, or movement of the faucal pillars. Lindau (1978, 1979) used cineradiographic tracings for four speakers of Akan (3 Akyem, 1 Asante). Lindau's results clearly indicate a more advanced tongue root and a lower larynx for +ATR vowels than -ATR vowels, with concomitant expansion of the pharynx. In contrast, -ATR vowels retract the tongue root and raise the larynx, leading to a small pharyngeal space. Due to the pharyngeal expansion, she proposes the phonetic feature [expanded] instead of [advanced tongue root] as it captures more of the articulatory properties. The following diagram (**Figure 1**, reproduced from Lindau, 1978) displays X-ray tracings of

¹ The phonological feature Retracted Tongue Root [RTR] is used in opposition to Advanced Tongue Root [ATR] by some researchers instead of binary values of a single [ATR] feature (e.g., Pulleyblank, 2002). In addition, [RTR] has been used for vowel distinctions in languages outside Africa to refer to tongue root distinctions, such as in Altaic languages (Li, 1996; Kang & Ko, 2012; Kang, 2018).

eight vowels of Akan (Akyem dialect), showing the position of the jaw and the tongue root. It can be seen that the +ATR vowels [i e u o] (IPA: ɨ ɛ ɯ ɔ) show a more advanced tongue root than their –ATR counterparts [ɪ ɛ ʊ ɔ] (IPA: ɨ ɛ ɯ ɔ), but that the raised tongue body is similar (note that ɔ and ɨ are alternate transcriptions for ʊ and ɪ).² In an earlier study of one speaker of the Akyem dialect, Lindau (1975) also examined low vowels, and found the same distinction between [a] and [ɜ].³ Jacobsen (1978) recorded eight speakers of DhoLuo (Nilotic, Kenya) and found that +ATR vowels have a higher tongue and wider pharyngeal cavity than –ATR vowels. However, he found that speakers vary in terms of which articulatory parameter they employ to achieve this, either tongue height or the tongue root position. There is no larynx displacement distinction. This fits with similar results for Ateso (Lindau 1975), and Jacobsen (1978) suggests that East African ATR systems may be different from West African ATR systems in this regard.

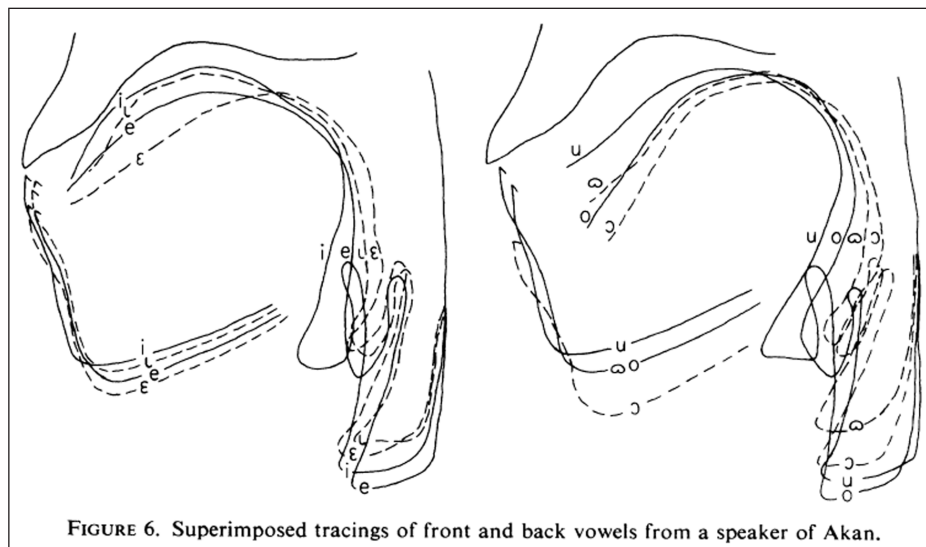


FIGURE 6. Superimposed tracings of front and back vowels from a speaker of Akan.

Figure 1: Lindau (1978, p. 551) cineradiographic tracings for one Akyem dialect speaker. Solid lines indicate +ATR vowels; dashed lines indicate –ATR vowels.

² We employ the vowel transcriptions [ɨ ɛ ɯ ɔ] instead of the IPA symbols [ɨ ɛ ɯ ɔ] as the diacritics to indicate ATR distinctions are difficult to discern.

³ Hess (1992) reports that Lindau (1975) examined the low vowel /a/ and its harmonic counterpart [ɜ], and determined that in the Akyem and Asante Twi dialects, [ɜ] does not have an advanced tongue root position. But this leaves out the fact that the Akyem speaker also had [ɜ] in words preceding +ATR vowels. Lindau describes [ɜ] as +ATR and [a] as –ATR. She states (1975, p. 50) that ‘the [ɜ] is articulated with a more advanced tongue root and lower larynx than [a]. It is also interesting to note that the low [ɜ] has the highest larynx position of all eleven vowels in Akyem.’ We examined the word list Lindau used (1978, p. 138) and found that while the [ɜ] preceded high vowels, the [a] was in the word sá [sɜ] ‘to dance,’ which purportedly contrasts with sá [sa] ‘to cure.’ However, these actually have the same vowel but different tone: ‘to cure’ is [sà]. This may explain the difference in larynx height. Furthermore, the word ‘to dance’ is shown with an optional glottal stop. Although glottal stops can occur pre-pausally in Akan, perhaps the tokens this speaker produced did contain glottal stops, which could have affected the vowel quality. It seems the [a] for this one speaker are –ATR fronted allophones of /a/.

Tiede (1996) used MRI to map Akan vowels for one male speaker of the Asante Twi dialect. He confirmed the tongue root advancement and lower larynx height for +ATR vowels of earlier work, and also showed that +ATR vowels have a larger pharyngeal expansion. The –ATR vowels showed pharyngeal constriction, not just a narrowing of the pharyngeal space due to tongue root and larynx position. Ultrasound has also proven to be an effective imaging technique for ATR contrasts. Studies on Kinande (Bantu, DR Congo) (Gick, Pulleyblank, Campbell, & Mutaka, 2006), Dagbani (Gur, Ghana) (Hudu, 2010, 2014), Lopit (Nilotic, South Sudan) (Billington, 2014), Yoruba (Benue-Congo, Nigeria) (Allen, Pulleyblank, & Ajíbóyè, 2013), and Akan (Kirkham & Nance, 2017) confirm distinct tongue root positions for +ATR and –ATR vowels. Finally, Edmondson, Padayodi, Hassan, and Esling (2007) used laryngoscopy to examine the production of Kabiyè (Gur, Togo) and Akan vowels. They determined that there is an “open epilaryngeal space, a less retracted tongue and neutral larynx height” for [+constricted] (or +ATR vowels), whereas [–constricted] (–ATR vowels) exhibit “a flatter forward-bending laryngeal sphincter angle, a more retracted tongue, and raised laryngeal structures.” Edmondson and Esling (2006) also report compression of the arytenoids and aryepiglottic folds forwards and upwards for –ATR vowels. Most articulatory studies involve a small number of speakers; the largest number is eight in Jacobsen’s (1978) study of DhoLuo.

1.2 Acoustics

Numerous phonetic studies have attempted to determine the acoustic properties of ATR vowel distinctions. Halle and Stevens (1969, p. 211) report that “The clearest and most consistent acoustic consequence of widening the vocal tract in the vicinity of the tongue root is a lowering of the first-formant.” This has been confirmed in numerous studies since. Lindau’s (1978) examination of Akan vowels reported that F1 is the main distinguishing acoustic parameter. [+ATR] vowels have lower F1 than their [–ATR] counterparts. The F1/F2 vowel plot is shown below. The chart in **Figure 2** shows that the +ATR vowels /i u e o/ are well-distinguished from the –ATR counterparts /ɪ ʊ ɛ ɔ a/, but other vowels such as [e] and [ɪ] appear to be acoustically similar, though not identical.

Hess (1992) is a detailed study of the acoustic properties of Akan vowels, based on data from a speaker of the Kwawu dialect (similar to Asante Twi). She examined four acoustic parameters: Formant frequency, formant bandwidth, vowel duration, and relative amplitude of spectral components. The two measures most strongly correlated with [ATR] were F1 and F1 bandwidth. But since F1 is also a correlate of vowel height distinctions, she deemed F1 bandwidth as the more reliable measurement for distinguishing ATR, at least for vowels that have similar formant frequencies. [+ATR] vowels had narrower first formant bandwidths than [–ATR] vowels, correlated with the third harmonic. [+ATR] vowels had a weaker third harmonic than [–ATR] vowels. The F1 bandwidth distinction was useful for distinguishing [+ATR] [e, o] compared to [–ATR] /ɪ, ʊ/, as their F1 values are similar, as seen on the following vowel plot in **Figure 3**. This plot also includes the [æ], which is similar to [ɛ], but with higher F1.

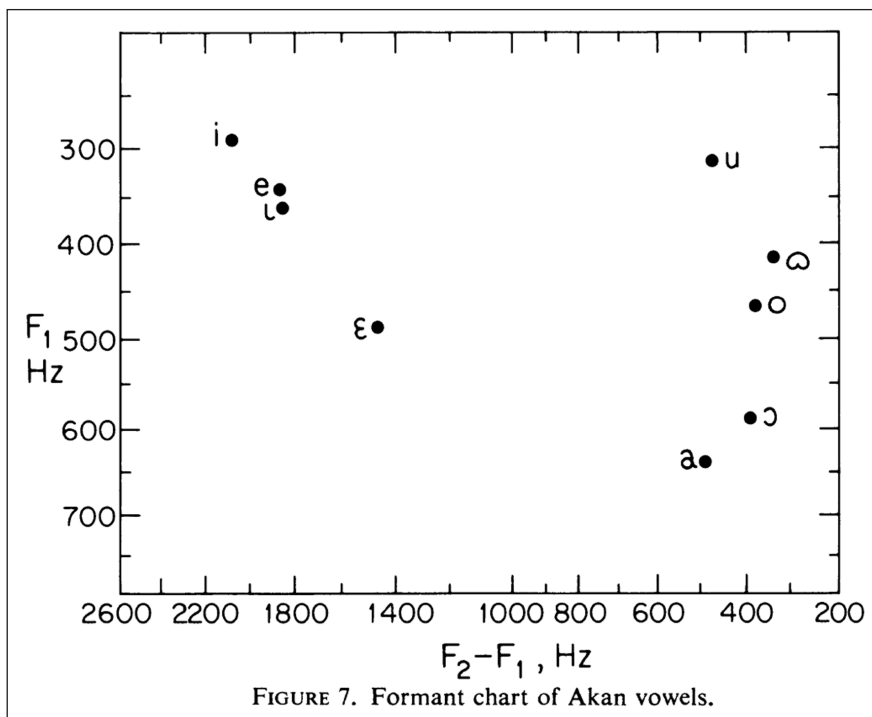


Figure 2: Lindau (1978, p. 552) formant chart for four speakers (three Akyem, one Ashanti) dialect.

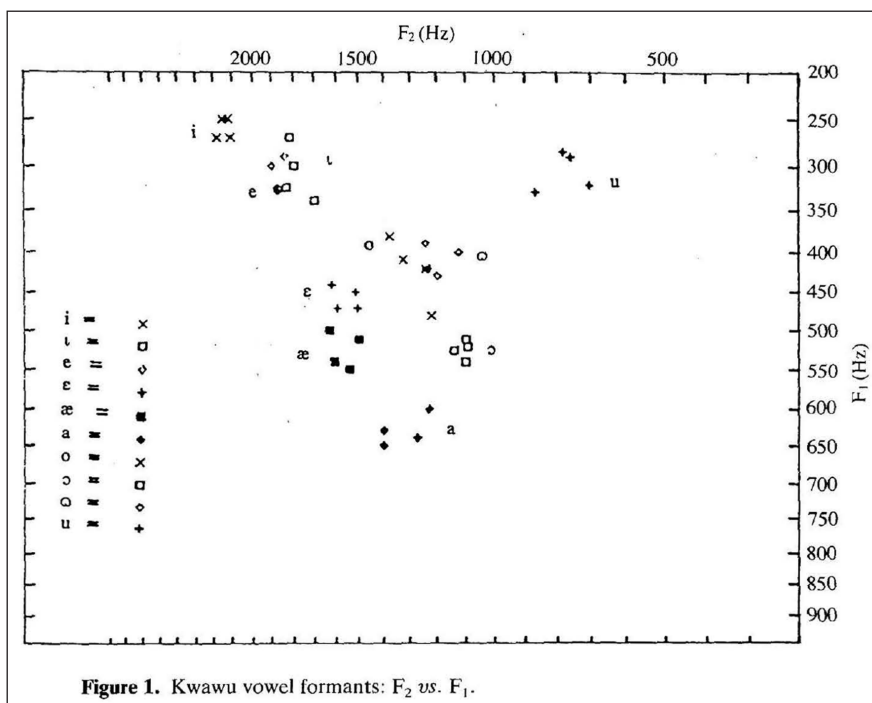


Figure 3: Hess (1992, p. 480) formant chart for one speaker of Kwawu dialect.

Appiah-Padi (1994) analyzed data from nine male speakers of the Asante Twi dialect. She found that besides [+ATR] vowels having lower F1 than [-ATR] vowels, [+ATR] vowels were also longer, except for the [ɛ]/[e] pair. Unlike Hess (1992), F1 bandwidth was not found to distinguish ATR pairs of vowels. Furthermore, a discriminant analysis showed that all ATR pairs were well distinguished using formants, bandwidth, duration, and amplitude, but, in contrast, [ɪ]/[e] and [ʊ]/[o] showed much poorer discrimination ratios.

Kirkham and Nance (2017) provided the F1/F2 vowel space for six speakers of the Akuapem dialect of Akan. They reported similar patterns to previous studies in terms of the similarity of [ɪ] and [e], as well as [ʊ] and [o]. They also showed that speakers differ in the position of the vowel [ɜ]. For three male speakers, it is positioned close to [ɛ] or [ɪ], whereas for two female speakers, it is close to [e]. For one female speaker, it is close to [a].

