Laboratory Phonology
Journal of the Association for Laboratory Phonology

Ozburn, A., Giovio Canavesi, G. F., & Akinbo, S. K. (2023). Perception of ATR in Dàgáárè. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 14(1), pp. 1–42. DOI: https://doi.org/10.16995/labphon.8919

Open Library of Humanities

Perception of ATR in Dàgáárè

Avery Ozburn*, Department of Language Studies, University of Toronto Mississauga, Mississauga, ON, Canada, avery.ozburn@utoronto.ca

Gianna Francesca Giovio Canavesi, Department of Language Studies, University of Toronto Mississauga, Mississauga, ON, Canada, gianna.gioviocanavesi@mail.utoronto.ca

Samuel Kayode Akinbo, Department of English, Linguistics, & Writing Studies, University of Minnesota, Minneapolis, MN, USA, samuel.akinbo@utoronto.ca

*****Corresponding author.

This paper reports on two related perception studies about the property Advanced Tongue Root (ATR) in Dàgáárè (Mabia; Ghana). We examine how well native speakers are able to distinguish ATR contrasts as well as the effects of harmony and disharmony on perception, thereby testing hypotheses that have been made in the literature about the perceptual motivations of harmony systems. We find that, as expected, ATR mid vowels and Retracted Tongue Root (RTR) high vowels are the hardest to distinguish in Dàgáárè, but contrary to expectations, harmony does not improve accuracy in discriminating ATR contrasts. Nonetheless, we find the accuracy on disharmonic disyllabic forms is significantly worse than the accuracy in monosyllabic forms, which may indicate that disharmony hurts perception. We examine the implications for our understanding of the motivations of harmony systems and discuss how this paper contributes to the very minimal existing literature on perception in African languages.

Laboratory Phonology: Journal of the Association for Laboratory Phonology is a peer-reviewed open access journal published by the Open Library of Humanities. © 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.
OPEN ACCESS

1. Introduction

One of the most well-studied phenomena in phonological theory is vowel harmony, given the difficulties that its long-distance nature poses to theoretical frameworks and to our understanding of language processing (Archangeli & Pulleyblank, 1994; Pulleyblank, 1996; Ringen & Vago, 1998; Mailhot & Reiss, 2007; Nevins, 2010). Within the vowel harmony literature, there has been much speculation on the motivations for the existence of harmony, with hypotheses that harmony promotes ease of articulation and/or perception. For example, a well-known dissertation on rounding harmony (Kaun, 1995) posits that asymmetries in the nature of triggers and targets in rounding harmony are motivated by relative ease and difficulty of perceiving rounding contrasts on low versus high vowels. Similarly, Ohala (1994) suggests that harmony may be phonologized due to listener hypocorrection, in which listeners attribute coarticulatory effects to the target.

Vowel harmony involving the feature Advance Tongue Root (ATR), or expanded pharynx, is one of the most widely studied phonological phenomena in African languages (Lindau, 1975, 1978; Edmondson, Padayodi, Hassan, & Esling, 2007; Esling, 2014). As the name suggests, ATR is commonly associated with advancement versus retraction of the tongue root, though articulatory studies have shown that the mechanisms vary both across and within languages (Ladefoged, 1968; Gick, Wilson, Koch, & Cook, 2004; Beltzung, Patin, & Clements, 2015). Acoustically, ATR is most commonly associated with differences in F1, with higher F1 in Retracted Tongue Root (RTR) vowels than in their ATR counterparts (Przezdziecki, 2005; Starwalt, 2008). Phonological systems of ATR harmony are commonly known as "cross-height", because these systems require agreement in ATR across vowel heights (Snider, 1984; Archangeli & Pulleyblank, 1994). Particularly interesting in such systems is that it is often reported (anecdotally) by non-native-speaker linguists that certain contrasts, particularly high RTR vowels and mid ATR vowels at the same backness, are impossible to distinguish beyond their phonological patterning.

Within the ATR harmony literature specifically, there are several claims that rest on the perceptibility or lack thereof of certain ATR contrasts (Aralova, 2015; Szeredi, 2016; Rose, 2018). For example, a typological survey paper by Rose (2018) suggests that the lack of perceptibility of the ATR contrast in high vowels may explain why languages with an ATR contrast in high vowels almost always have ATR harmony systems. However, very few systematic studies of ATR perception have been conducted: There is one by Fulop, Kari, and Ladefoged (1998), and one very recently published by Rose, Obiri-Yeboah, and Creel (2023). Indeed, as elaborated on in Section 5.4, perception in general (beyond ATR) is highly under-studied in African languages. As such, little is known about how native speakers of ATR languages perceive ATR contrasts. Instead, most of the claims in the literature about certain ATR contrasts being difficult

to distinguish come from the impressionistic observations of non-native-speaking linguists misidentifying sounds in languages that have ATR contrasts, rather than the perceptions of speakers of languages that natively have these contrasts. This means that confirming Rose's (2018) typologically-based hypothesis requires investigating two claims with limited evidence: That ATR contrasts are hard to perceive in high vowels and that ATR harmony helps with ATR perception.

Fulop et al. (1998) found that the ATR contrast in high vowels in artificially manipulated Degema stimuli was poorly differentiated compared to mid-vowel ATR contrasts. This study focused only on which of the first two formants was most reliable in ATR differentiation. More recently, Akan listeners were found to be highly accurate at differentiating ATR in all CV syllable types, including high vowels (Rose et al., 2023). This latter result challenges the idea that ATR contrasts are difficult to perceive in high vowels. However, these are the only two existing studies on ATR perception by native speakers, and so additional experiments on more languages will offer valuable insight into the property of ATR.

In this paper, we report on two experiments to investigate the perception of ATR in Dàgáárè (Mabia; Ghana). The Dàgáárè-speaking listeners in each of our experiments participated in an ABX task in which they were asked to distinguish between nonce words A and B, which differed in height and/or ATR on one or more vowels. We predicted that participants would have greater difficulty with trials with high vowels that differ in ATR (e.g., ki vs. kɪ is expected to be more difficult than ke vs. kɛ) and with disharmonic items (e.g., kɪke vs. keke). Predictions will be discussed in more detail after presenting the experimental designs (sections 3.2 and 4.2).

This paper is organized as follows. Section 2 provides background on the vowel system and harmony system of Dàgáárè and on perceptual accounts of harmony. Section 3 describes the research methodology, while Section 4 presents the results. Section 5 discusses the implications of the results to our understanding of ATR perception and harmony, while Section 6 concludes.

2. Background

2.1. Dàgáárè background

Dàgáárè is a Mabia (formerly known as Gur; Bodomo, 2017, 2020) language of the Niger-Congo phylum with about 1.5 million speakers in Ghana and Burkina Faso (Angsongna & Akinbo, 2022). Dàgáárè is predominantly spoken in northwestern Ghana and southern Burkina Faso. There are four broad dialects of Dàgáárè, namely Northern Dàgáárè [dàgàrà], Southern Dàgáárè [wá:lí], Western Dàgáárè [bìrìfɔ̀], and Central Dàgáárè [dàgáárì]. Previous impressionistic fieldwork research suggests that Southern, Western, and Northern dialects have nine vowels (Bodomo, 1997; Kuubezelle & Akanlig-Pare, 2017; Ali, Grimm, & Bodomo, 2021), but recent

acoustic and articulatory studies show that the Central dialect has a tenth vowel /ə/, which is the ATR counterpart of /a/ (Ozburn, Akinbo, Angsongna, Schellenberg, & Pulleyblank, 2018; Lloy, Akinbo, Angsonga, & Pulleyblank, 2019; Angsongna & Akinbo, 2022). The vowel inventory is shown in **Table 1**, with the ATR low vowel in parentheses since it may not exist in all dialects.

	ATR			**RTR**		
	Front	**Central**	**Back**	**Front**	**Central**	**Back**
High	i		u	ɪ		ʊ
Mid	e		o	ɛ		ɔ
Low		(ə)			a	

Table 1: Vowel inventory in Dàgáárè.

Dàgááre vowels can be grouped into ATR and RTR classes, as shown in **Table 1**. Like all Mabia languages, Dàgáárè has ATR harmony: Vowels within a word obligatorily agree in their ATR/RTR feature. Within words and in clitic groups in the language (i.e., with a single lexical item and any suffixes or clitics dependent on it), vowels must be harmonic. In this case, the root vowels trigger ATR harmony, targeting all vowels of affixes or clitics. Consequently, the vowels of affixes and clitics are either ATR or RTR depending on the root vowel. Some examples are shown in **Table 2**.

	[ATR]		**[RTR]**	
a.	tʰí:-rí	vomit- SG	sí:-rì	honey-SG
	tʰùù-rì	picking-NOM	sù:-rì	asking-NOM
	kpé-rè	slicking-NOM	wɛ́g-ɛ̀	log-SG
	kʰóg-ó	chair-SG	pɔ́g-ɔ́	woman-SG
b.	tì = bíí-rí	1PL.POSS = child-PL	tì = tìì-rí	1PL.POSS = tree-PL
	tì = lúg-ò	1PL.POSS = -pillar-PL	tì = pɔ́g-ɔ́	1PL.POSS = woman-PL
	tì = dè-rí	1PL.POSS = houses-PL	tì = wɛ́g-ɛ̀	1PL.POSS = log-PL

Table 2: ATR harmony in Dàgáárè (Angsongna & Akinbo, 2022).[1]

All vowels undergo ATR harmony in the Central dialect (Angsongna & Akinbo, 2022), but there is debate about the participation of the central vowel /a/ in ATR harmony in the nine-vowel dialects (Bodomo, 1997; Kuubezelle & Akanlig-Pare, 2017; Ali et al., 2021). As such, we exclude low vowels from the present study, so that this issue has no bearing on our experiments.

[1] Abbreviations: SG singular; NOM nominative; 1PL 1st person plural; POSS possessive; PL plural.

Disharmonic vowel combinations are impossible within a word, with the exception of compounds, which can involve ATR + RTR or RTR + ATR combinations. Consider the examples of ATR + RTR compounds below in **Table 3**, representing examples from all varieties of the language. Given that such disharmonic compound words are possible, we in fact expect speakers to be able to adequately distinguish disharmonic sequences, as they are exposed to them in the language.

	[ATR] root		**[RTR] root**			**Compound**	
a.	bí-é	child-SG	dɔ́ɔ́	man.SG	>	bì-dɔ́ɔ́	boy (lit. child-man)
	zû	head.SG	kɔ́mɔ́	hair	>	zú-!kɔ́mɔ́	hair (on head) (lit. head-hair)
	[RTR] root		**[ATR] root**			**Compound**	
b.	gbɛ́-rì	leg-SG	múní	lower torso	>	gbɛ́-múní	heel (lit. leg-lower.torso)
	bá-á	dog-SG	léé	small	>	bà-léé	puppy (lit. dog-small)

Table 3: ATR + RTR compound nouns, all dialects (Kuubezelle, 2013; Angsongna, 2023).

Disharmonic compounds in the opposite order, with RTR followed by ATR noun roots, are attested in all dialects except in Dagara. In the Dagara variety, the ATR roots in a compound can trigger harmony which targets the preceding RTR roots, as shown in **Table 4**. In contrast, no harmony happens in ATR + RTR compounds like those in **Table 3**, even in Dagara. Consequently, tongue-root harmony is considered ATR-dominant and regressive in varieties like Dagara (Casali, 2003).

[RTR] root		**[ATR] root**			**Dagara compound**	**Central Dàgáárè compound**	
pɔ́g-ɔ́	woman	lé	smallness	>	pò:lé	pɔ̀glé-é	daughter (lit. woman-smallness)
píír-ì	sheep	lé	smallness	>	pílè	pílé-é	lamb (lit. sheep-smallness)
gbɛ́r-ì	foot	bí-é	child	>	gbébìr	gbɛ́bìr-í	toe (lit. foot-child)
jɔ̀g-úú	pumpkin	bí-é	child	>	jò:bìr	jɔ̀gbír-í	pumpkin seed (lit. pumpkin-child)

Table 4: RTR + ATR compound nouns, Dagara and Central Dàgáárè (Somé, 1982, p. 183; Kuubezelle, 2013; Angsongna, 2023).