Other notable acoustic studies of ATR distinctions include Jacobson (1978) on Dholuo (Nilotic, Kenya) and Guion, Post, and Payne (2004) on Maa (Nilotic, Kenya). Guion et al. found that [+ATR] vowels have lower first formant values and relatively less energy in the higher frequency regions than [-ATR] counterparts. Fulop, Kari, and Ladefoged (1998) analyzed data from six speakers of Degema (Edoid, Nigeria). They found that all [+ATR]/[-ATR] vowel pairs were distinguished by F1, except the low vowel pair. Starwalt (2008) investigated 11 African languages with ATR contrasts from the Kwa and Bantu families, with five speakers of each language. She found considerable variation among languages and speakers, but the most consistently reliable acoustic measurement distinguishing the contrast was F1. [+ATR] vowels tended to have lower F1 than their [-ATR] pairs. F1 bandwidth and normalized A1–A2 measurements were only marginally different in some languages. [+ATR] vowels also tended to have lower center of gravity measurements.

There are reports in the literature of ‘breathy’ or ‘hollow’ phonation associated with +ATR vowels (Berry, 1957; Casali, 2008). Meanwhile –ATR vowels are described as ‘brighter’ (Fulop et al., 1998), ‘creaky,’ or ‘tense’ (see Denning, 1989 for an overview), but several acoustic studies cast doubt on voice quality as a distinguishing feature. Hess (1992) finds no evidence for breathy phonation in Akan as measured by the difference in amplitude between H2 and F0. Guion et al. (2004) used electroglottographic data from one Maa speaker and found a slightly less constricted glottis for [+ATR] than [-ATR] vowels, which may be indicative of a phonatory difference, but they could find no discernible voice quality distinctions in recordings. Local and Lodge (2004) found that for a speaker of the Tugen dialect of Kalenjin, breathiness was associated with –ATR vowels, auditorily perceptible as well as established through measurements of contact quotient of the glottal cycle made from electrolaryngographic recordings. They otherwise found that F1 was the primary acoustic measure distinguishing the vowels.

Finally, it is worth noting the acoustic similarity reported between certain vowels that are distinguished by both ATR and height: [ɪ]/[e] and [ʊ]/[o]. Lindau (1987, p. 51) states that “...

there may be a case for concluding that for practical purposes, these two pairs of vowels have merged phonetically in Akan.” However, the tongue root positions are different, and the height of the larynx is distinct between the +ATR and –ATR pairs, so articulatorily, they are produced differently, despite being acoustically similar. Stewart (1971, p. 200) reports that in some Akan dialects these vowels have merged, but it is not clear if that means an acoustic/perceptual merger or a merger in articulation. Hess (1992) also noted an acoustic merging of front –ATR /ɪ/ and front +ATR /e/, but states that “They remain phonetically distinct in that they are produced in characteristic ways,” so articulatorily they are different even if their acoustics makes them similar or merged. Hess further reports they are distinguished by F1 bandwidth measurements. Nevertheless, the acoustic similarity may have repercussions for their perception. Anecdotal evidence points to perceptual difficulty. Casali (2017) reports that the high –ATR [ɪ] and mid +ATR [e], as well as the [ʊ] and [o], are frequently confused by fieldworkers and this has led to mischaracterizations of inventories. The term “fieldworkers” implies people who do not speak languages with ATR contrasts. Therefore, they may be guided by their own first language perceptual system, and this does not mean that speakers of languages with ATR contrasts cannot distinguish them (Casali, 2017, p. 84). As for the [a]/[ɜ] pair, Casali (2012) reports that “The difference between [a] and [ɜ] was also clearly perceptible to our Akuapem language consultant, who described [ɜ] as sounding like the vowel [ɛ]” (where the transcription [ɜ] corresponds to [ɜ]). In the acoustic F1/F2 vowel space for many Akan speakers, [ɜ] is closer to [ɛ] than it is to [a]. He does not report on the perception of the other similar vowels for this consultant. Kpogo (2022) reports that there is an urban/rural distinction in the current Asante Twi dialect with respect to the [+ATR] harmonic counterpart of /a/. While rural speakers have a distinct tenth vowel ([ɜ] or [ɜ]), urban speakers produce [e] instead, exhibiting merger with the [e] corresponding to phonemic /e/. In addition to Akan, there are reports of acoustic similarity between certain vowels in other ATR systems, particularly the high [–ATR] and mid [+ATR] pairs [ɪ e] and [ʊ o] (e.g., Omamor, 1988 on Okpe and Uvwie; Koffi, 2018 on Anyi; Casali, 2003, 2008 for general discussion). It should be noted that with the exception of Kpogo (2022), who recorded data from 62 speakers, most acoustic studies are based on a small number of speakers, often less than 10.

1.3 Perception

Despite the plethora of phonetic studies on articulation and acoustics in languages with ATR harmony, there is very little research that examines perception, particularly perception by speakers of ATR languages.

Lindau (1975) conducted a study of the perception of Akan vowels by four non-Akan speaking phoneticians. Tokens of each vowel were produced by an Akyem dialect speaker (who apparently had both –ATR [ɜ] and +ATR [ɜ]) in a monosyllabic CV word in a carrier phrase and then the

vowel was repeated three times (e.g., meseɛ bu ʊ ʊ ʊ). The task required the four listeners to position a vowel produced in isolation on an F1/F2 chart. The composite results show that [o] and [ʊ] were perceived as very similar but still distinct, but there was a perceptual merger for [e] and [ɪ], and for [ɜ] and [ɛ]. Still, Lindau notes that the native speaker who produced the vowels was able to perceptually distinguish between [o] and [ʊ], and between [e] and [ɪ] on two separate occasions, once after a three-week interval and once after a four-week interval.

Hess (1992) conducted a small perception study as part of her acoustic assessment of Akan vowels. Specifically, she tested two Akan speakers on an identification task of single CV syllables in isolation containing [ɪ] [e] [ʊ] [o]. It is not stated if the syllables were real words or not. These vowels were chosen because each front or back pair are acoustically similar, especially for F1, suggestive of a merger. Speakers were tested on 40 items, with 10 tokens of each vowel. One speaker made two errors and the other speaker eight errors (87.5% accuracy), but it is not reported on which vowels or what kinds of errors. Hess concluded that despite the acoustic similarity of some of the vowels, they are perceptually distinct.

Fulop et al. (1998) report a perception study on Degema (Edoid, Nigeria) which has a 10 vowel ATR system (+ATR i e u o ə and –ATR ɪ ɛ ʊ ɔ a). Five Degema listeners performed a vowel identification task on synthesized vowels that manipulated F1, F2, and F3 in order to test how formants are used to distinguish vowels. Other elements such as fundamental frequency and phonation type (relative formant amplitudes) were kept constant. Listeners were given a written prompt of a particular vowel and asked to identify that vowel among a selection of tokens. Results showed the –ATR mid vowels ɛ ɔ as distinct from all other vowels, but considerable overlap within three groups of vowels: [i ~ ɪ ~ e], [u ~ ʊ ~ o], and [a ~ ə]; the study reports that listeners “do not behave alike when trying to synthesize any of the other vowels” (p. 96).

Kingston, Macmillan, Walsh Dickey, Thorburn and Bartels (1997) also used synthesized tokens designed to replicate variation in tongue root position and accompanying voice quality. This was done by manipulating F1, and the percentage or quotient of the glottal cycle in which the glottis was open and the overall tilt of the source spectrum. The experimental subjects were American English listeners. They report that F1 integrates perceptually with spectral effects of extreme lax or tense phonation.

A recent study (Ozburn, Giovio Canavesi, & Akinbo, 2022) examines perception of ATR distinctions in Dàgáárè, a Mabia language of Ghana, and is the only other study besides the current one to test perception with a large number of speakers (over 20) of a language with ATR contrasts. Stimuli were kV and kVkV nonce words using front vowels [i ɪ e ɛ] in an ABX task. Results showed low to medium rates of accuracy for three types of vowel contrasts (ATR i ~ ɪ, e ~ ɛ, and ATR/height ɪ ~ e), ranging from lowest performance with the ɪ ~ e contrast to highest with e ~ ɛ contrast. Disharmonic bisyllabic tokens resulted in lower accuracy rates

than monosyllables, but harmonic bisyllabic forms did not improve accuracy compared to monosyllables.⁴

Thus, while a few studies have examined perception of ATR vowel qualities in listeners who do not speak languages with ATR distinctions, or have examined synthetic vowel quality in listeners who do, there are very few perceptual tests of natural speech of ATR vowel qualities with listeners who speak languages with ATR distinctions. There are a few anecdotal reports that certain vowel distinctions may pose difficulties. The current study aims to fill this gap.

1.4. Akan vowel system and ATR harmony

Akan is a central Tano language of the Kwa language family. It has several dialects including Asante Twi, Akuapem, Akyem, and Fante. Specific descriptions of ATR vowel harmony in Akan date back as early as Christaller (1875), taken up again in Berry (1957), Stewart (1967, 1971, 1983), Schachter and Fromkin (1968), Clements (1981, 1984), Dolphyne (1988), and Casali (2012), among others. Recent papers by Akan-speaking linguists include Owusu (2014) and Abakah (2016).

Berry (1957) divided the vowels into two sets and noted that prefixes alternate between these sets according to the vowel of the root. Akan has nine phonemic vowels that contrast for ATR. The low –ATR vowel /a/ is unpaired with a phonemic +ATR low vowel, but does have an allophonic counterpart, transcribed as [ɜ] or [æ], and produced via vowel harmony.⁵

(1)			front	back
	high	+ATR	i	u
		–ATR	ɪ	ʊ
	mid	+ATR	e	o
		–ATR	ɛ	ɔ
	low	–ATR	a	

In general, all vowels in a word match for ATR:

⁴ The Ozburn et al. (2022) study was conducted online in 2021, whereas the current study was run prior to the COVID-19 pandemic in 2018 (Experiment 1) and early 2020 (Experiment 2), and was therefore able to be conducted in person in Ghana.

⁵ Stewart claims that the a/ɜ contrast is phonemic based on pairs such as *ntam* ‘oath’ vs. *ntɜm* ‘between.’ However, *ntɜm* derives from *ntá* ‘twin/pair’ + *mù* ‘inside,’ suggesting that the /ù/ harmonized the /á/ and then was deleted, leaving its low tone on the nasal. See also Appiah-Padi (1994) on this point. Even though Stewart acknowledges this derivation, he still views the distinction as phonemic. Nevertheless, the view that these two vowels are phonemic is a minority opinion. Kirkham and Nance (2017) state that they are phonemic in the Akuapem dialect based on earlier reports (Ladefoged, 1968; Dolphyne, 1988), but those reports do not distinguish phonemic from allophonic status clearly, indicating only the presence of [ɜ].

(2)	-ATR		+ATR	
a.	òwú	‘she delivers a baby’	d. òwú	‘death’
b.	èffè	‘vomit’	e. èffé	‘home’
c.	bòní	‘evil’	f. bòsómè	‘moon’
	<i>*bɔni</i>		<i>*ɛfie</i>	

In addition, prefixes and vowel-initial suffixes harmonize with the root for ATR, and therefore exhibit alternations between –ATR and +ATR:

(3)	-ATR		+ATR	
a.	wó-dà	‘you sleep’	c. wú-dì?	‘you eat’
b.	ò-tɔ̀-ì	‘s/he bought it’	d. ò-tú-ì	‘s/he dug it up’

Asymmetry in the system is noted by various authors. When /a/ precedes a +ATR vowel, it is realized as +ATR [ɜ] (or [æ] depending on transcription): e.g., *kɜri* ‘weigh’. However, when /a/ follows a +ATR vowel, it is realized as the –ATR [a]: *bisa* ‘ask’ or *èdʒá* ‘fire’, which can be construed as disharmonic forms. The vowel [ɜ] only occurs preceding a +ATR vowel and never alone in a monosyllable. This asymmetry prompted Casali (2012) to argue that Akan vowel harmony operates in a regressive direction. This argument is bolstered by two additional properties. Consonant-initial –ATR suffixes do not harmonize with a +ATR root (4a,c) and +ATR consonant-initial suffixes can trigger harmony leftwards, as shown by the agentive suffix -ni (4d). See also Schacter and Fromkin (1968, p. 62) on this point.

(4)	a. kúnú-nòm	‘husbands’	c. ò-sisi-fú	‘cheater’
	b. sɪká	‘money’	d. sɪkɜ́-ní	‘rich person’

Dolphyne (1988) treats [ɜ] as a [+ATR] allophonic variant of /a/. Clements (1981) argues that it is actually gradient coarticulation that raises /a/ to [ɜ] due to a following high [+ATR] vowel, but Hess (1992) and Casali (2012) dispute this analysis with acoustic evidence. Nevertheless, this vowel does show different distribution and behavior from the phonemic vowels. Although /a/ can become [+ATR] [ɜ], the harmonized vowel fails to trigger [+ATR] harmony to its left: *wò-bé-dí?* ‘they will eat’ vs. *wò-bé-kɜ̀rì* ‘they will weigh’. Therefore, in sequences of two /a/ preceding a +ATR high vowel, only the first one undergoes harmony. There are also reports that while /a/ becomes [ɜ] before high vowels, it does not do so when in a prefix before mid vowels in verbs (Stewart, 1967; Hess, 1992): *wà-bétú* ‘he has come and pulled it out’ vs. *wò-bísá* ‘he has asked’. We verified that this is the case with the Asante dialect. However, [ɜ] can occur before mid vowels in nouns, e.g., *ɜ̀kòd* ‘parrot’. Finally, there are reports of some dialects having other restrictions on mid +ATR vowels. Schacter and Fromkin (1968) report that Asante does not allow monosyllabic verb roots with just [e] or [o], and where the other dialects have these vowels in verb roots, it has [i] and [u].

1.5 Hypotheses on vowel perception in Akan

In order to test perception of ATR contrasts in Akan, we put forward two main hypotheses:

Hypothesis 1: Phonological contrast

Speakers of a language with ATR contrasts will have more difficulty distinguishing vowel pairs that have an allophonic relationship than those that have a phonemic relationship.

Hypothesis 2: Acoustic similarity

Speakers of a language with ATR contrasts will have difficulty perceiving vowel pairs that are acoustically similar, even if contrastive.

In general, phones that belong to separate phonemes are expected to be easier to distinguish than allophones of the same phoneme. Boomershine, Hall, Hume, and Johnson (2008) tested the perception of consonants that had either a phonemic or an allophonic relationship in English and Spanish. English speakers were better at perceiving the distinction between $d \sim \delta$ in a VCV context than Spanish speakers; this is a phonemic distinction in English but allophonic in Spanish. Similarly, Spanish speakers were better at perceiving the distinction between $d \sim r$ in a VCV context than English speakers, which is phonemic in Spanish but allophonic in English. However, other researchers have found that the phonological environment influences perception. Peperkamp, Pettinato, and Dupoux (2003) report that French speakers perform poorly at distinguishing the allophones $[\beta]$ and $[\chi]$ of $/\beta/$ preceding voiceless consonants. This is a context in which the voiceless allophone $[\chi]$ is expected. However, they perform well at distinguishing the same sounds in VC isolation contexts where the more widespread $[\beta]$ is expected. Overall, the results of these two studies suggest that allophonic pairs should be perceived as more similar to each other than phonemic pairs, but that context may impact the results if one allophone has a more restricted distribution. Applying this to the Akan case, the *Phonological contrast hypothesis* predicts that pairs that exhibit phonemic contrast for ATR are expected to be well distinguished. This includes the high vowel pairs $i \sim \text{ɪ}$ and $u \sim \text{ʊ}$, and the mid vowel pairs $e \sim \text{ɛ}$, $o \sim \text{ɔ}$. It also includes the acoustically similar pairs $\text{ɪ} \sim e$ and $\text{ʊ} \sim o$. Applying the general principle of phonemic contrast, the low vowel pair $\text{ɜ} \sim a$, where $[\text{ɜ}]$ is allophonic and only occurs preceding +ATR vowels, is expected to be poorly distinguished. However, given that $[\text{ɜ}]$ is a contextual +ATR vowel harmonic allophone of $/a/$, it is possible that better perception between $[\text{ɜ}]$ and $[a]$ will occur in an isolation context where the default $[a]$ is expected, following the results of Peperkamp et al. (2003).

The *Acoustic similarity hypothesis*, in contrast, predicts that the main driving factor in perceptual similarity will be acoustic similarity (Dubno & Levitt, 1981). That is, among phonologically contrastive categories, there may be differences in acoustic similarity which impact perception. This could also override allophonic relationships between segment pairs.

Furthermore, distribution and frequency in the lexicon can impact how segments in a phonemic contrast relationship are perceived. Hall, Letawsky, Turner, Allen, and McMullin (2015) found that phonemic fricatives in contexts where only one of the sounds is highly likely to occur are perceived as more similar than the same sounds in contexts where either is likely to occur. With respect to ATR, in some languages the high vowels are acoustically similar in the F1/F2 vowel space (e.g., Ikposo (Starwalt, 2008), Masaai (Quinn-Wriedt, 2013), Lopit (Billington, 2017)). In other languages, like Akan, they are more separated, and the mid +ATR vowels have lower F1 than the high [-ATR] vowels, what Casali (2012, p. 49) terms a ‘leapfrog’ pattern or what can be called a ‘flip’ in the vowel space. Therefore, in terms of acoustic similarity, high ATR pairs may not pose difficulty.

As reported above, however, certain vowels are very similar acoustically, even if distinguished by phonological properties of ATR and vowel height: [ɪ] / [e] and [ʊ] / [o]. There are anecdotal reports of fieldworkers having difficulty distinguishing these vowels (Casali, 2017), although it is not well known if speakers of languages with ATR contrasts can distinguish among them, except for the results of the Degema perceptual study (Fulop et al., 1998) and the Dàgáárè study (Ozburn et al., 2022), and anecdotal reports in Omamor (1988) and Koffi (2018). Koffi (2018) reports that Anyi speakers confused /e/ and /ɪ/ in literacy training. Omamor, a speaker of Uvwie, notes that acoustic measurements helped reveal that Uvwie had the vowels /ɪ/ and /ʊ/, which were hard to distinguish from other vowels. Notwithstanding, speakers of an ATR language are expected to perform better than speakers of a non-ATR language in detecting ATR contrasts because they have acquired a language that employs these contrasts, and so we assume their perceptual system is attuned to the perceptual cues of ATR contrasts. Although we do not directly test the ATR vs. non-ATR language speaker hypothesis in the study, this is the reason speakers of an ATR language were tested.

The acoustic similarity hypothesis would predict that the distinction between such vowels would be hard to perceive, despite their phonological feature contrast. As for the vowel [ɜ], its F1/F2 position suggests that it will be more likely to be confused with [ɛ] rather than [a] based on acoustic similarity. This is so even if [ɜ] maps to the phoneme /a/ and /a/ and /ɛ/ are differentiated by frontness and vowel height (in phonological features [low] and [back]). In order to calculate acoustic similarity more precisely, we measured Euclidean distance of vowel pairs using the F1/F2 measurements of vowel productions used in the stimuli in the experiments (**Table 1**). Each formant was z-scored prior to calculation so that F2, with its higher raw Hz values, did not dominate distances. The precise distances are provided in Appendix C, but the results reveal three broad categories of distance, with vowel pairs distinguished by both ATR and height features having the smallest distance. It is assumed that the smaller the Euclidean distance measurement, the greater the similarity between vowels.

Acoustic similarity	Vowel pairs	Category
Low similarity, dist. ≥ 1.00	i ~ ɪ, e ~ ɛ, u ~ ʊ, o ~ ɔ, a ~ [ɜ], ɛ ~ a	ATR or Height
Medium, dist. 0.61–0.99	i ~ e ɪ ~ ɛ u ~ ʊ ʊ ~ ɔ	Height
High similarity, dist. ≤ 0.60	ɪ ~ e ʊ ~ o ɛ ~ [ɜ]	ATR + Height

Table 1: Acoustic similarity: z-scored Euclidean distances between vowels. See Appendix C for details.

The *Phonological contrast hypothesis* does not predict a difference in perception based on the number of phonological features that two vowels may share or not share. Nevertheless, it is instructive to examine how phonological similarity calculations differ from acoustic similarity. Frisch, Broe and Pierrehumbert (2004) propose a phonological similarity metric using a formula based on distinctive phonological features and shared and unshared natural classes. Similarity = shared natural classes / shared natural classes + unshared natural classes. Using this method, the lower the number, the less phonologically similar the vowels are. The least phonologically similar phonemic vowels are the ATR + Height vowels ɪ ~ e and ʊ ~ o, while the other vowels have higher similarity (**Table 2**). See Appendix D for details. Note that because this metric of similarity is based on distinctive phonological features, only phonemes are included for assessment.

Similarity	Vowel pairs	Category
Low similarity, 0.16	ɪ ~ e ʊ ~ o	ATR + Height
High similarity, > 0.30	i ~ ɪ u ~ ʊ e ~ ɛ o ~ ɔ i ~ e u ~ o ɪ ~ ɛ ʊ ~ ɔ ɛ ~ a	ATR or Height

Table 2: Phonological similarity based on natural classes.

These two measures of similarity produce contradictory results for the ATR + Height vowels compared to the other vowels. The ATR + Height vowels have high acoustic similarity (that is, low Euclidean distance), but low phonological similarity. Acoustic similarity measurements suggest that distinguishing ATR + Height vowels will pose perception difficulties, while phonological similarity suggests they should be well-differentiated.

The predictions of the two hypotheses with respect to ATR differences are presented below (**Table 3**) with respect to predicted performance on a perceptual discrimination task. Both hypotheses predict that the non-low ATR vowels will show good discrimination in Akan, based on phonological contrast or the separation in acoustic space. The two hypotheses differ with respect to the other vowel pairs. The Phonological hypothesis predicts that those pairs distinguished by both ATR and vowel height will be easily discriminated, but the Acoustic similarity hypothesis predicts difficulty. The opposite predictions are made for the low ATR pair. As the relationship is allophonic, perceptual

Label	Vowel pairs	Phonological	Acoustic
Non-low ATR	i ~ ɪ e ~ ɛ u ~ ʊ o ~ ɔ	✓	✓
Height	i ~ e ɪ ~ ɛ u ~ ʊ ɔ ~ ɔ	✓	✓
ATR+Height	ɪ ~ e ʊ ~ o ɛ ~ ɔ	✓	✗
Low ATR	a ~ ɔ	✗	✓

Table 3: Predictions of Phonological and Acoustic similarity hypotheses.

difficulty is predicted under the Phonological hypothesis, whereas the Acoustic hypothesis predicts that, as they are not phonetically similar, good discrimination should result.

2. Experiment 1

2.1 Methods

2.1.1 Participants

Forty-one Akan speakers participated in the experiment (26 M, 15 F) in September 2018 in Ghana. Participants ranged in age from 19 to 28 years old. All subjects were recruited at the University of Education, Winneba, Ghana. Thirty of the speakers spoke the Asante Twi dialect, five spoke both Twi and Fante, three spoke Twi and Sefwi, one spoke Twi and Nafara, one spoke Fante, and one spoke Akuapem as well as the Mabia language Frafra. The participants were compensated at a rate consistent with local norms for study participation (5\$). Written instructions were presented in Akan and the researcher (Michael Obiri-Yeboah) spoke in Akan to the participants. Each participant completed a questionnaire on their language background and they also completed a separate music perception experiment (20 of the subjects preceding the vowel perception experiment and 21 subjects following the vowel perception experiment).

2.1.2 Stimuli

Stimuli were audio recordings of CV monosyllables. Monosyllables were chosen to avoid the influence of vowel harmony and to allow the testing of perception in isolation. One hundred twenty-four test trials and 54 filler trials were used in the experiment. All trials consisted of two consonant-vowel (CV) low-toned syllables. Sixty ‘same’ trials consisted of a test syllable paired with another production of the same syllable, e.g., kà ~ kà, where each syllable was recorded separately. Thirty syllables had the onset /k/ combined with all 10 vowels, so that there were three tokens of each CV combination. Three different tokens were used so that no ‘same’ pair contained the same two tokens (e.g., kà1 ~ kà2, kà2 ~ kà3, kà3 ~ kà1). Another 30 syllables had the onset /s/ combined with all 10 vowels, with three repetitions of each combination. C̀ syllables were extracted from bisyllabic C̀C̀V̀ words and the duration normalized to approach

duration of other vowels. This was done as [ɜ] only occurs preceding +ATR vowels, and the speaker was not able to produce this vowel in CV in isolation.

Sixty-four ‘different’ trials consisted of low-toned CV syllables with either [k] (32) or [s] (32) onsets. These syllables combined different vowels counterbalanced in both orders with the same onset (e.g., ki ~ ke and ke ~ ki). Each vowel pair occurred in four trials, with both onsets and both orders. Vowel combinations were divided into four classes: 1) Point vowels, 2) ATR contrasts, 3) Height contrasts, and 4) ATR + Height contrasts as in **Table 4**. Point vowels were selected to be maximally distinct and serve as a confirmation of general task understanding. These are the vowels most dispersed in the acoustic vowel space (highest and lowest F1/F2).⁶ Height contrasts were selected as a counterpoint to ATR. They also differ on one phonological feature and are distinguished primarily by F1. Finally, those vowels which differed on both ATR and Height were vowels that are the most acoustically similar. Rounding and back features were kept constant for these contrasts, although [ɛ] is [+front, -back] and [ɜ] is central or [-front, -back].

	Different vowels	Number of test items
Point	i ~ u u ~ a a ~ i	12
ATR	i ~ ɪ u ~ ʊ e ~ ε o ~ ɔ [ɜ] ~ a	20
Height	i ~ e u ~ o ɪ ~ ε ʊ ~ ɔ ε ~ a	20
ATR + Height	ɪ ~ e ʊ ~ o ε ~ [ɜ]	12

Table 4: ‘different’ vowel combinations.

Fifty-four Filler items consisted of CV syllables with one of 10 onsets drawn from /f b m t d tɕ s k kp h/ and one of 10 vowels. All these onsets except [kp] occur in Akan, but [kp] is common in other Ghanaian languages. There were 28 same items consisting of two different tokens, and 26 different items, 20 with the same vowel but a different consonant, and six with a different vowel and a different consonant. Fillers were included to provide variety and to disguise the experimental manipulation, so they were not analyzed.

Stimuli were recorded by a multilingual speaker (Michael Obiri-Yeboah) of Gua (Kwa, Ghana) and Akan (Akuapem dialect), pronouncing the tokens as if speaking Gua. The Gua vowel system is very similar to that of Akan. It has the same nine contrastive vowels and a tenth allophonic vowel that is also the [+ATR] counterpart of /a/. It has +ATR dominant regressive harmony like Akan (Casali, 2012; Obiri-Yeboah, 2021). We reasoned that a perceptual task using Akan monosyllables would be too easy, such that participants would perform well on most vowel

⁶ Euclidean distance measures are $i \sim u = 1.888$, $i \sim a = 3.67$ and $u \sim a = 3.074$, all greater distance measures than other vowel pairs in the experiment.

pairs. By using a language with a vowel system very close to Akan's, we expected participants to perceptually map their Akan vowel system onto the Gua vowels they heard, as per second language models of vowel perception, such as the Perceptual Assimilation Model (Best, 1995; Tyler et al., 2014) or the L2 Linguistic Perception Model (L2LP) (Escudero, 2009). Gua high -ATR and mid +ATR vowels are acoustically similar and are 'flipped' in acoustic space, as in Akan. Ultrasound images from the same speaker of Gua shows tongue root advancement/retraction (Myers et al., to appear). We measured the F1/F2 properties of the vowel tokens used in the study (both fillers and test items), as shown in **Figure 4**; mean values are indicated by the position of the vowel label. These results are similar to those reported for Akan, although [u] is pronounced with higher F2 following coronal consonants [s] and [tɕ] than following [k] (the three points further back).