Overall, harmony in Dàgááre is root-controlled, making it difficult to determine dominance and directionality directly. The Dagara dialect is also a root-controlled system, but with its harmonizing of compounds, it shows an underlying ATR-dominant and regressive nature to the harmony. Other dialects, particularly Central Dàgááre, do not harmonize compounds at all, which is common in many harmony systems. It would be incautious to make a strong statement about the whole language based on evidence from a single dialect. That said, we note that nothing in Central Dàgááre is inconsistent with underlying ATR dominance and regressive directionality. We cautiously propose that all varieties of Dàgááre have this basic nature, but the evidence only appears in Dagara, because it is the only variety known to harmonize in compounds. However, we note that further investigation into and characterization of the underlying nature (i.e., where root-control is irrelevant) of this harmony system is still needed, and we leave that to future research.

2.2. Perception of (vowel) harmony background

Claims about how harmony is motivated by perceptibility are relatively common in the literature (e.g., Kaun, 1995; Ohala, 1994), but there has been very little perceptual work investigating these hypotheses directly. In this section, we overview some of the research on these hypotheses.

In Finley's (2012) study, English speakers were tasked with learning a rounding harmony pattern in an artificial language; while this study is not directly a perception task, it relates to perceptibility claims because the artificial pattern had high vowel harmony triggers in one condition and mid vowel harmony triggers in the other. The latter pattern is the more common one cross-linguistically: Kaun (1995) posits that, rather than being the mere result of something like historical change or language relatedness, this asymmetry exists because (1) mid vowels are less likely to be perceived and produced as canonically round, and (2) rounding harmony enhances perception of the [round] feature by repeating it across the domain (i.e., word). Participants were found to perform better (i.e., they more accurately learned the pattern and correctly generalized it more often to novel instances) in the mid trigger condition, implying that the bias towards perceptual enhancement exists in the learning of harmony patterns.

Kimper (2017) also conducted multiple experiments wherein English-speaking participants listened to a nonce word and then had to report whether a particular target vowel was present in that word. The nonce word always contained a transparent vowel in some position, while its other vowels were either harmonic or disharmonic for height and backness with each other and with the target vowel. Participants responded more quickly and more accurately when the target vowel was harmonic with the nonce word's vowels along some dimension. Again, this result implies that harmony facilitates perception of particular vowel features, though the effect was attenuated when the transparent vowel interceded between any of the harmonic vowels.

Similarly, Kimper (2011) conducted an AX task with English speakers, using the feature ATR crossed with a height feature. He hypothesized that ATR harmony maximizes discriminability,

leading to faster and more accurate responses when both vowels in a disyllabic stimulus are distinct than when only one. He also hypothesized that this advantage would be reduced when there is disagreement in height between the vowels. Both of his hypotheses were supported by his experimental results.

Other studies have returned results that seem to reject the aforementioned hypotheses. Grosvald (2009) investigated the perception and production of anticipatory vowel-vowel (VV) ([i/a]-[ə]) coarticulation by English speakers. While some participants did consistently show coarticulation on the target (ə) in their productions, and correctly perceived [i]-/[a]-colouring on the [ə] in others' productions, the within-subject correlation between production and perception was positive but weak. These findings oppose Ohala's (1994) theory that vowel harmony arises from listeners' perception of VV coarticulation becoming grammaticalized over time. In another study looking at VV coarticulation (Busà & Ohala, 1999), American English and Italian speakers listened to V1-C-V2 sequences and C-V(2) sequences, where V2 was on a continuum between [i] and [u], and V1 (if present) was always [i]. Participants perceived V2 as being more back in the disyllabic context (i.e., preceded by [i]) even when there was a long interval between V1 and V2. That this dissimilatory effect persists even over long periods indicates that, rather than stemming from compensation for perceived coarticulation carrying over from the V1 (which could reasonably only happen in fluid speech between vowels that are temporally close to each other), this is the result of more general contrast effects in perception. Again, we find opposition to Ohala's (1994) theory: If it is the case that contrast effects (basically listener hypercorrection) are at play when one listens to VV sequences, then how could vowel harmony ever arise if it is supposedly the result of listener hypocorrection? Some additional contrary evidence is also found in electroencephalography (EEG) studies. In these studies, event-related potentials (ERPs) between 50–240ms after stimulus onset are thought to correspond to activation of phoneme representations for native speech (see Gansonre, Højlund, Leminen, Bailey, & Shtyrov, 2018 for a review). If vowel harmony aids the live perception of speech sounds, then vowel harmony ought to be associated with ERPs at these early time intervals; however, some EEG studies have found that vowel harmony is only associated with later-occurring ERP components (if any). For example, Tuomainen (2001) conducted a Finnish auditory segmentation task and found that vowel harmony was associated with an N400 (negative deflection 400ms after stimulus onset, which is thought to reflect higher-order types of language processing such as word recognition and integration [Junge, Boumeester, Mills, Paul, & Cosper, 2021]). This result indicates that vowel harmony (and disharmony) could help listeners correctly and more quickly identify boundaries between words (e.g., a speaker of an ATR harmony language would know immediately to separate [begedɪgɪ] into two harmonic units, [bege] and [dɪgɪ]). However, the result also indicates that vowel harmony does not help listeners identify individual sounds, in contrast to the aforementioned hypotheses. In brief, direct experimental investigations into the

effects of vowel harmony on speech perception are few, they are often conducted with speakers of non-harmony languages, and they have returned largely mixed results.

Turning now to how these proposals might be applicable to ATR, there are proposals specifically suggesting that vowel harmony aids the perception of ATR contrasts, particularly of ATR contrasts that are considered 'difficult', such as those on high vowels. These proposals have, for the most part, not been tested experimentally. However, we can consider what the proposals of harmony more generally would predict when applied to ATR harmony. Considering Kaun's (1995) claim that rounding harmony extends perceptibility of a feature, spreading from vowels where rounding contrasts are weakly perceived to those where the perceptibility is strong, we might expect ATR harmony to similarly be triggered by sounds on which ATR is difficult to perceive, and target those where the contrast is most difficult. However, it is unknown experimentally which ATR distinctions are the most difficult to perceive, as evidence on perception of ATR is typically anecdotal. Our experiments will contribute to filling this gap. Interestingly, while the rounding harmony that Kaun discusses tends to be parasitic on height (e.g., in the Yowlumne dialect of Yokuts, rounding harmony only applies when the trigger and target vowels are of the same height [Kuroda, 1967]), ATR harmony often behaves in a different way: It is not usually parasitic, but instead it is common for a certain class of vowels to be 'dominant' (Casali, 2008). This means that it is difficult to apply Kaun's discussion of instances where rounding contrasts are difficult to perceive directly to thinking about ATR. For instance, there are ATR systems where /i/ triggers harmony but /ɪ/ does not, because [+ATR] is the dominant feature (see Rose, 2018 for a review); presumably, if the contrast between these two vowels is difficult to perceive, it would be of perceptual benefit for both to trigger harmony, rather than just one. In terms of Ohala's hypocorrection hypothesis, we generally expect listeners to attribute coarticulatory ATR effects to the target vowel. Coarticulation, like ATR harmony, tends to be anticipatory (e.g., Hyman, 2002), which means that we might expect the initial vowels in disyllabic forms to be subject to hypocorrection. However, little is known specifically about ATR coarticulation, so the question of which types of segments we would expect to see as most perceptually susceptible to hypocorrection is a difficult one. Overall, then, linking these previous claims to ATR harmony is complex given substantial gaps in the current understanding of ATR, and so this paper aims to establish a baseline for perception of ATR such that some of these more complex issues can be addressed in the future.

In the most general case, however, disregarding triggers and targets, the expectation from these previous works is clear: If the perceptibility hypothesis holds, then when asked to distinguish two similar words, we expect participants to perform better with the assistance of harmony. For instance, distinguishing ɪ-ɪ sequences from e-e should be easier than distinguishing either of those from e-ɪ sequences: There are two opportunities to hear the ATR feature in the harmonic forms, but only one in the disharmonic forms. Such a result would be in line with Kimper's (2011) results with English-speaking participants.

2.3. Our study

To investigate this and related hypotheses, two ABX discrimination task experiments were conducted. The ABX task was selected because it is a popular paradigm in perceptual experiments, and because of its ease of explanation and presentation. In the first experiment, participants were tested on their ability to distinguish three vowel contrast types in monosyllabic (CV) and harmonic disyllabic (CVCV) nonce word contexts. This experiment served as a baseline to establish listeners' ability to discriminate the vowel contexts and to examine whether harmony helps discrimination. In the second experiment, participants were tested on a single contrast type in monosyllabic (CV) and both harmonic and disharmonic disyllabic (CVCV) nonce word contexts. This experiment served to examine whether disharmony impedes perception relative to harmonic and monosyllabic contexts. The next section presents more details about Experiment 1, followed by Section 4 which deals with Experiment 2.

3. Experiment 1

3.1. Methodology

3.1.1. Participants

Twenty-four native Dàgáárè speakers (of any dialect) participated in Experiment 1. Reading ability in English was also required to understand the experiment instructions. The experiments were conducted entirely online and programmed using jsPsych (JavaScript) (de Leeuw, 2015). Participants were recruited via word-of-mouth, and they provided informed consent by reading a consent form and checking a box to indicate consent prior to beginning the experiment. All participants were in Ghana and compensated $15CAD for their participation.

All participants completed a background questionnaire prior to the ABX task. This elicited information such as age, gender, and native dialect(s) of Dàgáárè. For Experiment 1, there were two female and 22 male participants, mean age 25.54 years. The following native dialects were represented among the participants: Dagaare (15 participants), Waale/Waali (10), Dagara (6), Dagaari (2), and Birifo (1).[2]

3.1.2. Materials

Nonce words of form CV and CVCV were produced by a phonetically trained male Dàgáárè native speaker. To increase comparability across conditions, the only consonant used was [k] (always as an onset), and the only tone used was low. We chose a stop onset to reduce interactions with surrounding vowels, and we chose a velar in order to avoid real words ([ti], [di], [pi], and [be] with low tone are all words in Dàgáárè). The voiceless velar stop was chosen over the voiced option to increase the sonority differential between the onset and the vowels of interest.

[2] Some participants indicated more than one native dialect.

Given that there are no known differences in harmony operation by backness in the language (i.e., both front and back vowels participate in harmony in the same way), we used only front vowels in this study to decrease the length of the experiment, while increasing the number of repetitions per A/B pair in the ABX tasks.

For Experiment 1, we used four nonce words in the form CV (i.e., [ke, kɛ, ki, kɪ]) and eight in the form CVCV (i.e., [keke, kɛkɛ, kiki, kɪkɪ, keki, kike, kɛkɪ, kɪkɛ]). All nonce words were harmonic, making them possible but non-existent Dàgáárè words.

Stimuli were recorded by a phonetically trained male native speaker of Dàgáárè, using a Shure WH30 headset microphone at the sampling rate of 48.1 kHz in WAV format. The microphone was connected to a Zoom Q8 video recorder. The audio from the microphone was saved as a separate file at the same time as the video file. Six to eight tokens of each nonce word were recorded. The best recordings were selected, determined based on representativeness of the intended category as judged by the authors and the phonetically trained native speaker. Individual trials never included identical tokens. For all selected tokens, all authors and the phonetically trained native speaker agreed that the IPA transcription matched the intended nonce word. Stimuli were segmented in Praat (Boersma & Weenink, 2022).

To ensure that the vowels in the stimuli were accurately produced, we plotted them prior to running the perception experiment. Formants were measured by a Praat script at the midpoint of each vowel, with the vowel boundaries determined as where clear periodicity began and ended. This includes all vowels in both experiments and in both positions, and it was particularly important given that some of the stimuli in Experiment 2 were illicit as (non-compound) Dàgáárè words. The formant plot is shown in **Figure 1**, and a table of formant means and standard deviations is shown in **Table 5**. A graph of the vowel durations is provided in **Figure 2**. All graphs and statistics in this paper were done using R (R Code Team, 2022) in RStudio (RStudio Team, 2022).