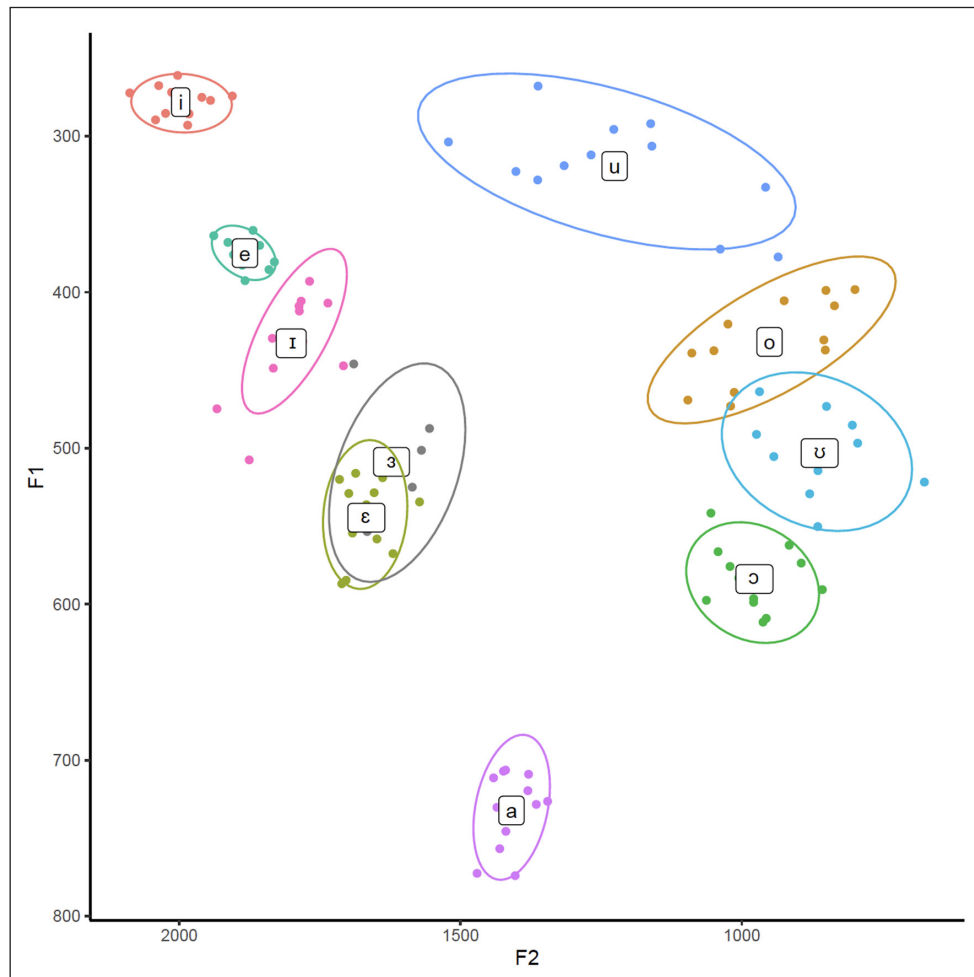


Figure 4: F1/F2 measurements of vowel tokens used in Experiment 1. Ellipses show 85% Confidence Interval.

2.1.3 Procedure

An auditory AX discrimination task (same/different) was presented to participants using PsychoPy on a laptop through headphones. The interstimulus interval (ISI) was set at 500 ms. The duration of the ISI can affect processing (Babel & Johnson, 2010; Pisoni, 1973, 1975; Werker & Logan, 1985, among others). Short ISIs are thought to encourage acoustic processing whereas longer ISIs are thought to favor phonological or lexical processing (Piccinini & Arvaniti, 2019). Babel and Johnson (2010) determined that ISIs 500 ms or above are sufficient for phonological processing. We, therefore, selected an ISI that could encourage phonological processing, but was still short enough that acoustic processing might also result, so as not to bias processing towards a particular hypothesis. Selecting an ISI that was too short would favor acoustic processing, while choosing one that was too long would favor phonological processing. Participants were told that they would hear words from another language and would be asked if the two words sounded the same or different. They pressed button ‘a’ on the keyboard for same (*adekorɔ* in Akan) or ‘s’ for different (*soronko* in Akan). Items were presented in random order to the participants. There was a short practice trial (with no feedback) to habituate participants to the experiment. The experiment took 15–25 minutes to complete.

2.2 Results

One participant was removed from the results, as they consistently showed poor performance across all items (5.2 SD below mean accuracy). The results of the remaining 40 participants are shown below. **Figure 5** shows the results of the ‘different’ test items, with proportion “different” responses (hits) along the y-axis. Responses to all the items are included in Appendix A.

Figure 5 suggests that point vowels were well discriminated, as were all ATR pairs, including the allophonic +ATR [ɜ] and its –ATR counterpart [a]. Height pairs were also well discriminated, but with slightly lower rates for the round vowels. However, the ATR + Height combination pairs were poorly discriminated, with responses of “different” that were not far above the false-alarm baseline.

For purposes of analysis, data were converted to d' scores (**Figure 6**) by subtracting z-scored false alarms from z-scored hits for a particular vowel pair. We used all “same” trials to create the false alarm baseline. To correct for values of 1 and 0, which yield z-values of \pm infinity, we added or subtracted $1/60/2 = 1/120$ —half of an item in the condition with the most items, the “same” condition. These d' scores were then subjected to a within-subjects analysis of variance (ANOVA) with Vowel Type (point vowels; ATR-differing; height-differing; both-differing) as the predictor. Analyses were conducted in R (R Core Team, 2020).

The effect of Vowel Type reached significance ($F(3,117) = 297.7, p < .0001$). To assess the meaning of this result, we compared all vowel types to each other with Bonferroni correction for multiple comparisons. Bonferroni correction sets the criterial p-value at .0083 for six comparisons.

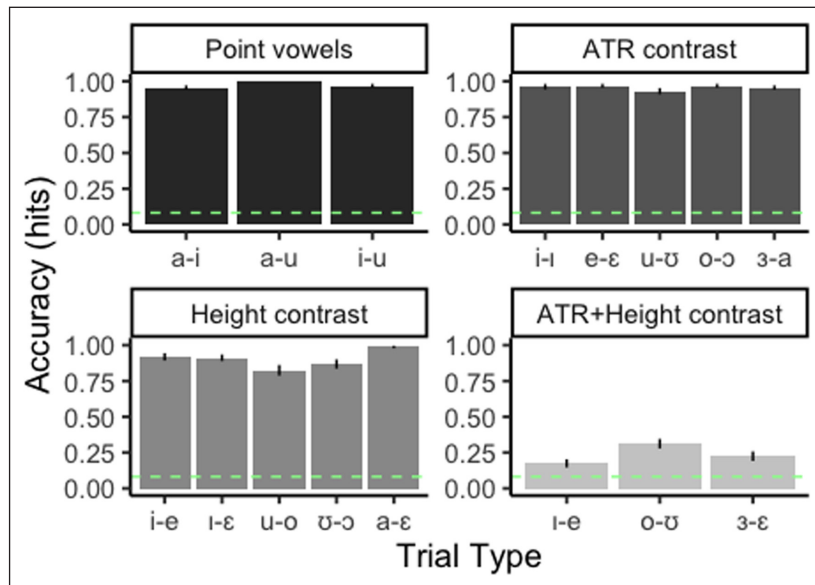


Figure 5: Experiment 1, proportion hits to all tested vowel contrasts, with standard errors. Dashed line represents the false-alarm baseline, that is, how often listeners responded “different” when the two vowels heard were actually the same vowel. Most figures were constructed using the ggplot2 library (Wickham, 2016) in R (R Core Team, 2020).

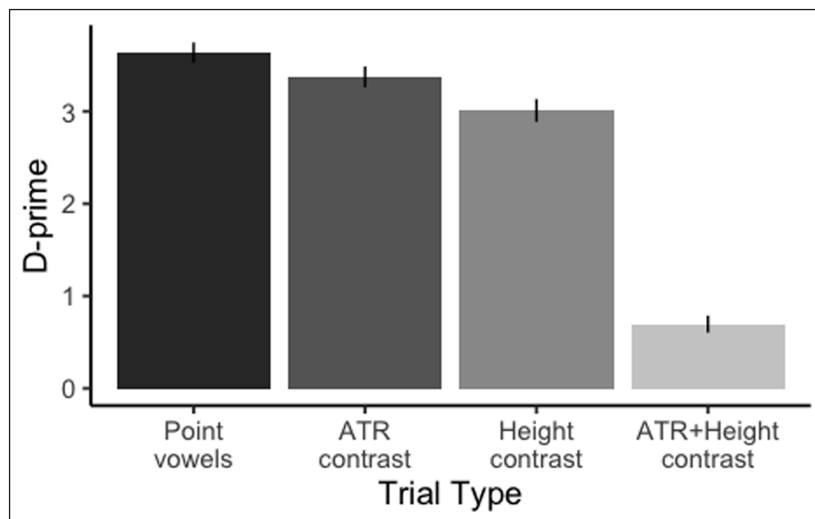


Figure 6: Experiment 1, d' scores for each vowel contrast type with standard errors.

Point vowels were marginally better discriminated than ATR vowels ($t(39) = 2.62, p = .01$; marginally significant after Bonferroni correction) and significantly better discriminated than Height vowels ($t(39) = 5.76, p < .0001$). Height vowels were discriminated significantly worse than ATR vowels ($t(39) = 3.26, p = .002$), and the ATR+Height combination showed weaker

performance than all other conditions ($t(39) \geq 19.16, p < .0001$). Nonetheless, all conditions, including the ATR+Height combination, exceeded chance performance (chance meaning that hits = FAs, for a $d' = 0$; $t(39) \geq 7.57, p < .0001$).

To better understand the source of the outcome, we examined hit rate differences amongst individual pairs. The rate for all three of the ATR+Height pairs fell well below the other conditions, and indeed none of the 40 participants achieved 100% detection. The highest accuracy in the ATR+Height condition was by a single participant who scored 58% on “different” trials. In fact, two participants detected none of the ATR+Height differences, even though they performed very well on the other pairs in the experiment. Yet the $a \sim [3]$ pair, despite being an allophonic distinction, was very well discriminated, with 95.6% of changes detected (SD = 9.6%; 33/40 participants, or 83%, detected differences on all four trials with this contrast). A follow-up test on d' values for individual vowels revealed that the allophonic $a \sim [3]$ pair was as well-discriminated as the other, phonemic ATR pairs considered together ($t(39) = 1.73, p = .09$), with a d' value of 3.59 (SD = .816), numerically *higher* than the d' of the other four ATR pairs ($d' = 3.39, SD = .730$).

Finally, to provide further support for the Acoustic Hypothesis, we conducted a significance test of the correlation between the Euclidean distances between vowel pairs and accuracy on different trials. There was a significant positive linear relationship between the two measures ($t(11) = 4.12, r = .779, p = .002$), as shown in **Figure 7**.

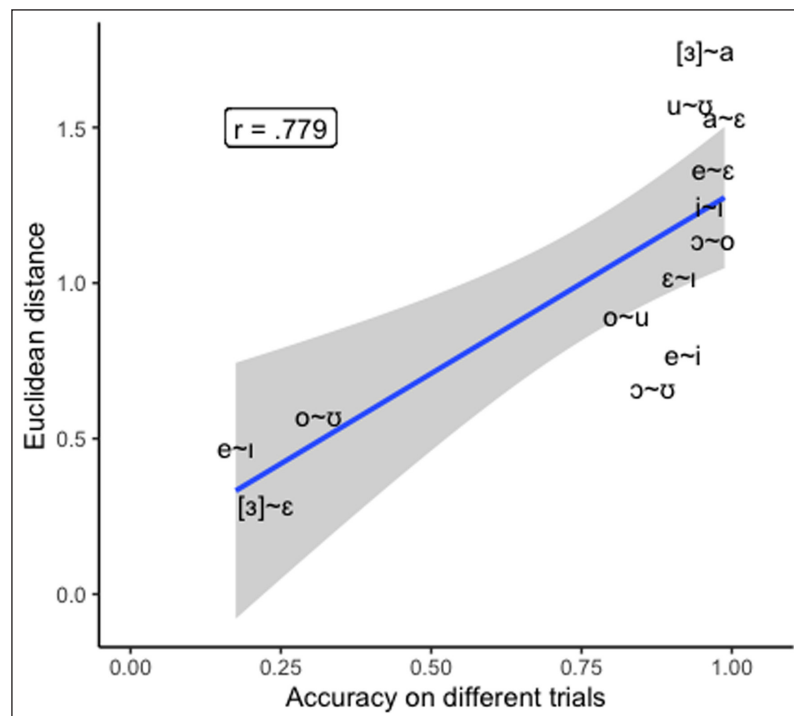


Figure 7: Experiment 1, accuracy on different trials x acoustic Euclidean distance scores.

2.3 Discussion

The results were not consistent with the Phonological contrast hypothesis (Table 5). Indeed, participants performed poorly on three phonological contrasts that were acoustically similar. In contrast, participants were able to distinguish ATR pairs with ease, whether they were the non-low phonemic pairs or the low pair that constituted a single phoneme. The pairs with which participants had difficulty were precisely those predicted by the Acoustic similarity hypothesis. Moreover, the fact that the pair $e \sim \text{ɪ}$ exhibits the numerically poorest discrimination is congruent with earlier reports (Lindau, 1975; Hess, 1992) that this vowel pair is merging acoustically in Akan, while the round vowel pair $[o \sim \text{ʊ}]$ is less similar.

	Vowel pairs	Phon. Hyp.	Acoustic Hyp.	d'	% different
Non-low ATR	$i \sim \text{ɪ} \quad e \sim \text{ɛ} \quad u \sim \text{ʊ} \quad o \sim \text{ɔ}$	✓	✓	3.39 (0.73)	95.7%
Height	$i \sim e \quad u \sim \text{ʊ} \quad \text{ɪ} \sim \text{ɛ} \quad \text{ʊ} \sim \text{ɔ} \quad \text{ɛ} \sim a$	✓	✓	3.01 (0.78)	90.3%
ATR + Height	$\text{ɪ} \sim e \quad \text{ʊ} \sim \text{ɔ} \quad \text{ɛ} \sim \text{ɜ}$	✓	✗	0.70 (0.58)	23.7%
Low ATR	$a \sim \text{ɜ}$	✗	✓	3.59 (0.82)	95.6%

Table 5. Experiment 1 predictions compared to results.

The result for the low ATR allophonic pair was in line with Peperkamp et al. (2003), who found that allophonic fricatives in French are perceived well in isolation contexts. We suspect that $[\text{ɜ}]$ is likely being perceptually ‘matched’ to $/\text{ɛ}/$, as they are highly similar. If $[\text{ɜ}]$ is perceptually ‘matched’ to $/\text{ɛ}/$, this explains why the $a \sim \text{ɜ}$ pair is easy to distinguish, as the $a \sim \text{ɛ}$ pair had a very high accuracy rate. The strong performance on the non-low ATR pairs is in line with both hypotheses. Unlike some other ATR languages, high vowels in Gua and Akan have relatively large F1 differences, due to the ‘flip’ in vowel space of high –ATR and mid +ATR vowels. This acoustic separation leads to better perception.

The fact that acoustic similarity drives the perception over phonological contrast could be due to several factors. One possibility is the nature of the task itself. The AX task may be more suited to acoustic perception than phonological discrimination. Subjects are simply asked if two tokens are the same or different with little contextual information. Relatedly, the interstimulus interval of 500 ms may not have been long enough to favor phonological processing. Although we did not want to bias the result one way or another, a longer ISI might have facilitated phonological processing. A second issue with the task is that participants were told the words were from another language, so this might have discouraged Akan phonological processing, even if the tokens sounded like Akan vowels. As the words were in fact from another (albeit closely-related) language, Akan speakers might have been at a disadvantage at distinguishing between highly similar vowels if they had slightly different pronunciations.

A third possibility is that orthographic similarity influenced responses to auditory stimuli. Modern Akan orthography uses seven vowels instead of nine. Precisely, the phonemic vowels that are hard to distinguish are written with a single character. Both *ɪ* and *e* are written ‘e’ and *ʊ* and *o* are written ‘o’. Historically, the Gold Coast orthography distinguished these similar vowels with diacritics so that *ɔ̄* (*u*) contrasted with *o* and *ɛ̄* (*i*) with *e*. However, the Standardized Akan orthography, intended to bridge dialects, no longer uses diacritics (Akan Language Committee, 1995). It is possible that the absence of visual differentiation between these vowels leads literate Akan listeners to conflate them auditorily. Of course, this is a chicken-and-egg problem, in that it is not clear whether this orthographic change led to the perceptual difficulties or whether the vowels were already perceived to have merged or be similar. Yet, allophonic [ɜ] and [a] are both written as ‘a’, and this pair was successfully discriminated, so orthographic influence cannot be the whole story.

A fourth possible explanation for the acoustically-driven outcome is that it is shaped by the frequency and distribution of the vowels. First, vowel harmony in polysyllabic words ensures that *o* and *ʊ*, or *e* and *ɪ* generally do not cooccur within the same word. So, distinguishing between them may not be required that often if the presence of other vowels that are more easily discerned in the word signifies the –ATR or +ATR status of other vowels in the word.⁷ The AX task with monosyllables may have been difficult precisely because there were no other contextual clues on which to rely. As previously noted, /e/ and /o/ are reported to be rare in monosyllables. Berry (1957, p. 125) states they are rare in CV monosyllables in the Akuapem dialect and of doubtful occurrence in Asante. Dolphyne (1988, p. 98) also reports this for Akuapem and some subdialects of Fante, and notes that /o/ occurs in a few words only in Asante. Most of the participants in our study were Asante speakers. Although it is generally assumed that Proto-Kwa had a nine-vowel system (Williamson, 1973), Abakah (2016) suggests that /e o/ arose historically via vowel harmony from a seven-vowel system. If this is correct, it would explain their more limited distribution. This does not necessarily mean that these vowels are not phonemic, however. Although rare in monosyllables, they do occur in some verb stems in some dialects (e.g., *s̄e* ‘say’ *d̄ze* ‘receive’), and they do appear word-finally in polysyllabic nouns (e.g., *bròbè* ‘taro’, *dédè* ‘noise’, *m̄kó* ‘pepper’ or *̀búrò*: ‘maize’. As vowel harmony is regressive (Casali, 2012), /e o/ do appear to be phonemic. Nevertheless, it is possible that predictability of distribution could hinder perception in monosyllabic CV context even with phonemes if [–ATR] *ɪ ʊ a* are expected over the less common [+ATR] *e o ɜ*. This is in line with Hall et al. (2015), who found that phonemic fricatives in more predictable contexts in English are perceived as more similar than the same sounds in less predictable contexts.

⁷ Akan does have cross-word vowel harmony (Schachter, 1989; Dolphyne, 1988; Hess, 1992; Kügler, 2015) that applies regressively and affects the last vowel of the preceding word, so disharmonic words can arise in sentential contexts due to this process.

To address several of these issues, we designed a second experiment. In order to address the concern that monosyllabic contexts may not be as representative, the second experiment tested bisyllables. Bisyllables also provide the opportunity to test out disharmonic sequences to assess whether disharmonic sequences are more difficult to distinguish than harmonic sequences. To encourage phonological processing, the second experiment was an ABX task, where listeners were asked to determine whether nonce word X better matched nonce word A or nonce word B. To best assess listeners' Akan spoken language knowledge, the vowels were recorded as nonce words in Akan, and listeners were told that these were to be construed as Akan words, ruling out language mismatch as an explanation for the results of Experiment 1.

3. Experiment 2

3.1 Methods

3.1.1 Participants

Forty-one Akan speakers participated in the experiment (21 M, 20 F) in February and March, 2020, in Ghana. Participants ranged in age from 19 to 36. All subjects were recruited at the University of Education, Winneba, Ghana, but none were the same as those in the first experiment. Thirty-five of the speakers spoke the Asante Twi dialect, one spoke Akuapem, one spoke Akyem, one spoke Kwahu, and three were bidialectal (Asante/Akyem (two) or Asante/Akuapem (one)). The participants were compensated (5\$). Written instructions were presented in Akan and the researcher (Michael Obiri-Yeboah) spoke in Akan to the participants. Each participant completed a questionnaire on their language background and they also completed a separate music perception experiment (20 of the subjects preceding the vowel perception experiment, and 21 subjects following the vowel perception experiment).

3.1.2 Stimuli

Experiment 2 consisted of 272 trials. Each trial consisted of three words, two of which were different tokens of the same word, whereas the other contained vowels that mismatched along one or more dimensions. Two consonant frames were used – gVbV and tVkV – and all tokens had low tone. These frames did not correspond to any actual Akan words, as verified by two Akan speakers. In order to include [ɜ] in the tokens, this vowel was produced preceding the vowel [i] in tokens such as [tɜki] or [kɜti], and then the syllable in which it appeared was spliced with another syllable. No other tokens were spliced in order to keep the tokens sounding as natural as possible.

Set A consisted of 96 trials where each token presented contained two *identical* vowels, e.g., A. gùbù B. gòbù X. gùbù. The vowel pairs compared were the same as used in Experiment 1, encompassing ATR, Height, and ATR + Height differences, except point vowels were not used, and we dropped the [ɛ] ~ [a] height combination to reduce the number of trials. There were

eight trials per combination so that the order of A and B were switched and X was matched to either A or B, and both consonant frames were employed. All of the bisyllables in these trials were necessarily harmonic given that the vowels in each word are identical.

Set B consisted of 64 *harmonic non-identical* vowel trials, where harmonic means that within a word, the vowels agreed for ATR, and non-identical means that within a word, the two vowels were not identical. These consisted of trials where either A or B had non-identical vowels, the other word had identical vowels and the X always had identical vowels (e.g., A. gubu B. gubo X. gubu or gobo). This meant that there was only one vowel across all three words that was different. The trials contained words in which both orders internal to the word (gubo and gubu) were used, and they also varied whether the non-identical word was A or B, resulting in eight combinations. Adding in the two different consonant frames, there were 16 trials per pair. The harmonic restriction meant that only height differences were employed in these trials: $i \sim \text{ɪ} \ u \sim \text{ʊ} \ e \sim \text{ɛ} \ o \sim \text{ɔ} \ [ɜ] \sim \text{a}$.

Set C consisted of 112 *disharmonic non-identical* vowels. They followed the same format as Set B, except that the vowels within the word mismatched for ATR, and either had identical height (e.g., A. gubu B. gubu X. gubu), or mismatched for both ATR and height (e.g., A. gubo B. gobo X. gobo). We did not test the $[\text{ɛ}] \sim [\text{ɜ}]$ distinction in this set to reduce the number of trials, due to concern over the length of the experiment. Note that disharmonic within a word does not necessarily mean phonotactically illegal in Akan. As harmony is regressive and +ATR dominant, sequences of -ATR +ATR are phonotactically illegal within a word, but the reverse sequence of +ATR -ATR is disharmonic but phonotactically legal, as in the examples in (4). Not all vowel combinations occur in this order though. The stimuli are summarized in **Table 6**:

	Vowel pairs	Trials
Set A Harmonic Identical	gobu/gubu	
ATR	$i \sim \text{ɪ} \ u \sim \text{ʊ} \ e \sim \text{ɛ} \ o \sim \text{ɔ} \ [ɜ] \sim \text{a}$	40
Height	$i \sim \text{e} \ u \sim \text{o} \ \text{ɪ} \sim \text{ɛ} \ \text{ʊ} \sim \text{ɔ}$	32
ATR + Height	$\text{ɪ} \sim \text{e} \ \text{ʊ} \sim \text{o} \ \text{ɛ} \sim [ɜ]$	24
Sub-Total		96
Set B Harmonic non-identical	gobu / gubu	
Height	$i \sim \text{e} \ u \sim \text{o} \ \text{ɪ} \sim \text{ɛ} \ \text{ʊ} \sim \text{ɔ}$	64
Set C Disharmonic non-identical	gobu / gubu or gobo / gobo	
ATR	$i \sim \text{ɪ} \ u \sim \text{ʊ} \ e \sim \text{ɛ} \ o \sim \text{ɔ} \ [ɜ] \sim \text{a}$	80
ATR + Height	$\text{ɪ} \sim \text{e} \ \text{ʊ} \sim \text{o}$	32
Sub-Total		112
Total		272

Table 6: Experiment 2 list of stimuli.

Stimuli were recorded by a multilingual speaker (Michael Obiri-Yeboah) of Gaa (Kwa, Ghana) and Akan (Akuapem dialect), pronouncing the tokens as if speaking Akan. This change in the language of the stimuli from Experiment 1 ensured that the subjects knew they were processing Akan. Results should therefore reflect Akan L1 processing and not processing of an unfamiliar language with a similar vowel system. The acoustic F1/F2 vowel space of the tokens is provided below in **Figure 8**, measured from both syllables of the identical vowel tokens in the experiment. As expected, there is overlap between [i] and [e], and between [ɛ] and [ɜ]. There is very little overlap between [u] and [o]. This is in line with other reports that the acoustically similar front vowels are closer than the back ones in Akan.

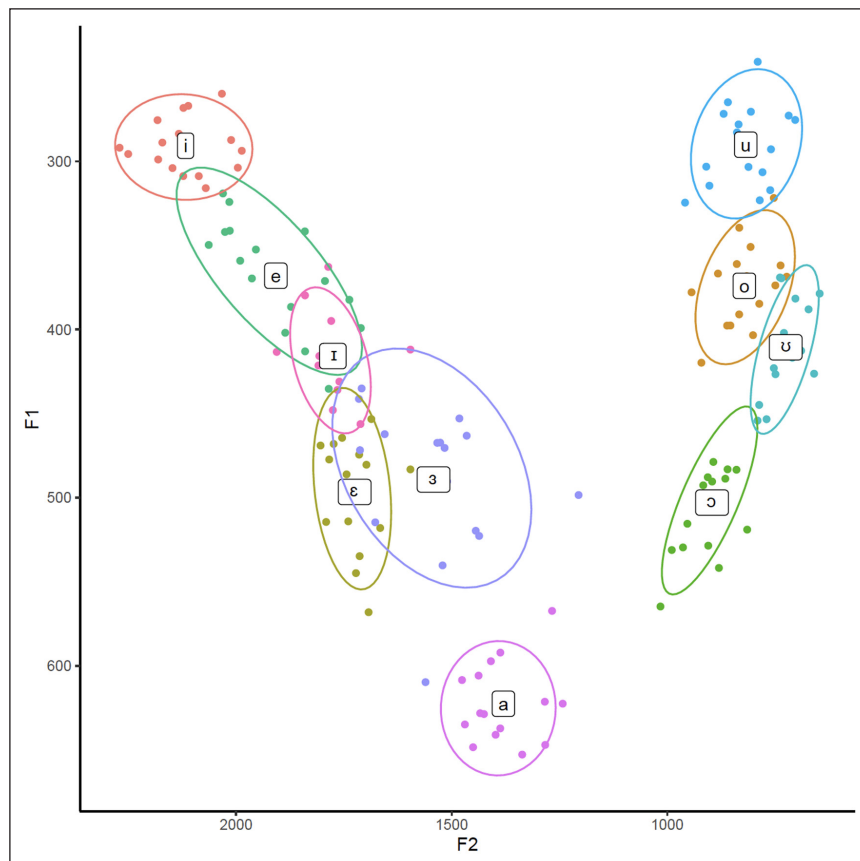


Figure 8: F1/F2 measurements of vowels in identical tokens in Experiment 2; Ellipses show 85% Confidence Interval.