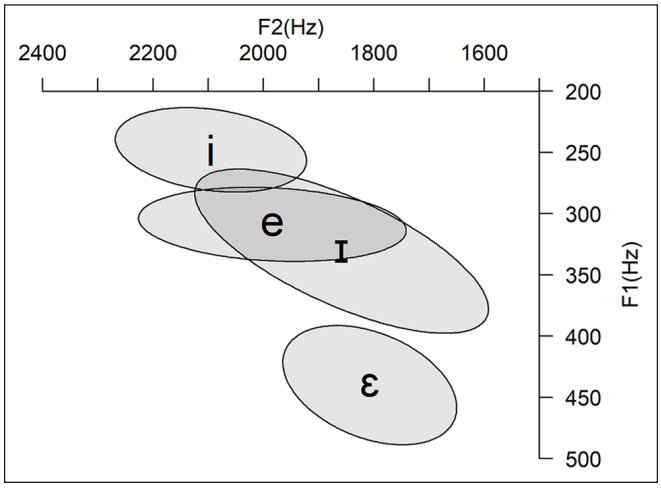

Figure 1: Formant plot for front vowels in Dàgáárè.

		i	ɪ	e	ɛ
F1	Mean	247.90	330.58	308.73	440.01
	SD	22.42	43.59	19.70	31.63
F2	Mean	2095.11	1858.47	1983.77	1807.46
	SD	112.65	172.91	157.62	102.27
Duration	Mean	94.05	84.40	107.73	130.42
	SD	23.17	20.35	15.38	18.50

Table 5: Mean and Standard Deviation of formant values and duration.

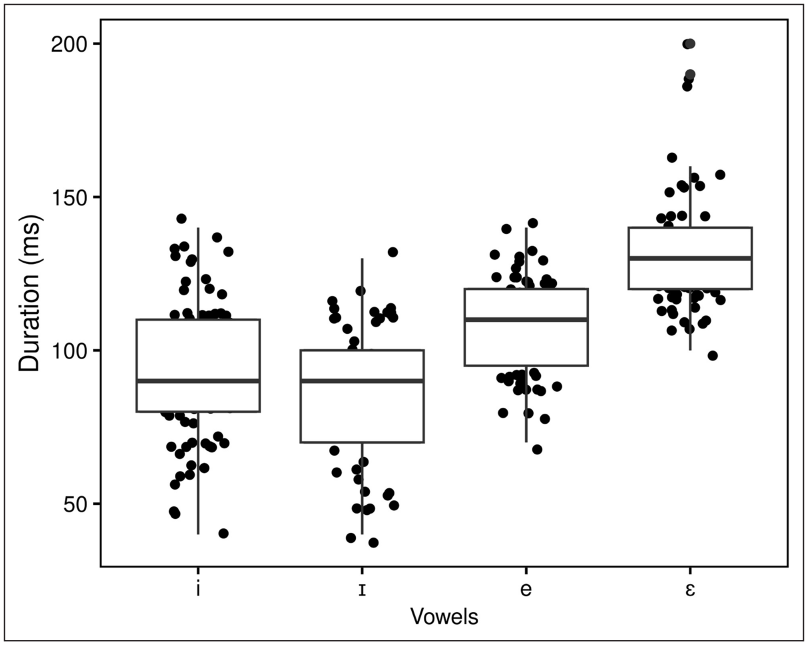

Figure 2: Vowel durations in the stimuli.

3.1.3. Procedure

The experiment consisted of a brief practice block followed by two experimental blocks of equal length. At the start of the practice block, participants were instructed that their task was to listen to sequences of three words ("A", "B", "C", where participants were shown "C" for the word that represents "X" in "ABX") and report whether word "C" was the same as word "A" or word "B". They were informed that they would have only two seconds to provide a response before the experiment would advance to the next trial automatically. They were told that they would not be penalized for giving wrong answers (response correctness was not indicated to them). Following the instructions, participants were given eight practice trials in a randomized order before proceeding to the first main experimental block.

During each trial, the participant heard three stimuli in succession, with a 500ms inter-stimulus interval (during which the screen was blank). Note that the stimuli used in each block were all preloaded before the trials started playing, such that inter-stimulus intervals would always be consistent, even if the participant experienced internet problems. The letters "A", "B", and "C" appeared on the screen while the word was auditorily presented. After the third stimulus was presented, a response screen appeared asking the participant to report whether "C" was the same as "A" or "B" by clicking the corresponding button on the screen. They had 2000ms to respond, and the inter-trial interval (between the participant's response and the start of the next trial) was also 2000ms. Trials were divided into two blocks, and participants were given the opportunity to take a short break between blocks (as well as after the training block).

Participants ran the experiments online, on their own computer or smartphone. They were asked at the start of the experiments to wear headphones to ensure that they heard all the stimuli as clearly as possible. They were also given a volume check, wherein they heard a tone and were told to adjust their headphone volume until it could be heard clearly. They were given the option to click to hear the tone as many times as necessary to adjust their volume appropriately.

The "A" and "B" pairings were arranged into three groups based on vowel contrast type (i/ɪ, e/ɪ, e/ɛ), as well as four groups based on context type (monosyllables, before harmonic high (i/ɪ), before harmonic mid (e/ɛ), and before harmonic cross-height (e/ɪ)). Note that "before harmonic cross-height" refers to a context with the cross-height pair e/ɪ, which differ in both ATR and height but were shown to overlap in the formant space. This means that, in this context, the ATR vowels [i] and [e] occur in the context [e], while the RTR vowels [ɪ] and [ɛ] occur in the context [ɪ]. Contrast and context, therefore, had the same options, aside from the context 'monosyllabic'; contrast type refers to which pair (i/ɪ, e/ɪ, e/ɛ) occurs in V1 position, while context type refers to which of the same pairs (i/ɪ, e/ɪ, e/ɛ) occurs in V2 position, and was monosyllabic if no second vowel was present (in monosyllables).

These contrast and context types resulted in 12 pairings (see **Table 6**). Each pairing had four possible configurations (two word orderings [e.g., A = ki and B = kɪ, or A = kɪ and B = ki] × two ABX configurations [ABB, where X = B, or ABA, where X = A]), giving 48 unique trial types. Given that all disyllabic forms were harmonic, all disyllabic pairings had two differences; for example, kiki versus kɪkɪ differ in both the first and second vowels.

Both blocks contained two repetitions of all 48 unique trial types in a randomized order for each participant. Each instance of a nonce word used a different token, so that, for example, even if A and X were the same nonce word, they were always different tokens of it (e.g., A = ki-1 and X = ki-2 in repetition 1, A = ki-3 and X = ki-4 in repetition 2). In total, there were 48 (unique trial types) × two (repetitions) × two (blocks) = 192 test trials. At the start of each block, there was a screen that repeated the instructions for the participant.

Vowel Contrast Type (V1)	Context Type (V2)			
	monosyllabic	Disyllabic		
		before harmonic high (i/ɪ)	before harmonic mid (e/ɛ)	before harmonic cross-height (e/ɪ)
i/ɪ	k**i** vs. k**ɪ**	k**i**ki vs. k**ɪ**k**ɪ**	k**i**ke vs. k**ɪ**k**ɛ**	k**i**ke vs. k**ɪ**k**ɪ**
e/ɪ	k**e** vs. k**ɪ**	k**e**ki vs. k**ɪ**k**ɪ**	k**e**ke vs. k**ɪ**k**ɛ**	k**e**ke vs. k**ɪ**k**ɪ**
e/ɛ	k**e** vs. k**ɛ**	k**e**ki vs. k**ɛ**k**ɪ**	k**e**ke vs. k**ɛ**k**ɛ**	k**e**ke vs. k**ɛ**k**ɪ**

Table 6: Word pairings for Experiment 1 (vowels of interest bolded).

3.2. Hypotheses

Experiment 1 consisted of three vowel contrast types and four contexts. As noted, ATR contrasts in high vowels are predicted to be particularly difficult to perceive based on typological findings (Rose, 2018), and high RTR/mid ATR contrasts (e.g., e/ɪ) have been found to be difficult in previous experiments on other languages (Rose et al., 2023). Additionally, high RTR/mid ATR vowels are acoustically overlapping in Dàgáárè (see **Figure 1** in the Materials section). As such, we predicted that e/ɛ contrasts will have the highest accuracy, with significantly lower accuracy in the other contrast types. Given that Rose et al. (2023) found that acoustic overlap was the greatest predictor of perception results, we predict that the cross-height e/ɪ contrast will have the lowest accuracy rates.

In terms of contexts, harmony is said to aid in perceptibility, and all our disyllabic contexts were harmonic, giving participants more opportunities to hear the harmonic feature: They have two opportunities to hear ATR/RTR instead of just one, because both vowels differ. This prediction was discussed in Section 2.2 as the most general case that we would expect based on perception-related hypotheses in the harmony literature and based on previous experiments like Kimper (2011). As such, we predict that all disyllabic contexts should have higher accuracy than the monosyllabic context. Within the disyllabic harmonic contexts, we noted above that 'context' is effectively the same as 'contrast', just in the V2 position rather than V1. As such, we predict that accuracy should reflect the same contrast-specific considerations noted above. Specifically, since mid vowels are predicted to be the easiest to accurately distinguish, in part due to the lack of acoustic overlap of ɛ with any of the other vowels, the harmonic mid context (e/ɛ) should have the greatest accuracy, followed by harmonic high (i/ɪ), followed by cross-height (e/ɪ), in the same way as for the contrast types.

The hypothesis follows because mid vowels are expected to have the strongest perceptual contrast, they should also be the best place to hear the effects of harmony in that ɛ is quite distinct from the other vowels in the language, and so they should improve performance the most in harmonic disyllables.

3.3. Results

The proportion correct for each participant for each trial type was calculated. Null responses (i.e., trials where a participant did not respond) were excluded from this calculation; this decision resulted in the exclusion of one participant in Experiment 1 whose responses all registered as null. As such, the results are based on 23 participants in Experiment 1. Of the participants who were included, a total of 184 trials (average of eight per participant) were excluded for being null responses. We then calculated the mean and standard deviation of reaction time for each participant, for all trials with non-null response, and excluded any trials where a participant's reaction time was greater than 2.5 standard deviations away from their mean. This resulted in the exclusion of an additional 125 trials total (average of 5.4 per participant).

Responses were submitted to a linear regression mixed effects models on proportion correct (Baayen, Davidson, & Bates, 2008; Jaeger, 2008). All statistical models throughout the paper were implemented using the function lmer from the package lme4 (Bates, Mächler, Bolker, & Walker, 2015). Proportion correct was calculated for each subject for each condition (i.e., for each subject in each context/contrast combination). The range in total numbers of tokens that went into each proportion correct was 10 to 16, depending on how many null responses and responses excluded due to reaction time a participant had for a given condition. The mean number of included tokens in each proportion correct for Experiment 1 was 14.88 and the standard deviation was 1.28.

For all models, the random effect was a by-participant random intercept; item effects were not included as random intercepts due to collinearity with the fixed effects, and the models did not converge with random slopes. It is worth noting that the Participants random effect, along with the fact that our conclusions are based on differences between stimulus pairs, means that any extra by-subject variability due to online presentation should not have affected the results.

The fixed effects were Context and Contrast. Context has levels i/ɪ, e/ɪ, e/ɛ, and monosyllabic (reference level); Contrast has levels i/ɪ, e/ɪ, e/ɛ (reference level). The default dummy coding scheme for categorical variables was used. We also ran a version of the Experiment 1 model for comparison purposes that included the interaction between context and contrast. Using a model comparison ANOVA, we determined that the BIC of the model without the interaction term was lower than the model with the interaction (BIC –274.87 vs. –254.92). As such, the model without interactions performed better, and is therefore the one reported here.

Note that we did not consider whether X was A or B as an effect, because the experiment was counterbalanced by this property, and so any effect of this nature (e.g., if any participants were naturally biased towards choosing A over B) would be balanced over the course of the experiment and would not affect the results.