3.1.3 Procedure

An auditory ABX discrimination task was presented to participants using PsychoPy on a laptop through headphones. The interstimulus interval (ISI) was set at 700 ms between the first and second and between the second and third stimuli. This longer ISI was aimed to encourage

phonological processing (Piccinini & Arvaniti, 2019). Participants were told that they would hear new, made-up words in Akan. They were asked if the last word sounded the same as the first or the second. They pressed button ‘b’ on the keyboard for ‘first’ (*baako* in Akan) or ‘m’ for ‘second’ (*mmienu* in Akan). Items were presented in random order to the participants. There was a set of four practice trials, half answer ‘first’ and half answer ‘second’, using distinctly-different words (i.e., *teke-gubu-teke*) and providing accuracy feedback to acclimate participants to the procedure. If participants missed one of these four easy items, the practice trials were repeated. After one repetition of the practice trials, participants continued to the main experiment regardless of score.

3.2 Predictions

Based on the results of Experiment 1, we make several predictions for the results of Experiment 2 in accordance with the Acoustic Hypothesis (**Table 7**). With respect to the identical vowel pairs (Set A), we expect to see excellent discrimination for ATR contrasts and for Height contrasts, but poor performance for vowels contrasting in both ATR and Height. Given the outcome of Experiment 1, we also expect that ATR pairs will be discriminated slightly better than Height pairs. Compared to Experiment 1, the bisyllables may also afford speakers more opportunity to detect distinctions, since there is not reliance on a single vowel, possibly leading to high accuracy.

As for the non-identical vowel tokens (Sets B and C), we expect an overall poorer discrimination rate for these sets than identical (Set A), as there are two opportunities to hear the vowel distinctions (*gibi* vs. *gebe*) with identical vowels, but only one opportunity with the non-identical categories (*gibi* vs. *gibe*). At a more detailed level, the role of vowel harmony should impact the results. The Harmony hypothesis (**Table 8**) predicts that Set B, where all vowels match for ATR and are harmonic, should show better discrimination than disharmonic forms (that is, Set C).

3.3 Results

As with Experiment 1, the acoustically similar pairs had lower rates of accurate perception compared to the ATR pairs and the Height pairs (**Figure 9**). Detection rates for ATR pairs and Height pairs in the Identical stimuli (Set A) were high. For completeness, accuracy for individual sound pairs are shown in Appendix B.

For maximum parallelism to Experiment 1, we computed d' for this experiment as well. The d' values for ABX tasks are computed slightly differently, as follows. Correct trials where the expected response was the A (first) item were counted as hits, while correct trials where the expected response was the B (second) item were counted as correct rejections. Correct rejections were converted to false alarms ($FA = 1 - CR$) and then hits and FAs were used to compute d' .

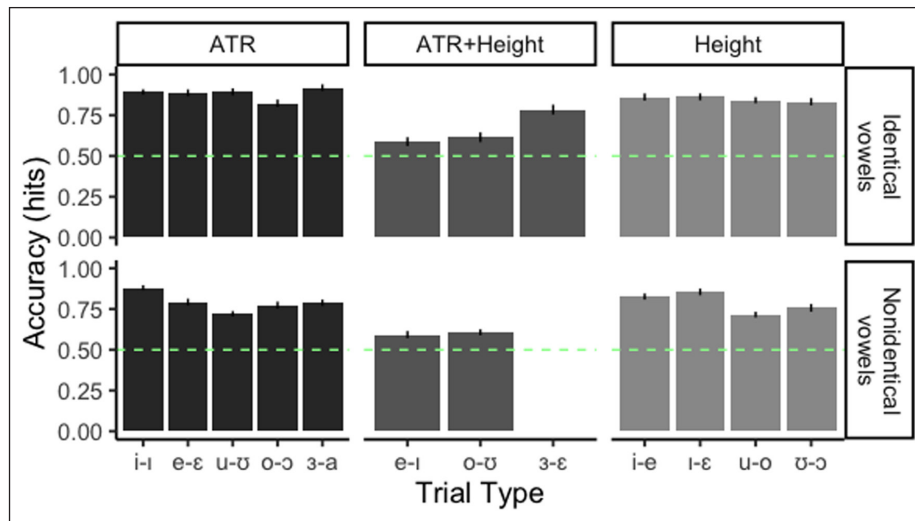


Figure 9: Experiment 2 accuracy values, with standard errors. Dashed line indicates chance performance (.50). Identical-vowel trials (e.g., gubu-gubu-gubu) showed high accuracy for ATR+Height contrasts, with lower overall accuracy for ATR+Height-differing items. Non-identical vowels (e.g., gubu-gubu-gubu; Sets B and C) showed similar patterns, again with Height (harmonic) and ATR (disharmonic) accuracy higher than ATR + Height (disharmonic).

We conducted an ANOVA on d' values with two within-subjects predictors: Vowel Identity within the differing word (identical vowels, nonidentical vowels) and vowel Contrast Type (ATR, Height, or ATR + Height). Trials with ɜ-ε were dropped as these did not have a counterpart in the differing-vowel case. Recall that Bonferroni correction for three comparisons yields a p -value of .0083. There was an effect of Contrast Type ($F(2,76) = 102.9, p < .0001$), such that ATR pairs showed marginally higher d' than Height pairs ($t(38) = 2.72, p = .01$; marginal after Bonferroni correction), and both exceeded the ATR + Height pairs (vs. ATR: $t(38) = 11.58, p < .0001$; vs. Height: $t(38) = 10.08, p < .0001$). There was also an effect of Vowel Identity ($F(1,38) = 84.7, p < .0001$), such that d' values were greater when the test word contained two identical vowels than when the test word contained two different vowels. Finally, the main effects were qualified by a two-way interaction ($F(2,76) = 15.04, p < .0001$). To understand the nature of the interaction, we compared the contrast types for each Vowel Identity. For identical vowels, ATR changes were better identified than Height changes ($t(38) = 3.16, p = .003$), and both were better identified than ATR + Height changes; (vs. Height: $t(38) = 8.82, p < .0001$; vs. ATR: $t(38) = 10.81, p < .0001$). The results for nonidentical vowels inform the Harmony Hypothesis, that harmonic nonidentical vowel stimuli would be easier to process than disharmonic stimuli. While both Height-changing stimuli (e.g., gubo; $t(38) = 8.64, p < .0001$) and ATR-changing stimuli (e.g., gubu; $t(38) = 8.88, p < .0001$) were better identified than ATR + Height-changing stimuli (e.g., gubo), the Height stimuli were no better detected than the (disharmonic) ATR stimuli

($t(38) = 0.06, p = .95$). While this might seem to suggest that disharmonic stimuli are not overall at a perceptual disadvantage, it is worth noting that the disharmonic ATR stimuli show a larger drop from the identical-vowel stimuli than the Height stimuli do. A follow-up ANOVA compared the size of this drop by testing the interaction of Contrast Type (ATR, Height; leaving out ATR + Height) and Vowel Identity. The interaction term was significant ($F(1,38) = 6.82, p = .01$), suggesting that the drop in performance in two-vowel stimuli was larger for ATR stimuli than for Height stimuli, which is consistent with a relative disadvantage of disharmony.⁸ This may be seen as partly consistent with the Harmony Hypothesis. Despite all of these differences, and low d' values in some cases, all six values shown in **Figure 10** exceeded chance ($t(38) \geq 4.28, p \leq .0001$), meaning that all vowel classes show better discriminability than chance.

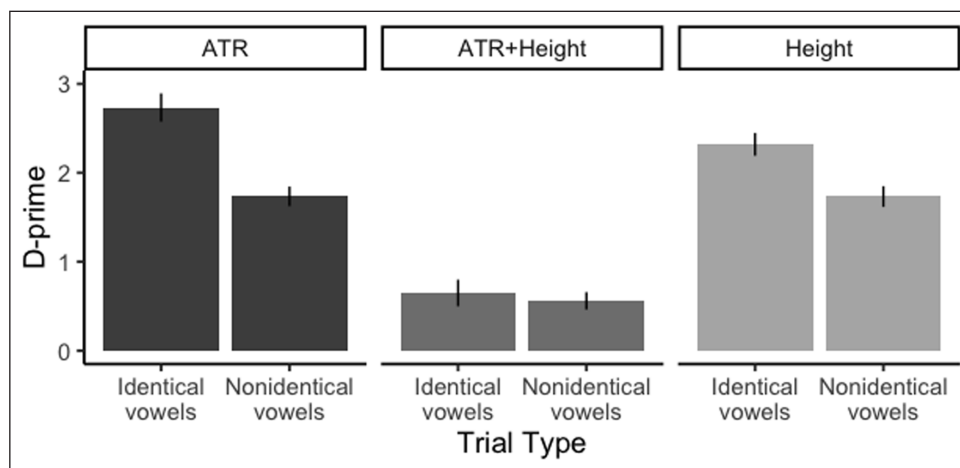


Figure 10: d' values for identical-vowel and non-identical vowel trials, collapsed across vowel pairs. The 3-ε trials have been omitted as they were only present in the identical-vowel condition.

As in Experiment 1, we continue to find that ATR + Height-differing vowels are difficult to distinguish perceptually. Among the ATR + Height contrasts, 3-ε appeared to be distinguished more accurately. To verify this statistically, we computed individual per-vowel d' scores for identical-vowel ATR + Height stimuli. The 3-ε contrast outperformed the other two ATR + Height

⁸ A reviewer asks if there are differences in performance on the disharmonic sequences depending on the order of the vowels, due to the regressive nature of vowel harmony. Poorer perception might be expected if the vowel order is -ATR + ATR (illicit) compared to +ATR -ATR (licit). This is indeed the case if the standard X is +ATR + ATR, but not when it is -ATR -ATR, suggesting that speakers may be perceptually compensating for regressive harmony. However, the order +ATR -ATR was harder to differentiate when X was -ATR -ATR than when it was +ATR + ATR, which does not appear to be due to directionality. We do not include directionality results for reasons of space, and due to the difficulty of making clear hypotheses and predictions. For example, if participants misperceive [ɪ] as [e] and [ʊ] as [o], or vice versa, there are a range of possible perceptual directionality outcomes.

contrasts ($F(2,76) = 19.01, p < .0001$; individually, it outscored $e-i, t(38) = 5.26, p < .0001$; and $o-u, t(38) = 5.50, p < .0001$).⁹

As in Experiment 1, to provide further support for the Acoustic Hypothesis, we assessed the significance of the correlation between the vowel pairs' Euclidean distances and their accuracy (identical trials only). As in Experiment 1, there was a positive linear relationship between the two measures ($t(10) = 3.60, r = .751, p = .005$), as shown in **Figure 11**.

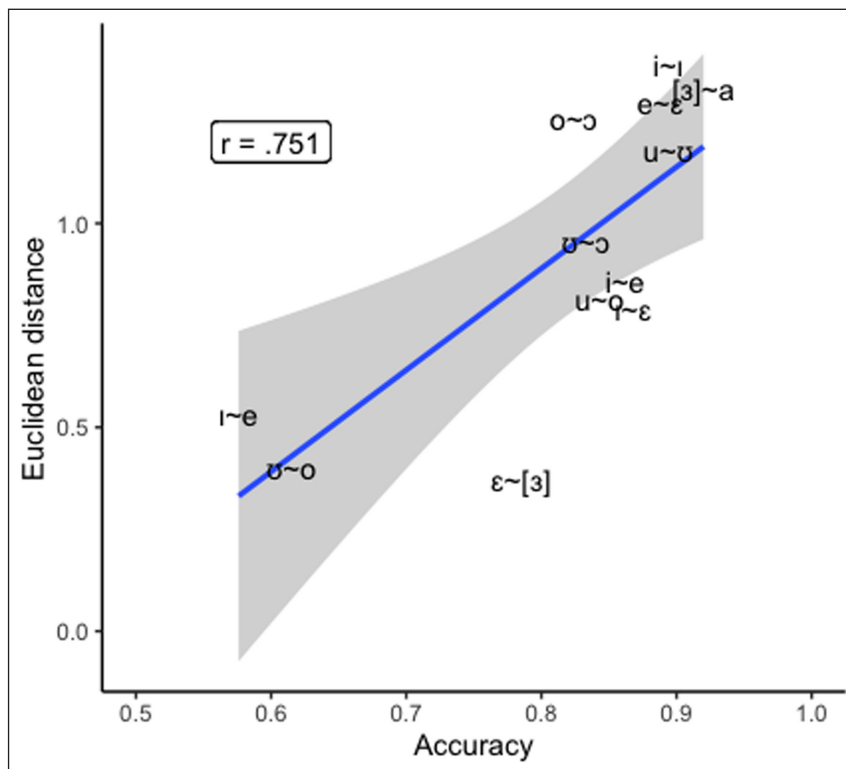


Figure 11: Experiment 2, accuracy x acoustic Euclidean distance scores.

As in Experiment 1, we wanted to know how well the only allophonic ATR contrast, $[3]-a$, fared. Did it show poor discrimination (in keeping with the Phonological hypothesis) or good discrimination (in keeping with the Acoustic hypothesis)? To test this, we computed per-vowel d'

⁹ While accuracy values appear different for ATR + Height between the two studies, this is primarily because Experiment 1 just shows accuracy on the “different” trials, while Experiment 2 shows accuracy on all trials. The d' values for the two studies are similar. While d' values are somewhat lower for Height-only and ATR-only in Experiment 2, given the higher memory load in Experiment 2 (having to remember four syllables before getting to X (A+B = 4 syllables), over a longer ISI, vs. Experiment 1's one syllable and shorter ISI), as well as the change in language, we are reluctant to ascribe too much practical significance to this difference.

scores for identical-vowel stimuli differing in ATR. A t-test comparing the [ɜ]~a pair ($d' = 3.85$, $SE = 1.38$) to the combination of the other four pairs ($d' = 2.73$, $SE = 1.07$) showed that [ɜ]~a had greater d' ($t(38) = 4.95$, $p < .0001$), suggesting that its allophonic status did not make it more difficult to discriminate. This is consistent with the Acoustic hypothesis—and our findings in Experiment 1—but not the Phonological hypothesis.