The reference levels in the model were chosen with our hypotheses in mind. We hypothesized that the e/ɛ contrast will have significantly better performance than other contrasts, due to the acoustic separation. The model, therefore, tests the other contrasts in comparison to this one,

and we expect performance will be worse on i/ɪ and e/ɪ compared to e/ɛ. In terms of contexts, we hypothesized that the three disyllabic contexts, all of which are harmonic, will show better performance than monosyllabic contexts, due to opportunities to hear the ATR feature on both vowels. The model compares the three disyllabic contexts against the monosyllabic one; we expect performance to be better on the disyllabic contexts. However, in order to test additional nuances, we also re-levelled the factors and ran the model with other reference levels. Doing so allows us to see whether the hypothesized accuracy trend of e/ɛ > i/ɪ > e/ɪ holds in the data, as not all of these comparisons can be directly tested with a single reference level. This method allows us to compare other levels of the factors to test our hypotheses in greater detail.

The monosyllabic context and e/ɛ context had approximately equal performance (mean = 69.9% and 70.1% respectively), while other contexts had worse performance (means 66.8% and 64.8%), though only the "cross-height" (e/ɪ) context was significantly different from monosyllabic ($p < 0.05$; see **Table 8**). Moreover, both other contrasts (e/ɪ and i/ɪ) had worse performance than the mid e/ɛ contrast (means = 70.1% vs. 65.7%, 68.0%), though again the difference only reached significance for the "cross-height" [ɪ]~[e] contrast (see **Table 8**). **Table 7** illustrates the mean and standard deviation of proportion correct for all context/contrast combinations in Experiment 1 (obtained using the function ddply from the package plyr (Wickam, 2011)), while **Tables 8–10** show the model output. In **Table 8**, the intercept represents the level at the reference level, which is the contrast e~ɛ in the monosyllabic context, while the remaining rows compare particular context and contrast levels to those reference levels. In order to compare all levels directly, we also repeated the model with all other possible

	Context (V2)	Contrast (V1)	mean(prop)	sd(prop)	Stimulus type
1	monosyll	e~ɛ	0.74	0.24	ke~kɛ
2	monosyll	e~ɪ	0.68	0.22	kɪ~ke
3	monosyll	i~ɪ	0.68	0.24	ki~kɪ
4	i~ɪ	e~ɛ	0.67	0.24	keki~kɛkɪ
5	i~ɪ	e~ɪ	0.68	0.20	keki~kɪkɪ
6	i~ɪ	i~ɪ	0.65	0.19	kiki~kɪkɪ
7	e~ɛ	e~ɛ	0.70	0.21	keke~kɛkɛ
8	e~ɛ	e~ɪ	0.70	0.19	keke~kɪkɛ
9	e~ɛ	i~ɪ	0.70	0.21	kike~kɪkɛ
10	e~ɪ	e~ɛ	0.69	0.17	keke~kɛkɪ
11	e~ɪ	e~ɪ	0.57	0.20	keke~kɪkɪ
12	e~ɪ	i~ɪ	0.68	0.20	kike~kɪkɪ

Table 7: Mean and standard deviation of proportion correct by Context and Contrast for Experiment 1.

mod1 <– lmer(prop ~ context+contrast + (1|subject),data = results_1)						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	**Significance**
(Intercept)	0.721	0.04	31.16	17.96	<2e-16	***
Context i~ɪ	–0.030	0.02	248.00	–1.49	0.138	
Context e~ɛ	0.002	0.02	248.00	0.12	0.905	
Context e~ɪ	–0.051	0.02	248.00	–2.491	0.013	*
Contrast e~ɪ	–0.044	0.02	248.00	–2.48	0.014	*
Contrast i~ɪ	–0.022	0.02	248.00	–1.22	0.224	

Table 8: Model output for Experiment 1 with monosyllabic as the reference level for context and e~ɛ as the reference level for contrast.

	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	**Significance**
(Intercept)	0.626	0.04	31.16	15.61	2.8e-16	***
Context i~ɪ	0.020	0.02	248.00	1.00	0.317	
Context e~ɛ	0.053	0.02	248.00	2.61	0.010	**
Context monosyllabic	0.051	0.02	248.00	2.49	0.013	*
Contrast i~ɪ	0.022	0.02	248.00	1.26	1.257	
Contrast e~ɛ	0.044	0.02	248.00	2.48	0.014	*

Table 9: Model output for Experiment 1 with e~ɪ as the reference level for contrast and context.

	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	**Significance**
(Intercept)	0.669	0.04	31.16	16.70	<2e-16	***
Context i~ɪ	–0.020	0.02	248.00	–1.00	0.317	
Context e~ɛ	0.033	0.02	248.00	1.61	0.109	
Context monosyllabic	0.030	0.02	248.00	1.49	0.138	
Contrast e~ɪ	–0.022	0.02	248.00	–1.26	0.210	
Contrast e~ɛ	0.022	0.02	248.00	1.22	0.224	

Table 10: Model output for Experiment 1 with i~ɪ as the reference level for contrast and context.

reference levels; **Table 9** shows e~ɪ as the reference level for both contrast and context, while **Table 10** shows i~ɪ as the reference level for both contrast and context. From these tables, we can conclude that Context e~ɪ significantly differs from the monosyllabic Context ($p < 0.05$) and e/ɛ Context ($p < 0.05$), Contrast e~ɪ significantly differs from the Contrast e/ɛ ($p < 0.05$), but no other differences are significant.

Figures 3 and **4** illustrate graphically the proportion correct by Context (faceted by Contrast and without facets respectively) and **Figures 5** and **6** illustrate proportion correct by Contrast (faceted by Context and without facets respectively). All results graphs (excluding the acoustic plots in Section 3.1.2) were created using the function ggplot in the package ggplot2 (Wickam, 2016). The facets sort the data into separate graphs by the faceted criterion, while the non-faceted versions show all data combined. For example, **Figure 3** shows three smaller graphs, one for each Contrast, and within each plot, the x-axis shows the Context and the y-axis shows the proportion correct, specifically for the given Contrast within that graph. **Figure 4** has the same information, but not divided by Contrast, so that all three Contrast levels are combined into a single graph, again with Context on the x-axis and proportion correct on the y-axis. The violin plots show the density at a particular point, such that a point on the y-axis where the 'violin' is wider, like just above the 0.8 line for the e/ɛ contrast in **Figure 4**, represents a proportion correct that was very common in the results. Similarly, narrow parts of the violin plot show places where the density

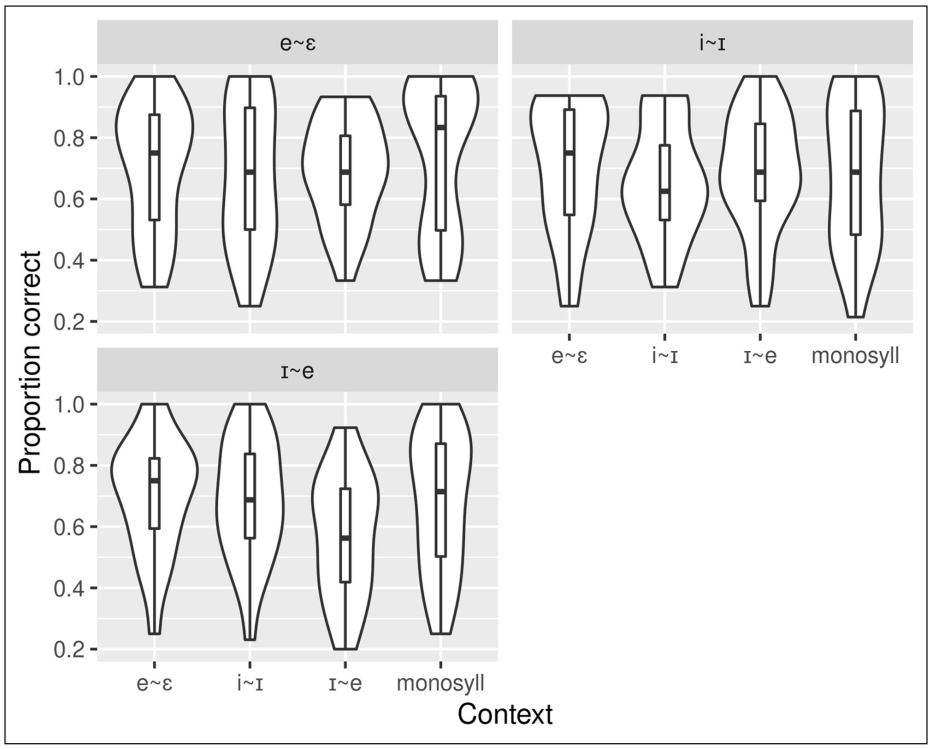

Figure 3: Proportion correct by Context (=V2) (facet by Contrast (=V1)).

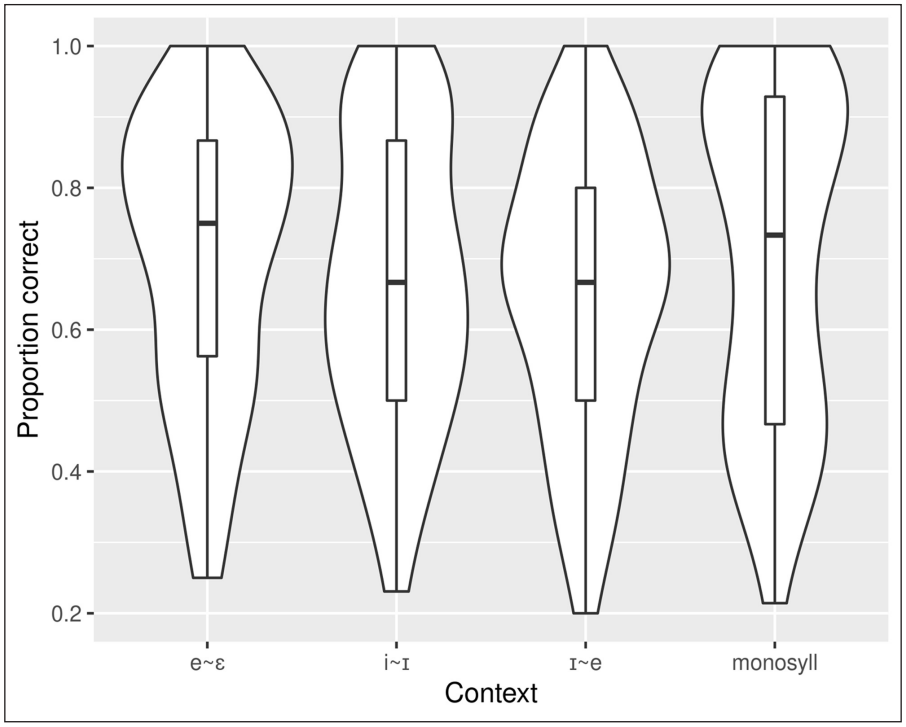

Figure 4: Proportion correct by Context (=V2).

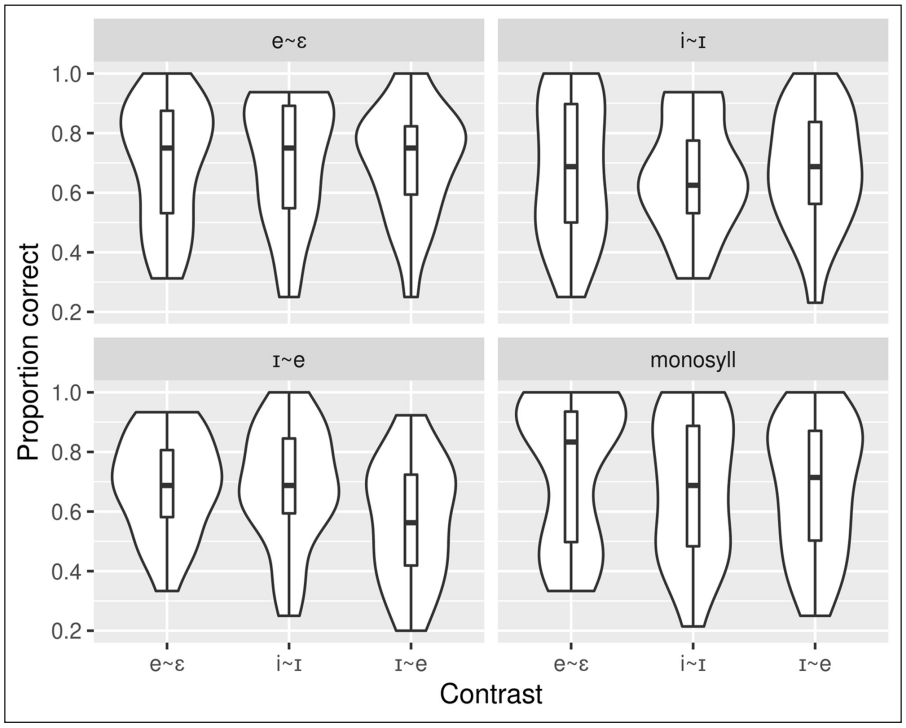

Figure 5: Proportion correct by Contrast (=V1) (facet by Context (=V2)).

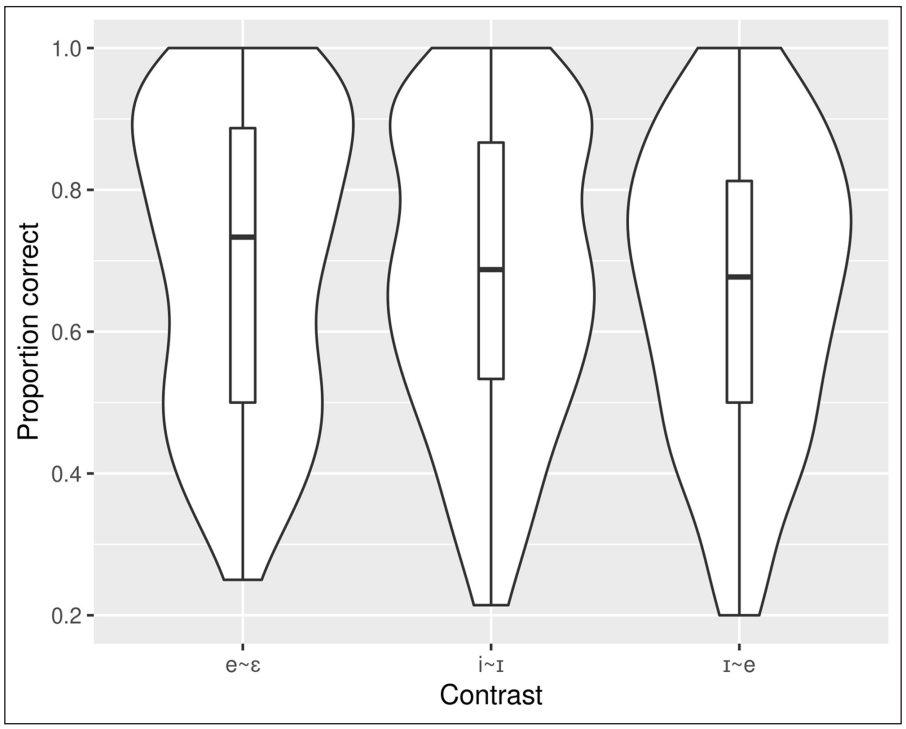

Figure 6: Proportion correct by Contrast (=V1).

of respondents with that proportion correct was lower. Inside each violin is a standard boxplot, where the box shows the location of the first to third quartile of the data, with the thick line in the middle showing the median. The 'whiskers' of the boxplot, which are the lines going up and down from the box, show the range of the data minus any outliers.

To visually check that the results were not due to a single participant with particularly low or high results, we also created a boxplot of results by-participant, shown in **Figure 7**. In this graph, we see that we have a combination of participants who do quite well (e.g., participants 2, 6, 7, etc.) and those whose results are lower (e.g., participants 1, 5, 8, etc.). There are no participants who are visual outliers in the results, in that none of the participants are consistently getting zero correct and we have a general spread of participants across the accuracy levels. As such, it does not appear to be necessary to remove any participants for being extreme outliers to the results.

To summarize these results, Experiment 1 found that participants had significantly lower accuracy on e/ɪ as both V1 (Contrast) and V2 (Context), compared to the monosyllabic and e/ɛ levels. The pairs e/ɛ and i/ɪ did not significantly differ from each other as contrasts, nor did they differ significantly from monosyllabic when looking at contexts. However, there was a trend for i/ɪ to have lower accuracy as both Contrast and Context compared to e/ɛ and monosyllabic.

In terms of the hypotheses, the trend matched what we expected within disyllabic words: Accuracy was highest for e/ɛ as Contrast/Context, followed by i/ɪ, followed by e/ɪ. Only e/ɪ

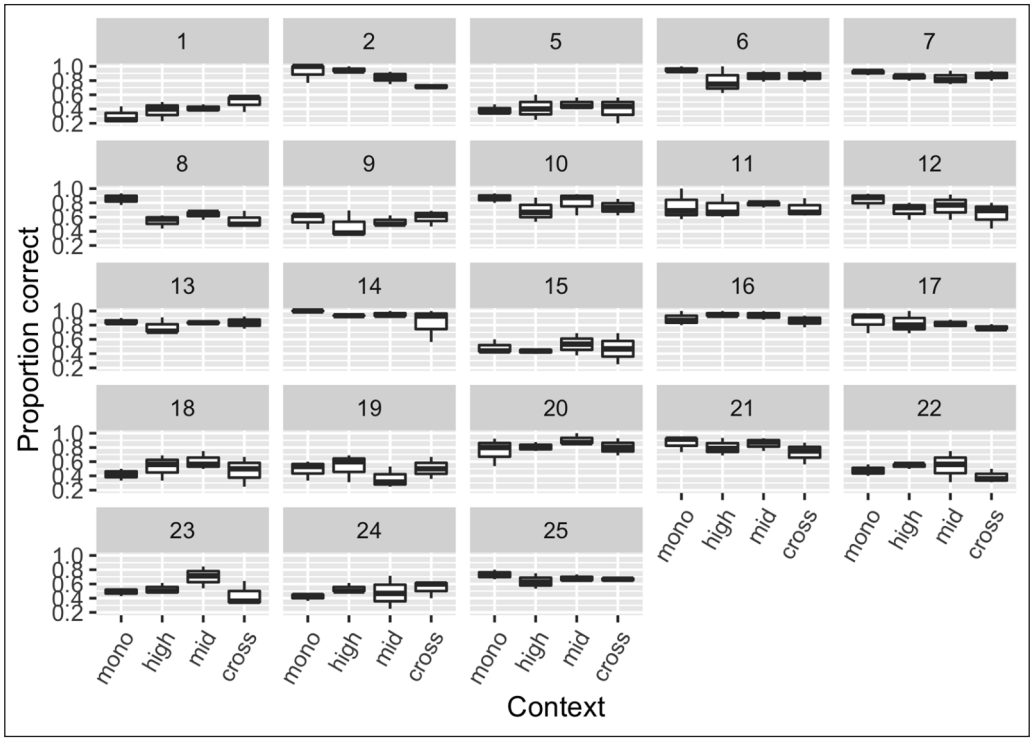

Figure 7: Proportion correct by Context for each participant.

was significantly different from e/ɛ. However, participants did not have greater accuracy on harmonic disyllabic than on monosyllabic words, in contrast to our prediction. Indeed, accuracy on the context e/ɪ was significantly lower than on monosyllabic contexts.

Experiment 1 suggests that e/ɪ is the hardest contrast for speakers. However, because comparisons were always between words with two differences in the disyllabic items, we cannot draw firm conclusions about the effect of harmony on perception. In contrast to our predictions, harmony does not aid perception, in that monosyllables are, on average, discriminated better than harmonic disyllables. However, Experiment 1 leaves open the possibility that harmony helps within disyllables. In other words, disyllables may be easier to discriminate when harmonic than when disharmonic. Experiment 2 addresses this possibility; it focuses on the e/ɪ contrast and includes disharmonic disyllabic items, in addition to monosyllabic and harmonic ones.

4. Experiment 2

4.1. Methodology

4.1.1. Participants

Twenty-five native Dàgáárè speakers (of any dialect) participated in Experiment 2. The same recruitment techniques and compensation applied as in Experiment 1. For Experiment 2, there

were three female and 22 male participants, mean age 25.5 years. The following native dialects were represented among the participants: Dagaare (17 participants), Waale/Waali (10), Dagara (7),[3] Dagaari (1), and Birifo (2).

Several participants in Experiment 2 had also participated in Experiment 1. Of the participants in both experiments, four participated in Experiment 1 only and three participated in Experiment 2 only, leaving the rest as participants in both experiments. Given the small number of participants who did only one of the two experiments, we cannot draw any conclusions about effects of completing both versus only one experiment.

4.1.2. Materials

The same speaker and recording equipment were used for the stimuli as for Experiment 1. Again, the only consonant was [k] and the only tone was low. However, Experiment 2 used only a single vowel contrast, [ɪ] versus [e]. Experiment 2 used two nonce words of the form CV (i.e., [ke, kɪ]) and four of the form CVCV (i.e., [keke, kɪkɪ, kekɪ, kɪke]), two of which were disharmonic. As with Experiment 1, all stimuli used were ones that all authors and the native speaker agreed matched the IPA transcription.

4.1.3. Procedure

The general procedure for Experiment 2, including the task, the instructions, the interstimulus intervals, and the blocks, were the same as for Experiment 1. The only difference was that participants were only given six practice trials in Experiment 2, and the number of stimuli within the block also differed.

For Experiment 2, there was only a single vowel contrast type (e/ɪ), which was the most difficult in Experiment 1, but there were six context types (monosyllables, before ɪ, before e, harmonic, after ɪ, and after e). This produced six "A" and "B" word pairings as outlined in **Table 11**. In each of the before and after ɪ and e trials, one of the two nonce words was disharmonic.

monosyllables	**before ɪ**	**before e**	**harmonic**	**after ɪ**	**after e**
kɪ vs. k**e**	k**ɪ**kɪ vs. k**e**kɪ	k**ɪ**ke vs. k**e**ke	k**ɪ**kɪ vs. k**e**ke	kɪkɪ vs. kɪk**e**	kekɪ vs. kek**e**

Table 11: The six "A" and "B" configurations for Experiment 2 (vowels of interest bolded).

The experimental design was the same as in Experiment 1, except that the two test blocks contained 48 trials each (six word pairings × two word orderings (e.g., A = ke and B = kɪ, or

[3] Note that there were not enough Dagara-speaking participants to tell whether the presence of harmony in certain compounds in their dialect (discussed in Section 2.1) affected their perception of the disharmonic nonce forms in our experiment.

A = kɪ and B = ke) × two ABX trial types (ABA or ABB) × two repetitions of each trial with a different token configuration in each one), for a total of 96 trials in the experiment.

4.2. Hypotheses

Experiment 2 had monosyllabic contexts (kɪ vs. ke), harmonic disyllabic contexts (kɪkɪ vs. keke), and four types of disharmonic disyllabic contexts (before ɪ, before e, after ɪ, after e), all using the vowel contrast that was the most difficult for listeners in Experiment 1. We predict that, compared to monosyllabic contexts, accuracy should be greater in harmonic contexts and lower in disharmonic contexts, given that harmony is said to aid perceptibility. Specifically, listeners might hypocorrect the disharmony in perception, such that we expect listeners to have the most difficulty with the four contexts comparing a disharmonic disyllabic nonce word to a harmonic word that differs in only a single vowel. Beyond hypocorrection, this result would also be predicted by the discussion in Sections 2.2 and 3.2 about one versus two differences: The harmonic context has a pair of words where both vowels differ, giving two opportunities to hear ATR, while the two nonce words in each disharmonic context differ in only a single vowel, resulting in fewer opportunities for listeners to be certain of the ATR features in each word.

Since harmony in Dàgáárè is root-controlled and bidirectional, we do not necessarily predict any directionality effects. However, at least in the Dagara variety of Dàgáárè, ATR is posited to be the dominant feature (Casali, 2003). Further, the harmony is said to be regressive (again, at least in Dagara), as noted with discussion of compounds in Section 2.1. Due to the apparent ATR dominance, we might expect [e] to show greater pressure to trigger harmony compared to [ɪ], because it is ATR. Under a theory where harmony is phonologized perceptual compensation, we might expect the degree of perceptual compensation to reflect properties like ATR dominance and regressivity. If so, then we might predict that disharmonic forms involving [e] in V2 position would be more likely to be inaccurately perceived as harmonic. Such contexts should then be more difficult and therefore show lower accuracy. In other words, if (a) [e] is more likely to trigger harmony, because it is ATR, (b) regressive is a more likely directionality, and (c) these facts are due to perceptual hypocorrection, then kɪke vs. keke should be particularly difficult to distinguish. Indeed, we expect listeners to incorrectly harmonize the former in perception.