	Vowel pairs	Acoustic	Phonological	d'	Accuracy
Identical vowels					
Non-low ATR	i ~ ɪ e ~ ε u ~ ʊ o ~ ɔ	✓	✓	2.73 (1.07)	87.3%
Height	i ~ e u ~ o ɪ ~ ε ʊ ~ ɔ ε ~ a	✓	✓	2.35 (0.85)	84.9%
ATR+Height	ɪ ~ e ʊ ~ o ε ~ ɜ	✗	✓	0.67 (0.98)	60.0%
Low ATR	a ~ [ɜ]	✓	✗	3.85 (1.38)	92.0%

Table 7: Experiment 2, Acoustic hypothesis vs. Phonological hypothesis predictions compared to results.

Non-identical vowels		Harmony hypothesis	d'	Accuracy
Non-low ATR	i ~ ɪ e ~ ε u ~ ʊ o ~ ɔ	✗	1.74 (0.68)	78.9%
Height	i ~ e u ~ o ɪ ~ ε ʊ ~ ɔ ε ~ a	✓	1.74 (0.75)	78.7%
ATR+Height	ɪ ~ e ʊ ~ o	✗	0.56 (0.61)	60.1%
Low ATR	a ~ [ɜ]	✗	2.01 (1.20)	78.8%

Table 8: Experiment 2, Harmony hypothesis predictions compared to results.

3.4 Discussion

As with Experiment 1, the acoustically similar phonemic pairs ɪ ~ e and ʊ ~ o had lower rates of accurate perception compared to ATR and Height pairs. This was true for both Identical and Non-identical vowel trials. Further, the [ɜ] ~ a ATR pair was again perceived well, even though this is an allophonic contrast. Like Experiment 1, these results are consistent with the Acoustic hypothesis whereby similar acoustic vowels exhibit lower perceptual accuracy. This is especially striking for the [o] ~ [ʊ] contrast, as these vowels are acoustically similar but do not show as great an overlap in the acoustic vowel space. This is the case even if the vowels are produced in a different articulatory manner and even if they are distinguished by two phonological features. While degree of phonological similarity was not part of the hypotheses

tested, being distinguished by two phonological features contributes to a low phonological similarity rating, which is not congruent with the perceptual results for $i \sim e$ and $u \sim o$ pairs. As for the Harmony hypothesis, it was not the case that bisyllabic harmonic Height tokens were better perceived than bisyllabic disharmonic ATR tokens, as predicted. Performance on the non-identical tokens were similar for all ATR pairs and Height pairs, but the ATR + Height pairs again showed the weakest performance. However, there was a stronger drop in performance for the ATR pairs than the Height pairs, suggesting that disharmony did negatively impact perception for ATR distinctions.

However, unlike in Experiment 1, the $ɜ \sim \varepsilon$ pair showed much better perceptual accuracy than the other ATR + Height pairs in Experiment 2. We do not believe this is due to the vowels being significantly different between the two experiments, as they were pronounced similarly in both with a large overlap in F1/F2. This difference may be due to two aspects of Experiment 2. First, [ɜ] does not occur in monosyllables in Akan words, but [ɜ] can occur in initial position in bisyllables (albeit followed by +ATR vowels other than [ɜ]), so the bisyllabic environment aids perception. However, it is also possible that the spliced bisyllabic $gɜbɜ/tɜkɜ$ tokens may not have sounded as natural as other tokens, so when $gɜbɜ$ was compared to $gebe$, it may have been easier to detect different vowels. In Experiment 1, we cut out the first syllable [kɜ] from a bisyllabic word, but in this experiment, we had to splice two halves of two words together. Regardless, it is clear that overall, the ATR + Height pairs are more difficult to perceive than ATR differences or Height differences.

4. General Discussion

Overall, both experiments provided evidence inconsistent with the Phonological hypothesis and consistent with the Acoustic hypothesis. Both confirmed that the acoustically similar vowels [i] / [e] and [u] / [o] are poorly discriminated by Akan listeners, even if they are differentiated by two distinctive phonological features. Conversely, both experiments also confirmed that the acoustically distinct vowels [ɜ] / [a] are well discriminated, despite being in an allophonic rather than contrastive relationship.

Recall that there were several differences between the two experiments, which make their relative agreement all the more striking. They varied in number of syllables, monosyllabic (Experiment 1) vs. bisyllabic (Experiment 2); interstimulus interval, with 500 ms (Experiment 1), vs. a more “phonological” 700 ms in Experiment 2; words in the phonologically-similar language Gua in Experiment 1, vs. nonce words in Akan in Experiment 2; an AX design that allowed acoustic discrimination in Experiment 1 vs. an ABX design that required a more phonological level of comparison in Experiment 2. Despite these differences, acoustically similar (but phonologically distinct) pairs in both experiments were significantly harder to distinguish than the other pairs,

and the allophonic (but acoustically distinct) pair were at least as easy to distinguish as the phonemic counterparts.

The Ozburn et al. (2022) study on Dàgáárè also found that the ATR + Height pair [ɪ ~ e] was poorly discriminated, at least compared to [e ~ ε], in keeping with the overlap in the acoustic space of that pair, and similar to our results. However, discrimination was not significantly different for [i ~ ɪ] compared to either of the other two pairs. They also found that Dàgáárè listeners fared worse with bisyllable disharmonic stimuli than with monosyllables, but did not find that harmonic stimuli improved perception compared to monosyllables. The experiments employed an ABX design and had monosyllable and bisyllable stimuli in the same experiment, allowing for direct comparison between the two kinds of stimuli. Our Experiment 2 results did show lower accuracy for disharmonic bisyllables than harmonic identical forms, but not for harmonic non-identical forms. However, we are not able to make the same comparison between monosyllables and bisyllables due to the multiple different parameters outlined above.

5. Implications for the phonology of Akan

Akan appears to present a case of near-merger – the situation whereby speakers differentiate sounds in production, but report that they are ‘the same’ in perception tests (Labov et al., 1991). Indeed, phoneticians in earlier work often described the situation in Akan as a merger, particularly for the vowels [ɪ] ~ [e] due to the close F1/F2 acoustic measurements. However, X-ray and MRI data from the 1970s and 1990s, albeit based on a few speakers and different generations of speakers, confirm distinct tongue root productions for Akan vowels despite the acoustic similarity; recent ultrasound data (Kirkham & Nance, 2017) confirm this for most vowel pairs (i/ɪ u/ʊ o/ɔ o/ʊ), but were not able to confirm it for the e/ε or e/ɪ distinctions due to speakers adopting different articulatory strategies.¹⁰

This apparent near-merger raises the question of learnability. If the vowels are so hard to perceive as different, how do Akan speakers learn to produce them differently? A recent study of first language acquisition by nine Akan-speaking children (Asante-Twi dialect) aged three to five years (Amoako, 2020) shows that these particular vowel distinctions may not necessarily be the ones posing problems. Her results indicate that all +ATR vowels have a lower production

¹⁰ Kirkham and Nance (2017) describe the situation as follows: “An examination of individual speaker data shows two divergent patterns amongst the Twi speakers: GF01 and GF02 produce /e/ slightly more advanced than /ε/, whereas GM01 and GM02 produce /e/ slightly more retracted than /ε/. These patterns remain the same whether we use values extracted from the vowel midpoint or values extracted 80% into the vowel. There is also no clear correspondence between the use of these two articulatory strategies and any patterns in tongue height, which suggests that the Twi speakers may be particularly variable in the articulation of the /e ε/ contrast, despite greater acoustic consistency among speakers.”

accuracy than –ATR vowels across the age groups and children. In particular, [ɜ] (she transcribes this as [æ]) has the lowest accuracy rate of all vowels and is still not fully acquired by age five. It is frequently substituted with either [ɛ], which fits with acoustic similarity, or with [a], its harmonic counterpart, which may reflect lack of vowel harmony application for this vowel (as noted in Section 1.4, the contexts where /a/ is harmonized to [ɜ] are restricted). This shows that the acquisition of the allophonic vowel is delayed compared to the phonemic contrasts, which may be a reflection of its more limited contextual environment and possible ongoing merger with [ɛ] from /e/ (Kpogo, 2022), rather than difficulty in perception. Interestingly, the other [+ATR] vowels are commonly substituted with their [–ATR] counterparts, not the acoustically similar but height-differing [–ATR] counterparts. For example, the word èdzá ‘fire’ is produced as [èdzá] rather than [ídzá]. Although there are some sporadic cases of /o/ being pronounced as [u], or [e] as [ɪ], given what we know about adult perception, we might have expected a far stronger tendency for these vowels to be merged in children’s productions. Amoako (2020, p. 130) notes that the children have acquired vowel harmony, and vowel substitutions tend to respect it “even when there were errors in their production.” This provides some supporting evidence that vowel harmony is learned at a young age and appears to reinforce the distinction between vowels by keeping them separated within words. We hypothesize that vowel harmony may aid speakers in acquiring phonemic contrasts between acoustically similar vowels, such as ATR+Height pairs /ɪ e/ and /u o/. Nevertheless, if the author, as an adult Akan speaker, transcribed the vowels based on her own perception, substitution errors between [ɛ] and [ɪ] may not have been as readily perceived as between other vowels. Future acquisition work may benefit from using acoustic measurements.

6. Conclusion

This is a large-scale study of the perception of Advanced Tongue Root vowel contrasts by speakers of a language with ATR vowels and vowel harmony. Two experiments were conducted to test perception of ATR contrasts in Akan: An AX task and an ABX task. The AX task (Experiment 1) tested perception of monosyllables drawn from a related language (Gua), and participants were told it was a different language than Akan. The ABX task (Experiment 2) tested perception of bisyllables in the Akan language and participants were told it consisted of nonce Akan words. The results show that participants were able to accurately perceive all vowel pairs that differed based on only ATR in both experiments, even the pair a ~ ɜ that was not phonologically contrastive. However, acoustically similar vowels ɪ ~ e, u ~ o, and ɜ ~ ɛ in Experiment 1, were poorly perceived. This is so despite these vowel pairs being phonologically contrastive and differing for two phonological features, vowel height and ATR. Similar results obtain for ɪ ~ e, u ~ o in Experiment 2.

The results show that acoustic similarity in Akan vowels drives perception, rather than phonological similarity or phonemic status. It provides supporting perceptual evidence that some vowel pairs are so acoustically similar as to be practically merged, despite being produced differently based on tongue root and larynx position. This constitutes a case of near merger where production is different but perception shows a merger. The role of ATR vowel harmony is important in Akan for maintaining the distinction between similar vowels. These vowels commonly occur in words with other vowels that match them for ATR, respecting vowel harmony. There are thus few instances where distinguishing, say, [e] from [i], must occur in the absence of other vowel cues. Nevertheless, we know from historical evidence in other languages that vowels may exhibit merger despite the presence of vowel harmony, reducing nine vowel ATR systems to seven (Elugbe, 1989). It remains to be seen whether Akan will move definitively in this direction.

Acknowledgements

We are grateful to Associate Editor James Kirby and two anonymous reviewers for helpful comments and critiques on this paper. We thank audiences at the UC San Diego Center for Research in Language, the *Southern California Workshop on Phonology* at California State University Long Beach, and the *Annual Conference on African Linguistics 50* at the University of British Columbia for feedback. We are grateful to Aaron Braver for useful tips on experiment design, and Monica Apenteng Obiri-Yeboah for nonce word assessment and recording help. We thank the Department of Applied Linguistics and Prof. Avea Nsoh of the Colleges of Languages Education at the University of Education, Winneba in Ghana, who provided logistical support for this research. We thank all the Akan speakers who participated in the study. This research was supported by a grant from the UC San Diego Academic Senate.

Competing interests

The authors have no competing interests to declare.

References

- Abakah, E. N. (2016). Hypotheses on the diachronic development of the Akan language group. *Journal of Universal Language*, 17(1), 1–51. DOI: <https://doi.org/10.22425/jul.2016.17.1.1>
- Akan Language Committee. (1995). *Akan Orthography*. Bureau of Ghana Languages, Accra
- Allen, B., Pulleyblank, D., & Ajíbóyè, O. (2013). Articulatory mapping of Yoruba vowels: An ultrasound study. *Phonology*, 30(2), 183–210. DOI: <https://doi.org/10.1017/S0952675713000110>
- Amoako, W. K. (2020). *Assessing phonological development among Akan-speaking children*. MA Thesis. Vancouver: University of British Columbia. [<https://open.library.ubc.ca/soa/cIRcle/collections/ubctheses/24/items/1.0390995>].
- Appiah-Padi, R. (1994). Acoustic correlates of advanced tongue root. MSc. thesis, University of Alberta, Edmonton. DOI: <https://doi.org/10.7939/R3F766H2R>
- Babel, M., & Johnson, K. (2010). Accessing psycho-acoustic perception with speech sounds. *Laboratory Phonology*, 1(1), 179–205. DOI: <https://doi.org/10.1515/labphon.2010.009>
- Bakovic, E. (2000). *Harmony, dominance and control*. Doctoral dissertation, Rutgers University. DOI: <https://doi.org/10.7282/T3TQ60BJ>
- Berry, J. (1957). Vowel harmony in Twi. *Bulletin of the School of Oriental and African Studies*, 19(1), 124–130. DOI: <https://doi.org/10.1017/S0041977X0011924X>
- Best, C. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 171–204). Baltimore: York Press.
- Billington, R. (2014). ‘Advanced Tongue Root’ in Lopit: Acoustic and ultrasound evidence. In J. Hay & E. Parnell (Eds.), *Proceedings of the 15th Australasian International Speech Science and*

Technology Conference (pp. 119–122). Christchurch, New Zealand: Australasian Speech Science and Technology Association.