4.3. Results

As for Experiment 1, the proportion correct for each participant for each trial type was calculated. Null responses (i.e., trials where a participant did not respond) were excluded from this calculation, resulting in the exclusion of a total of 86 trials with null responses (average of 3.4 per participant) in Experiment 2. As for Experiment 1, we excluded any trials where a participant's reaction time was greater than 2.5 standard deviations away from their mean. This resulted in the exclusion of an additional 56 trials total (average of 2.2 per participant) in Experiment 2.

Responses were submitted to a linear regression mixed effects models on proportion correct (Baayen et al., 2008; Jaeger, 2008). As for Experiment 1, the random effect was a by-participant random intercept; item effects were not included as random intercepts due to collinearity with the fixed effects, and the models did not converge with random slopes. As before, it is worth noting that the Participants random effect, along with the fact that our conclusions are based on differences between conditions, means that any extra by-subject variability due to online presentation should not have affected the results. For Experiment 2, the fixed effect was Context, as there was only one contrast, as shown in **Table 11**. The levels were before e, after e, before ɪ, after ɪ, harmonic, and monosyllabic. Given that our experiment involved comparing monosyllabic contexts to disyllabic contexts and comparing harmonic contexts to disharmonic contexts, we ran the model twice, once with the monosyllabic context as the reference level and once with the harmonic context as the reference level. The former allows us to compare performance on monosyllabic contexts to all other (i.e., disyllabic) contexts, while the latter allows us to directly compare harmonic to disharmonic disyllabic contexts.

In Experiment 2, like in Experiment 1, all other contexts were again worse than the monosyllabic context, in terms of lower proportion correct, as shown in **Table 12**, though only slightly so for the harmonic context (means 68% vs. 67%). Using the model described above and illustrated in **Tables 13–14**, this trend was significant ($p < 0.05$) for the contexts before and after [e], and for the context before [ɪ]. In other words, kɪke vs. keke, kekɪ vs. keke, and kɪkɪ vs. kekɪ were significantly harder to distinguish than kɪ vs. ke. The difference from monosyllabic was not significant for the context after [ɪ] (kɪkɪ vs. kɪke) or for the harmonic context (kɪkɪ vs. keke). **Table 12** shows the means and standard deviations of proportion correct for each Context in Experiment 2, while **Tables 13–14** show the model output, comparing to monosyllabic and harmonic reference levels respectively. **Figure 8** illustrates the results graphically. In **Figure 8**, the x-axis is Context and the y-axis is Proportion Correct, parallel to the graphs in Experiment 1, and the graph in **Figure 8** is similarly a violin plot with a boxplot overlaid. Given that there was only one fixed effect, no faceted graphs were created. The graph shows clearly how harmonic

Context	mean(prop)	sd(prop)
monosyllabic	0.68	0.21
harmonic	0.67	0.19
before ɪ	0.59	0.15
before e	0.57	0.17
after ɪ	0.61	0.18
after e	0.55	0.13

Table 12: Mean and standard deviation of proportion correct by Context for Experiment 2.

mod2 <– lmer(prop ~ context + (1|subject),data = results_2)						
	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**	**Significance**
(Intercept)	0.677	0.03	85.43	19.469	<2.00E-16	***
Context harmonic	–0.002	0.04	120.00	–0.053	0.958	
Context before ɪ	–0.087	0.04	120.00	–2.239	0.027	*
Context before e	–0.106	0.04	120.00	–2.709	0.008	**
Context after ɪ	–0.064	0.04	120.00	–1.636	0.104	
Context after e	–0.127	0.04	120.00	–3.256	0.001	**

Table 13: Model output for Experiment 2 with monosyllabic as the reference level.

	Estimate	**Std. Error**	**Df**	**t value**	**Pr(>|t|)**	**Significance**
(Intercept)	0.675	0.03	85.43	19.410	<2e-16	***
Context monosyllabic	0.002	0.04	120.00	0.053	0.958	
Context before ɪ	–0.085	0.04	120.00	–2.186	0.031	*
Context before e	–0.104	0.04	120.00	–2.656	0.001	**
Contrast after ɪ	–0.062	0.04	120.00	–1.583	0.116	
Contrast after e	–0.125	0.04	120.00	–3.203	0.002	**

Table 14: Model output for Experiment 2 with harmonic as the reference level.

and monosyllabic contexts are comparable (the violins and boxes are in similar places with similar densities), but the disharmonic ones are generally situated at a lower proportion correct, with the density (wider part of the plot) at lower proportion correct values. This graph therefore visually reflects the means and statistical results described earlier.

Similarly to Experiment 1, we created a boxplot of results by participant, shown in **Figure 9**. Again, there is a spread of results, but it does not appear that any participant is a particular outlier who might be unduly affecting results. It is worth noting that there are a few participants in **Figure 9** who have higher accuracy on harmonic items than on monosyllabic ones, namely participants 4, 6, 10, 14, 20, 21, and 23.[4] This pattern was the predicted one, but is not the

[4] We thank the associate editor for this observation.

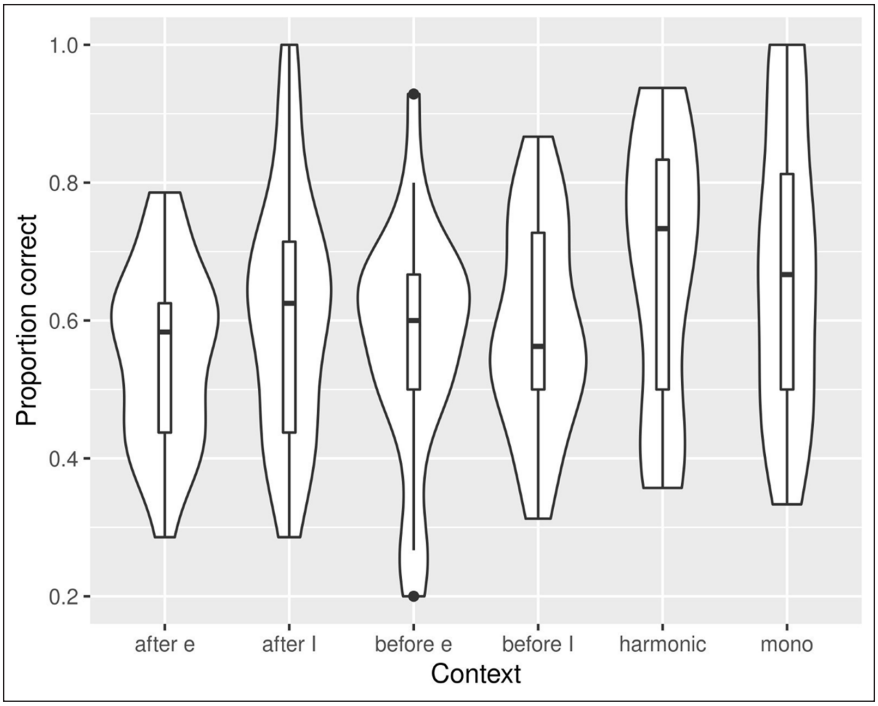

Figure 8: Proportion correct by Context for Experiment 2.

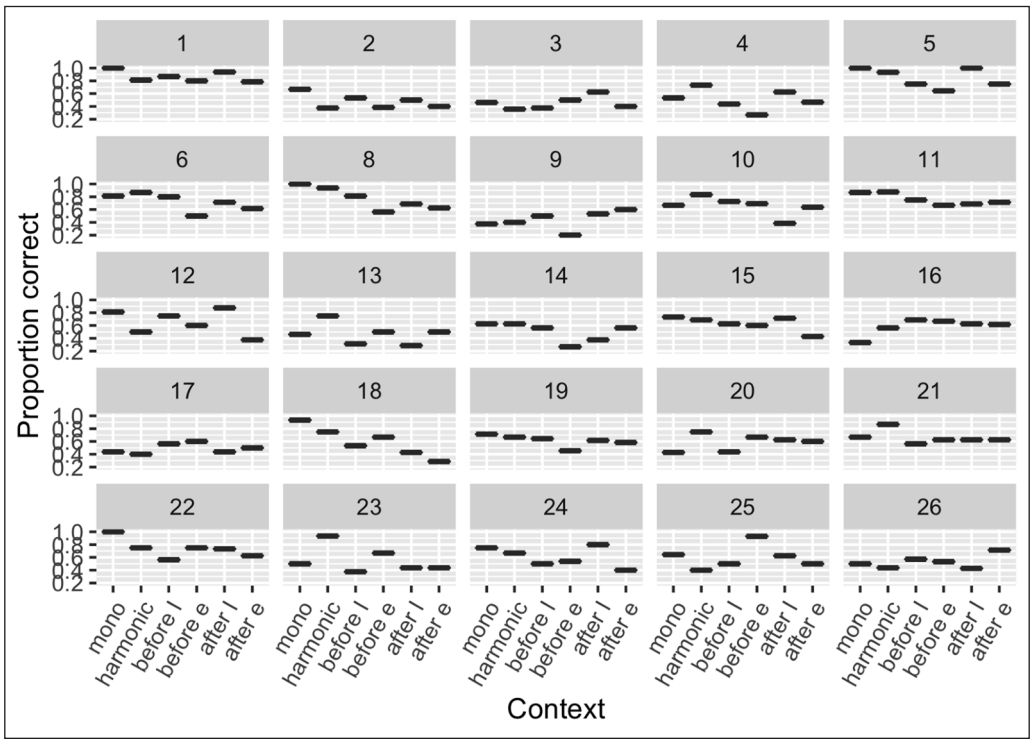

Figure 9: Proportion correct by Context for each participant.

majority one. We point it out because it suggests that individual variation may be an interesting direction to explore in future work.

To summarize, Experiment 2 showed that most of the disharmonic contexts have significantly lower accuracy than the harmonic and monosyllabic contexts, indicating that participants did worse at distinguishing disyllabic disharmonic stimuli than at distinguishing monosyllables or harmonic disyllables. These results are shown directly in the statistical comparisons in **Tables 13** (for the comparison to monosyllabic) and **14** (for the comparison to harmonic disyllables).

5. Discussion

In these experiments, we examined ATR perception, focusing on how harmony and disharmony affect the perception of different contrasts in Dàgáárè. In this section, we summarize the results, draw conclusions relative to our hypotheses, and discuss the implications for our understanding of the property ATR and perceptual motivations of harmony systems.

5.1. Summary of results

In terms of both contrast and context, only the "cross-height" (ɪ vs. e) factor was significant in Experiment 1. In other words, all trials involving [ɪ] versus [e] were more difficult than those with the other vowel pairs, with significantly lower accuracy rates. This result is expected, as the formant plot showed more overlap for [ɪ] versus [e] than for any other vowel pairs, and previous studies (e.g., Rose et al., 2023) have also found cross-height pairs (mid ATR with high RTR) difficult to differentiate in other languages. This fact motivated the choice of [ɪ] versus [e] as the contrast examined in Experiment 2.

In both experiments, all disyllabic contexts were harder for participants than monosyllabic contexts, in terms of lower accuracy rates, though in most (harmonic) cases not significantly so. This result may have simply been an effect of memory, as participants heard twice as many syllables in disyllabic than in monosyllabic trials. Specifically, participants may have had greater difficulty remembering what A and B were when it came time to listen to C, simply because they had already heard more syllables. It is worth noting that the inter-stimulus interval (500ms) was relatively long, meaning listeners needed to hold A in their memory for 1.5 seconds before responding. This long interval is likely to force participants to use phonological representations instead of auditory impressions, though we note that because participants hear competitors aloud in an ABX task, unlike in real-world language perception, it remains possible that they use auditory information in this task in a way that they do not in real-world phonological perception.

Notably, in contrast to Rose et al. (2023), accuracy rates were low across-the-board in these experiments, suggesting that these contrasts are generally difficult to distinguish even for (at least some) native speakers. The individual speaker graphs in **Figures 7** and **9** show that some

participants in our experiments did have very high accuracy, but there was a wide range in accuracy levels between participants, and on average the accuracy was low in our experiments. Even for the [e] versus [ɛ] contrast, which has no acoustic overlap according to **Figure 1**, participants in Experiment 1 showed less than 75% accuracy. We can therefore conclude from Experiment 1 that, even in cases without overlap, ATR is a difficult property to distinguish in Dàgáárè. Nonetheless, the comparison to the Rose et al. (2023) experiment should be made with caution, because of the differences in experiment design: That experiment was an AX task, while ours was ABX. Furthermore, their experiment was in-person before the pandemic, while ours was online during the pandemic. We return to this point later in the discussion.

With the caveat that memory is known to play a role in ABX tasks (Gerrits & Schouten, 2004), harmony does not appear to help perception of these difficult contrasts, given that performance was worse (though generally not significantly so) in harmonic disyllabic contexts than in monosyllabic ones. That said, disharmony does appear to hurt perception, with three of the four disharmonic contexts in Experiment 2 having significantly lower accuracy compared to monosyllables and harmonic disyllables.

Interestingly, the harmonic disyllables in Experiment 2 were [kɪkɪ] versus [keke], which was also the pair for the e/ɪ contrast in the e/ɪ context for Experiment 1. However, this pair had very different accuracy in the two experiments, as we can see in **Table 7** (for Experiment 1) versus **Table 10** (for Experiment 2). This pair had by far the lowest accuracy in Experiment 1, at 57%, whereas in Experiment 2, it had a much higher accuracy, at 67%, which is more comparable to the results from Experiment 1 for other harmonic pairs and even some of the monosyllabic pairs. We will return to this issue in the general discussion in Section 5.3.

5.2. Returning to the hypotheses

5.2.1. Experiment 1

Experiment 1 consisted of three vowel contrast types and four contexts. We predicted that e/ɛ contrasts/contexts would have the highest accuracy, followed by i/ɪ and then e/ɪ. For contexts, we predicted that the monosyllabic contexts would have the lowest accuracy, given fewer opportunities to hear the contrast, and that the disyllabic contexts would follow the same pattern as the contrasts.

In keeping with our predictions, we found that accuracy for both contrasts and contexts is highest for e/ɛ, followed by i/ɪ and then e/ɪ, but of those, only the difference between e/ɛ and e/ɪ was significant; the i/ɪ contrast and context were not significantly different from either of the other pairs. While the monosyllabic context did have slightly (non-significantly) lower accuracy compared to the e/ɛ context, it had higher accuracy than both other disyllabic harmonic contexts, significantly so for the e/ɪ context. This result is in direct contrast to our prediction

that disyllabic contexts, due to the presence of harmony, should have higher accuracy than monosyllabic contexts. It may have been due to memory considerations, with more load on participants' memory before making a decision for disyllables because they are longer. We will discuss more detail about possible explanations for these patterns in the general discussion in Section 5.3.

5.2.2. Experiment 2

Experiment 2 had monosyllabic contexts, harmonic disyllabic contexts, and four types of disharmonic disyllabic contexts, all with the vowel contrast that was most difficult for listeners in Experiment 1 (e/ɪ). We predicted that, compared to monosyllabic contexts, accuracy should be greater in harmonic contexts and lower in disharmonic contexts. We predicted possible lower accuracy in disharmonic forms in which [e] was the second vowel, due to the regressive, ATR-dominant underlying nature of the system.

In contrast to our predictions, but in line with the results of Experiment 1, the harmonic context had slightly (but not significantly) lower accuracy compared to the monosyllabic context. In line with the predictions, accuracy was lower in all disharmonic contexts than in the monosyllabic context, significantly so for three of the four disharmonic contexts. Further, while the difference was not significant, we did see a trend towards disharmonic forms involving a constant [e] (i.e., where both forms have [e] in V1 or both have [e] in V2) having lower accuracy compared to those involving constant [ɪ], and the one disharmonic context that was not significantly different from monosyllabic was one of the ones with constant [ɪ]. This is in line with our prediction. There was no significant directionality effect, again in line with our prediction; however, in terms of trends, accuracy was lower after [e] compared to before [e], but the reverse was true with [ɪ]. This trend is the opposite of what we predict based on a hypocorrection hypothesis given a regressive, ATR-dominant system.

5.3. General discussion

As noted in the introduction, we expected possible listener-based motivation for ATR harmony, namely that there should be some sort of perceptual motivation for harmony to occur. This was not apparent in Experiment 1, given that disyllabic harmonic forms did not improve accuracy over monosyllabic forms. In other words, having two opportunities to hear the ATR value (in harmonic disyllables) did not improve accuracy over having only one such opportunity (in monosyllables). Moreover, in Experiment 1, we were particularly interested in testing the hypothesis following Kaun (2004) that certain vowels are good triggers or targets based on perceptibility of the contrast. For example, since e~ɛ are the most perceptibly distinct, we would expect kɪkɛ vs. keke to be significantly better than kɪ vs. ke, due to the additional e~ɛ contrast

where the ATR distinction can be heard. Note that this issue does not directly require us to know which vowels are triggers or targets in the disyllabic forms, but simply relates to the fact that disyllabic forms provide two opportunities for hearing the ATR contrast, compared to a single opportunity in the monosyllabic forms. Additionally, e~ɛ is expected to be the easier place to hear the contrast given the acoustic distinctness of the vowel [ɛ], so the addition of this vowel into the disyllable adds to the expectation that kɪkɛ vs. keke would be easier than kɪ vs. ke. The results did not conform to this expectation.

Experiment 2, on the other hand, did provide some potential evidence for a perceptual advantage of harmony, showing that perception is significantly worse, and in some cases barely above chance, in most disharmonic contexts. There are three potential explanations for this result: (1) disharmony impedes the perception of ATR contrasts, (2) perception is more difficult for nonce words that are not possible words in the language, or (3) both vowels differ between A and B in the harmonic condition, but only one vowel does in the disharmonic conditions. This last option also connects to the potential of identity effects, since the harmonic disyllables in Experiment 2 had identical vowels in both syllables of each stimulus, while the disharmonic ones did not.

The second explanation is not plausible considering that disharmonic disyllabic forms are attested in compounds and sometimes in cliticization in Dàgáárè, as exemplified in **Table 4**. Since listeners are used to hearing such combinations in compounds, these nonce words would in fact be possible in the language, and so listeners should be able to adequately distinguish them. That said, we do not necessarily know how listeners are processing these disharmonic stimuli; it is possible that they have a processing cost due to not being possible as non-compound words in the language. As for the third possible interpretation, it still comes down to disharmony negatively affecting perception, as there are the same number of contrasts in the disharmonic conditions as in the monosyllabic one (i.e., one opportunity to hear [ɪ] versus [e]), yet performance is nonetheless worse in the disharmonic conditions. For example, when listeners hear [kɪke] versus [keke], they have the first syllable to hear the distinction between the two words, while the second syllable is the same in both words. This means that, in theory, they have the same amount of information to process the [ɪ] versus [e] distinction as in the forms [kɪ] versus [ke], yet they nonetheless perform worse on such forms. As such, the third possible explanation is effectively equivalent to the first, with the caveat that memory (disyllables being longer than monosyllables means more to hold in memory before a decision) and identity (harmonic disyllables in Experiment 2 had two identical syllables) may have an effect. Note that both of these points still point to a potential advantage of harmony and/or disadvantage of disharmony.

Thus, Experiment 2 suggests that disharmonic sequences, even though they exist in the language, make perception of difficult ATR/height contrasts even more difficult. This connects

to the idea that harmony is due to listeners misperceiving disharmonic sequences. However, they seem to misperceive such sequences primarily in a way that is inconsistent with the phonological patterns. Specifically, listeners have the hardest time with a pair that differs in both ATR and height. If listeners have a particular tendency to misperceive [ɪ] as [e] and vice versa in harmonic contexts, then it is unclear why [ɪ] harmonizes to [i] rather than to [e]. This suggests that phonologization of misperception, as discussed by Ohala (1994), may not be the right explanation for the link between perception and ATR harmony. For example, if listeners are more likely to mishear [ɪ…e] as [e…e] than as [i…e], then it is the former that we would expect to see as the result of harmony, if harmony is the result of hypocorrection. Instead, the relationship between perception and harmony seems to be modulated through avoidance of misperception, similar to the suggestion for rounding harmony by Kaun (2004). If [i] and [ɛ] are the maximally distinct vowels, then their inclusion in a form should allow greater differentiation; a language allowing just [ɪ…ɛ] and [i…e] will have fewer misperception issues, since ATR values can be heard on [ɛ] and [i] respectively. While performance on differentiating these forms was not significantly better than for monosyllables in Experiment 1, that is expected, given that the finding for Experiment 2 was that harmony does not help perception but disharmony hurts. Nonetheless, this idea does have some issues. First, [ɪ…ɪ] and [e…e] are allowed in the language and are difficult to differentiate, as found particularly in Experiment 1. This fact makes any perceptual explanation for harmony quite difficult to motivate. Additionally, some languages have been said to have merged high RTR vowels with mid ATR vowels but either kept them distinct for the purposes of harmony (e.g., Aziza, 2008 on Urhobo; Omamor, 1988 on Okpe) or allophonically continue to harmonize them (e.g., Andersen, 1999 on Mayak). If the difficulty in differentiating that pair motivates harmony, then we would expect such a merger to make harmony redundant, and so it is odd that it would so consistently remain part of the grammar after such mergers. However, as ATR is already known to be different across languages, it may be that such languages have different contrasts that are perceptually difficult. More research is needed on perception in other ATR systems.

Given these considerations about [ɪ…ɪ] and [e…e], it is worth returning here to the issue of the difference between this pair in Experiment 1 versus Experiment 2, where the accuracy rates in distinguishing [ɪ…ɪ] versus [e…e] were 57% and 67% respectively. It is unclear what caused this difference, but it is worth exploring further in the future to understand why these sounds are allowed in ATR harmony languages, despite being difficult to distinguish. One possible explanation for the difference is that this was the hardest pair in Experiment 1, where all stimuli were harmonic, but not in Experiment 2, which included disharmonic stimuli. Perhaps the relative difficulty of words to differentiate had an effect on the results. Alternatively, it may be because many participants completed both experiments, so the experience with this pair that

they had in Experiment 1 may have improved their ability to distinguish it in Experiment 2. Of course, in a real-world situation, semantic context may also help in differentiating real words with [ɪ…ɪ] and [e…e], so such considerations are also important in understanding the role of perception in the motivations for ATR harmony.

Returning to the issue of languages that are said to have merged high RTR vowels with mid ATR vowels (e.g., Aziza, 2008 on Urhobo; Omamor, 1988 on Okpe; Elugbe, 1983 on Edoid languages and proto-Edoid; Andersen, 1999 on Mayak), it is worth noting that, while our experiments are unable to provide any motivation for why these languages continue to show the ATR harmony systems they do after the merger, the experiments do potentially help us to understand why such mergers happen in the first place. As Casali (1995, p. 119) suggests, auditory similarity to and acoustic overlap with neighbouring vowels in the system could motivate the merging of high RTR vowels with mid ATR vowels. While the acoustic overlap is well-established and it is widely reported that [ɪ] and [e] are hard to distinguish, we noted in Section 1 that these perceptual comments are primarily anecdotal. Our study is among the first to confirm this fact experimentally with native speakers. The fact that it is common cross-linguistically for languages with ATR systems to merge these vowels could derive from the difficulty that even native speakers appear to have with their perceptual discrimination.

It is worth noting the standard orthography of most Ghanaian languages, including Dàgáárè, represents high RTR and mid ATR vowels the same way, as “e” and “o” for front and back respectively (cf. Bodomo, 1997). Thus, [ɪ] and [e] have the same spelling in Dàgáárè, which may affect native speaker perception and judgement. A plausible explanation for low accuracy in harmonic forms is the orthographic representation of the vowels [ɪ] and [e] with the letter “e”. By considering the phone-grapheme association of the letter “e”, the listeners might have perceived phonetically harmonic forms (e.g., [e]-[i]) as disharmonic (e.g., [ɪ]-[i]) and vice versa. However, participants were not shown any orthography, so it is unclear whether the spelling would have had any effect.