Billington, R. (2017). *The Phonetics and Phonology of the Lopit language*. Doctoral Dissertation, U. of Melbourne.

Boomershine, A., Hall, K. C., Hume, E., & Johnson, K. (2008). The Influence of allophony vs. contrast on perception: the case of Spanish and English. In P. Avery, E. Dresher & K. Rice (Eds.), *Phonological Contrast* (pp. 143–172). New York: Mouton de Gruyter. DOI: <https://doi.org/10.1515/9783110208603.2.145>

Casali, R. F. (2003). [ATR] value asymmetries and underlying vowel inventory structure in Niger-Congo and Nilo-Saharan. *Linguistic Typology*, 7, 307–382. DOI: <https://doi.org/10.1515/lity.2003.018>

Casali, R. F. (2008). ATR Harmony in African Languages. *Language and Linguistics Compass*, 2, 496–549. DOI: <https://doi.org/10.1111/j.1749-818X.2008.00064.x>

Casali, R. F. (2012). [+ATR] dominance in Akan. *Journal of West African Languages*, 39(1), 33–59.

Casali, R. F. (2017). High-vowel patterning as an early diagnostic of vowel-inventory type. *Journal of West African Languages*, 44(1), 79–112.

Christaller, J. G. (1875). *A grammar of the Asante and Fante language called Tshi <Chwee, Twi> based on the Akuapem dialect with reference to the other (Akan and Fante) dialects*. Gold Coast: Basel German Evangelical Mission. (Reprinted 1964, Ridgewood, New Jersey: Gregg Press).

Clements, G. N. (1981). Akan vowel harmony: A nonlinear analysis. *Harvard Studies in Phonology*, 2, 108–177. Bloomington: Indiana University Linguistics Club.

Clements, G. N. (1984). Vowel harmony in Akan: a consideration of Stewart's word structure conditions. *Studies in African Linguistics*, 15(3), 321–337. DOI: <https://doi.org/10.32473/sal.v15i3.107510>

Clements, G. N. (1985). Akan vowel harmony: a nonlinear analysis. In D. Goyvaerts (Ed.), *African Linguistics: Essays in Memory of M.W.K. Semikenke* (pp. 55–98). Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/ssls.6.06cle>

Clements, G. N. (2007). Africa as a phonological area. In B. Heine & D. Nurse (Eds.), *A Linguistic Geography of Africa*. (pp. 36–85). Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511486272.004>

Denning, K. (1989). *The Diachronic development of phonological voice quality, with special reference to Dinka and the other Nilotic languages*. Doctoral dissertation, Stanford University. <https://www.proquest.com/dissertations-theses/diachronic-development-phonological-voice-quality/docview/303795858/se-2?accountid=14524>

Dolphyne, F. A. (1988). *The Akan (Twi-Fante) Language: Its Sound Systems and Tonal Structure*. Ghana Universities Press, Accra.

Dubno, J., & Levitt, H. (1981). Predicting consonant confusions from acoustic analysis. *The Journal of the Acoustical Society of America*, 69(1), 249–261. DOI: <https://doi.org/10.1121/1.385345>

- Edmondson, J. A., & Esling, J. H. (2006). The valves of the throat and their functioning in tone, vocal register, and stress: laryngoscopic case studies. *Phonology*, 23(2), 157–191. DOI: <https://doi.org/10.1017/S095267570600087X>
- Edmondson, J. A., Padayodi, C. M., Hassan, Z. M., & Esling, J. H. (2007). The laryngeal articulator: source and resonator. *International Conference of Phonetic Sciences*, 16, 2065–2068.
- Elugbe, B. (1989). *Comparative Edoid: Phonology and Lexicon*. Delta Series No. 6. Port Harcourt: University of Port Harcourt Press.
- Escudero, P. (2009). Linguistics perception of “similar” L2 sounds. In P. Boersma & S. Hamann (Eds.), *Phonology in Perception* (pp. 151–190). Mouton de Gruyter. DOI: <https://doi.org/10.1515/9783110219234.151>
- Frisch, S. A., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity Avoidance and the OCP. *Natural Language & Linguistic Theory*, 22, 179–228. DOI: <https://doi.org/10.1023/B:NALA.0000005557.78535.3c>
- Fulop, S., Kari, E., & Ladefoged, P. (1998). An acoustic study of tongue root contrast in Degema vowels. *Phonetica*, 55, 80–98. DOI: <https://doi.org/10.1159/000028425>
- Gick, B., Pulleyblank, D., Campbell, F., & Mutaka, P. (2006). Low vowels and transparency in Kinande vowel harmony. *Phonology*, 23(1), 1–20. DOI: <https://doi.org/10.1017/S0952675706000741>
- Guion, S., Post, M., & Payne, D. (2004). Phonetic correlates of tongue root vowel contrasts in Maa. *Journal of Phonetics*, 32, 517–542. DOI: <https://doi.org/10.1016/j.wocn.2004.04.002>
- Hall, K. C., Letawsky, V., Turner, A., Allen, C., & McMullin, K. (2015). Effects of predictability of distribution on within-language perception. In S. Vinerte (Ed.), *Proceedings of the 2015 Canadian Linguistics Association*.
- Halle, M., & Stevens, K. N. (1969). On the feature ‘Advanced Tongue Root’. *MIT Research Laboratory of Electronics Quarterly Progress Report*, 94, 209–215. (Reprinted in M. Halle (2003). *From Memory to Speech and Back Papers on Phonetics and Phonology 1954–2002* (pp. 37–44). Boston: de Gruyter. DOI: <https://doi.org/10.1515/9783110871258.37>
- Hess, S. 1992. Assimilatory effects in a vowel harmony system: an acoustic analysis of advanced tongue root in Akan. *Journal of Phonetics*, 20, 475–492. DOI: [https://doi.org/10.1016/S0095-4470\(19\)30651-5](https://doi.org/10.1016/S0095-4470(19)30651-5)
- Hudu, F. (2014). [ATR] feature involves a distinct tongue root articulation: Evidence from ultrasound imaging. *Lingua*, 143(2), 36–51. DOI: <https://doi.org/10.1016/j.lingua.2013.12.009>
- Jacobsen, L. (1978). *DhoLuo vowel harmony: a phonetic investigation*. Doctoral Dissertation, UCLA. Published as *UCLA Working Papers in Phonetics* 43. <https://escholarship.org/uc/item/4537z5qp>
- Kang, S. (2018). *Tongue Root Harmony and Vowel Contrast in Northeast Asian Languages*. (Turcologica 112). Wiesbaden, Germany: Harrassowitz.
- Kang, S., & Ko, H. (2012). In search of the acoustic correlates of tongue root retraction in three Altaic languages: Western Buriat, Tsongol Buriat, and Ewen. *Altai Hakpo*, 22, 179–203. DOI: <https://doi.org/10.15816/ask.2012..22.009>

- Kingston, J., Macmillan, N. A., Walsh Dickey, L., Thorburn, R., & Bartels, C. (1997). Integrality in the perception of tongue root position and voice quality in vowels. *The Journal of the Acoustical Society of America*, 101, 1696. DOI: <https://doi.org/10.1121/1.418179>
- Kirkham, S., & Nance, C. (2017). An acoustic-articulatory study of bilingual vowel production: Advanced tongue root vowels in Twi and tense/lax vowels in Ghanaian English. *Journal of Phonetics*, 62, 65–81. DOI: <https://doi.org/10.1016/j.wocn.2017.03.004>
- Koffi, E. (2018). The acoustic vowel space of Anyi in light of the cardinal vowel system and the Dispersion Focalization Theory. In J. Kandybowicz, T. Major, H. Torrence & P. T. Duncan (Eds.), *African linguistics on the prairie: Selected papers from the 45th Annual Conference on African Linguistics* (pp. 191–204). Berlin: Language Science Press. DOI: <https://doi.org/10.5281/zenodo.1251729>
- Kpogo, F. (2022). A vowel shift in the Twi harmony system: a case of Urban Twi speakers. Poster presented at the 96th Annual Meeting of the Linguistic Society of America, Washington, D.C.
- Kügler, F. (2015). Phonological phrasing and ATR vowel harmony in Akan. *Phonology*, 32, 177–204. DOI: <https://doi.org/10.1017/S0952675715000081>
- Ladefoged, P. (1964). *A Phonetic Study of West African Languages: An Auditory-Instrumental Survey*. (West African Language Monographs, I.) Cambridge: Cambridge University Press.
- Li, B. (1996). *Tungusic vowel harmony: Description and analysis* (HIL Dissertations 18). Dordrecht: Holland Institute of Generative Linguistics.
- Lindau, M. (1975). [Features] for vowels. *UCLA Working Papers in Phonetics* 30. <https://escholarship.org/uc/item/7gv6z0vq>
- Lindau, M. (1978). Vowel features. *Language*, 54, 541–563. DOI: <https://doi.org/10.1353/lan.1978.0066>
- Lindau, M. (1979). The feature expanded. *Journal of Phonetics*, 7, 163–176. DOI: [https://doi.org/10.1016/S0095-4470\(19\)31047-2](https://doi.org/10.1016/S0095-4470(19)31047-2)
- Lindau, M. (1987). Tongue mechanisms in Akan and Luo. *UCLA Working Papers in Phonetics*, 68, 46–57. <https://escholarship.org/uc/item/56t0x9z3>
- Local, J., & Lodge, K. (2004). Some auditory and acoustic observations on the phonetics of [ATR] harmony in a speaker of a dialect of Kalenjin. *Journal of the International Phonetic Association*, 34(1), 1–16. DOI: <https://doi.org/10.1017/S0025100304001513>
- Myers, S., Obiri-Yeboah, M., de Jong, K., & Berkson, K. (to appear). Tongue root contrasts in Gua: evidence from articulatory imaging. *Indiana Working Papers on Speech Sound Articulation*.
- Obiri-Yeboah, M. (2021). Phonetics and Phonology of Gua. Doctoral Dissertation, University of California San Diego.
- Omamor, A. P. (1988). Okpè and Uvwie: a case of vowel harmony galore. *Journal of West African Languages*, 18(1), 47–64.
- Ozburn, A., Giovio Canavesi, G. F., & Akinbo, S. (2022). Perception of ATR in Dàgáàrè [dàgáàrì]. Ms. University of Toronto & University of Minnesota.

- Owusu, S. (2014). On exceptions to Akan vowel harmony. *International Journal of Scientific Research and Innovative Technology*, 1(5), 45–52.
- Peperkamp, S., Pettinato, M., & Dupoux, E. (2003). Allophonic Variation and the Acquisition of Phoneme Categories. In: B. Beachley, A. Brown & F. Conlin (Eds.), *Proceedings of the 27th Annual Boston University Conference on Language Development*, 2, 650–661. Somerville, MA: Cascadilla Press.
- Piccinini, P., & Arvaniti, A. (2019). Dominance, mode, and individual variation in bilingual speech production and perception. *Linguistic Approaches to Bilingualism*, 9(4/5), 628–658. DOI: <https://doi.org/10.1075/lab.17027.pic>
- Pike, K. (1947). *Phonemics: a technique for reducing languages to writing*. Ann Arbor: University of Michigan Press.
- Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics*, 132, 253–260. DOI: <https://doi.org/10.3758/BF03214136>
- Pulleyblank, D. (2002). Harmony drivers: no disagreement allowed. *Berkeley Linguistics Society*, 28, 249–267. DOI: <https://doi.org/10.3765/bls.v28i1.3841>
- Quinn-Wriedt, L. (2013). *Vowel Harmony in Maasai*. Doctoral Dissertation, University of Iowa. DOI: <https://doi.org/10.17077/etd.mi47g4f5>
- R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rolle, N., Lionnet, F., & Faytak, M. (2020). Areal patterns in the vowel systems of the Macro-Sudan Belt. *Linguistic Typology*, 24(1), 113–179. DOI: <https://doi.org/10.1515/lingty-2019-0028>
- Schacter, P., & Fromkin, V. (1968). A Phonology of Akan: Akuapem, Asante, Fante. *UCLA Working Papers in Phonetics* 9. <https://escholarship.org/uc/item/3fn134hv>
- Schacter, P. (1969). Natural assimilation rules in Akan. *International Journal of American Linguistics*, 35(4), 342–355. DOI: <https://doi.org/10.1086/465079>
- Starwalt, C. (2008). *The Acoustic correlates of ATR harmony in seven- and nine-vowel African languages: a phonetic inquiry into phonological structure*. Doctoral Dissertation, University of Texas, Arlington. <http://hdl.handle.net/10106/1015>
- Stewart, J. M. (1967). Tongue root position in Akan vowel harmony. *Phonetica*, 16, 185–204. DOI: <https://doi.org/10.1159/000258568>
- Stewart, J. M. (1971). Niger-Congo, Kwa. In T. A. Sebeok (Ed.), *Current trends in linguistics 7: Linguistics in sub-Saharan Africa* (pp. 179–212). The Hague & Paris: Mouton & Co. DOI: <https://doi.org/10.1515/9783111562520-008>
- Stewart, J. M. (1983). Akan vowel harmony: the word structure conditions and the floating vowels. *Studies in African Linguistics*, 14, 111–139. DOI: <https://doi.org/10.32473/sal.v14i2.107530>
- Tiede, M. K. (1996). An MRI-based study of pharyngeal volume contrasts in Akan and English. *Journal of Phonetics*, 24, 399–421. DOI: <https://doi.org/10.1006/jpho.1996.0022>
- Tyler, M., Best, C., Faber, A., & Levitt, A. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica*, 71(1), 4–21. DOI: <https://doi.org/10.1159/000356237>

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, 37(1), 35–44. DOI: <https://doi.org/10.3758/BF03207136>

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>. DOI: <https://doi.org/10.1007/978-3-319-24277-4>

Williamson, K. (1973). Some reduced vowel harmony systems. *Research Notes, Department of Linguistics and African Languages* 6, Ibadan:145–169.