As mentioned in Section 2.2, there is very little existing baseline in the literature for understanding ATR perception and its possible role in ATR harmony systems. This study helps to provide a baseline about ATR perception, including which vowels are most difficult to distinguish for native speakers of a language with ATR, the difference between perception of ATR in monosyllables versus disyllables, and the effects of harmony and disharmony. While there remain many further questions to explore, as discussed in Section 5.5, this study makes an important contribution that can later be built on to answer some of the more nuanced questions of how perception might motivate ATR harmony. This is particularly true given the dearth of perception studies on ATR, and on African languages more generally, discussed in more detail in the next subsection.

5.4. Perception in African languages

Perception studies on African languages (on any property, not just ATR) are incredibly sparse. Based on a systematic literature review of literature since 2000, we are aware of a total of only eight published, peer-reviewed journal articles directly testing perception of African languages by native speakers in that time frame. At the word level, lexical tone perception has been investigated in Dinka (Remijsen, 2013), Yoruba (Harrison, 2000), Mambila (Connell, 2000), and Shona (Kadyamusuma, 2012), and there are also studies of tonal spreading in Bemba verb forms (Kula & Braun, 2015) and vowel length contrasts in Civili (Ndinga-Koumba-Binza & Roux, 2009). At the sentence level, perception of prosody and/or semantic focus has been explored in Akan (Genzel & Kügler, 2020) and Sepedi (Turco & Zerbian, 2021). In addition to the sound-related monolingual perception articles mentioned above, there have been a handful of studies on word perception by Twi (L1)-English (L2) bilinguals (e.g., Neumann & Nkrumah, 2019). There is also one very recently published perception study that has recently been conducted on African languages (Rose et al., 2023), and a small number of studies published more than two decades ago (e.g., Fulop et al., 1998 for Degema; La Velle, 1974, Harrison, 1996, and Hombert, 1976 for Yoruba; Bladon, Clark, & Mickey, 1987 for Shona). Of these studies, only Rose et al. (2023) and Fulop et al. (1998) are on ATR perception.

Given the dearth of research in perception on African languages, the present study represents a much-needed addition to the literature. Many more studies of this sort are necessary in order to further the representation of African languages in the literature and to thoroughly investigate the phonetic and phonological properties of these languages. In particular, our results are quite different from what Rose et al. (2023) found for ATR perception in Akan, emphasizing the need for more studies on more languages. Many further questions remain about the perception of ATR and the possible perceptual underpinnings of ATR harmony, which cannot be answered without further perceptual work on more languages.

5.5. Future directions

5.5.1. Perception in other languages with ATR harmony

This study is among the first on the perception of ATR harmony, and there remain many questions. As noted, it is unclear whether ATR is a consistent phonetic property across languages, as evidence shows that different languages, and even different speakers within a single language, may use different properties to phonetically realize ATR distinctions (Ladefoged & Maddieson, 1996; Fulop et al., 1998; Edmondson et al., 2007; Beltzung et al., 2015). As such, similar experiments should be repeated on other languages with different ATR systems, to see whether the results for Dàgáárè extend to other languages. There is a well-known distinction between the behaviour of ATR in West African versus East African languages (e.g., Rose, 2018), and perception may help

us to understand these typological observations. Additionally, recent work by Rose et al. (2023) found that accuracy by Akan listeners on ATR/height pairs was higher in disyllabic harmonic (identical) pairs than in monosyllables, though the results were from separate studies (ABX and AX tasks separately), and so cannot necessarily be directly compared. Future work should investigate whether the different findings between our study and the one by Rose et al. are due to task type differences or language differences.

Moreover, a major typological finding of ATR systems is that languages with an ATR contrast in high vowels (e.g. [i] versus [ɪ]) almost always have ATR harmony, and such languages without an ATR contrast in mid vowels almost always harmonize mid vowels allophonically anyways (Rose, 2018). One well-known example is Kinande (Archangeli & Pulleyblank, 2002). The Dàgáárè results are somewhat puzzling from this typological perspective: if the high RTR versus mid ATR contrast (e.g. [ɪ] versus [e]) is the most difficult to perceive, why would a language without ATR mid vowels create them just to harmonize? This question needs further exploration in future research, particularly exploring whether the perceptual space is different in languages like Kinande, which lack phonemic ATR mid vowels, compared to a language like Dàgáárè, where those vowels are phonemic. With reference to Kinande, it seems that adding allophonic ATR mid vowels would further impede perception, as those vowels are difficult to differentiate from some of the phonemic ones, the RTR high vowels. Having fewer vowels in the vowel space should aid perception, and thus from a listener-based perspective on harmony, it is odd that allophonic harmony of this nature would so consistently occur cross-linguistically. This fact suggests that something else must be going on to explain why allophonic harmony would occur. One possibility is that in such vowel systems, it is the high vowels themselves that are too perceptually close, such that adding a harmonic equivalent of the mid vowel could aid perception rather than hinder it. Similar perception studies should be conducted on languages with such vowel systems, such as Kinande, to test this theory. Moreover, given the known differences between West and East African ATR systems, perception studies on any East African language would be a valuable addition to this literature.

5.5.2. Additional exploration of the results in Dàgáárè

There are a few additional explorations on perception of ATR in Dàgáárè that could be done in the future with other tasks or additional stimuli. First, future studies should examine the potential memory effects in an ABX task, in terms of the fact that ABX competitors are spoken aloud (unlike in real-world perception) and need to be kept in memory, but also in terms of the question of which type of stimuli are easier to remember. It is possible that harmony provides a memory advantage that means that harmonic disyllables are approximately equivalent on the ABX task to monosyllables, even when disharmonic disyllables show worse performance. We do

not have the information from this experiment to ask whether harmony modulates the effect of stimulus length, but it is worth examining in the future.

Further research should also conduct an experiment similar to Experiment 2, but where not all harmonic tokens have two identical vowels, in order to disentangle potential identity effects from the effects of harmony and disharmony. Adding additional disyllabic tokens, both harmonic and disharmonic, with other vowels would aid in separating competing hypotheses. Moreover, the higher accuracy for the pair [ɪ…ɪ] vs. [e…e] in Experiment 2 versus Experiment 1 should be explored further with variations on Experiment 2. This could be done by having more participants in the current version of Experiment 2 who did not complete Experiment 1, to test the idea that experience from Experiment 1 boosted accuracy on this pair in Experiment 2, or by including other harmonic stimuli in Experiment 2, to test the hypothesis that [ɪ…ɪ] vs. [e…e] being one of the easier pairs to distinguish in Experiment 2 caused the higher accuracy.

Additional research on Dàgáárè may want to have additional stimuli and additional participants to directly compare the dialects. We briefly checked the performance of our participants against their dialect, and there did not appear to be a clear pattern of performance based on dialect; some speakers of a given dialect did poorly and others did very well. However, there were too few participants to make any strong generalizations on this issue. It would be particularly valuable to recruit more participants of each dialect, particularly Central Dàgáárè and Dagara, given that Central Dàgáárè allows disharmonic compounds in both orders, but Dagara harmonizes RTR + ATR ones. This comparison was not possible given the small number of Dagara participants in this study, but for future work, it may allow us to disentangle potential effects related to participants' exposure to disharmony in their native language.

Finally, our experiments did not directly address the question of how the identity of one vowel in an item influences the perception of the other. In our experiment, what "contrast" and "context" essentially mean is just the identity of V1 and V2 respectively (where context is monosyllabic when there is no V2). Moreover, the model performed better without the interaction between contrast and context. Future research should address this question of vowel interactions, in order to look in greater depth at perception of harmony and of potential triggers and targets of harmony. Such an experiment could be done for instance with a continuum between two vowels, with the continuum put into different contexts and participants asked which vowel they hear for each step on the continuum. A study of this nature would greatly enrich our understanding of potential perceptual motivations for ATR harmony systems, but is left to future research.

5.5.3. Methodological notes

We end the discussion with a methodological note, namely that the study was conducted online due to COVID-19, while the limited previous ATR perception studies have been conducted in person. While web-based perception studies have been found to generally be comparable to

lab-based studies in terms of participant population representativeness (Woods et al., 2015) and overall task performance (Germine, Nakayama, Duchaine, Chabris, Chatterjee, & Wilmer, 2012), this is only insofar as proper controls are utilized to reduce unwanted variability (namely in stimulus presentation and participant attentiveness). Thus, it is possible that the methodological differences contributed to the lower accuracy rates seen in this study compared to previous ones. For example, although we did implement certain checks to see if participants were following instructions (e.g., directly asking them to wear headphones and providing a sound check such that they could adjust their device volume to a comfortable level, in order for them to hear the stimuli as clearly as possible), it is possible that these were not totally effective at enforcing compliance, which may have reduced our overall accuracy rates. Some additional controls that could have been utilized are attention check trials (very easy trials interspersed throughout normal experimental trials; correct response rates below 100% indicate a participant is likely either not paying attention or does not understand the instructions, and is thereby just responding randomly) and a more stringent headphone screener (e.g., having participants identify whether a set of tones are playing in their left or right ear, which would be difficult to answer correctly if they are listening through external speakers).[5] Future post-pandemic work should consider comparing participants from the same languages in online studies versus in-person ones, to see whether the substantial difference in accuracy rates between the Rose et al. (2023) study and ours is due to differences between the languages or differences between the methodologies.

Another aspect of our methodology to consider in regards to the results is the nature of the ABX task itself. First, it is particularly cognitively demanding, as the participant must hold three items simultaneously in memory for multiple seconds in order to successfully compare them. Depending on the length of the inter-stimulus intervals and stimuli themselves,[6] by the time participants are expected to provide a response, the auditory trace of the stimuli may have already left their phonological loop, and only the general phonological categories they had assigned to items A and B may remain in their working memory, giving an impression of a high degree of categorical perception (Gerrits & Schouten, 2004). Furthermore, participants have been shown to be biased towards stating that B is X in these studies (Schouten et al., 2003). Other similar paradigms like the AXB task and AX task also have their own problems, however. In AXB tasks, participants may be biased to report that A is X (Van Hessen & Schouten, 1999). As for AX tasks, while their working memory load is reduced since participants must only discriminate

[5] We thank an anonymous reviewer for their suggestions in regard to improving the experimental controls.

[6] A reviewer suggested that certain stimuli may be easier to store in memory based on their phonological structure in one of two possible ways: vowels sharing a specification for ATR might be easier to remember, or there may be a distinctiveness effect such that items with more phonological specifications are easier to remember. Our experiment suggests that, at least for ATR in Dàgáárè, the latter is likely not the case, as participants did significantly worse on most disharmonic trials. However, the possibility that sharing an ATR specification might help with memory, and that this memory effect could contribute to the motivations for ATR harmony, is worth exploring in future research.

between two items, participants can still be predisposed to report in extreme manners depending on how they interpret the task instructions: in some instances, they may report that the two items are "different" only if they are clearly (to them) across rather than within phoneme boundaries, again giving an overly strong impression of categorical perception (Repp, 1981; Gerrits and Schouten, 2004; cf. Han, 2009); in other cases, participants may be primed to hyper-focus on minute, irrelevant acoustic differences between stimuli, and thus would only report two literally identical tokens as being the "same" sound (Stevenson, 2015; Choi & Tsui, 2022). While some perception studies (e.g., Liu, Chen, & Kager, 2017) have had participants complete multiple discrimination tasks and compared their performance on each task, we opted to not do this in order to not make the experiment overly long and tedious to complete, particularly because we knew participants would likely be completing the experiment on their phones and possibly with spotty internet connectivity. However, future perception studies may still want to make use of different (or even multiple) tasks to get a fuller picture of this issue.

6. Conclusion

To conclude, this study has investigated the perception by native speakers of the property ATR in Dàgáárè. We found that ATR is a difficult property to differentiate on average for the native speakers in our study, even for vowel pairs with no acoustic overlap, and that disharmony impedes the perception of ATR contrasts. We argue that this study provides a listener-based motivation for the existence of ATR harmony, namely that harmony as a phonological pattern could be motivated by the disadvantage that disharmony creates for listeners in correctly interpreting the sounds they hear. That said, unlike what Rose et al. (2023) found for Akan, we found no perceptual advantage of harmonic disyllables compared to monosyllables, and so at least in Dàgáárè, it appears that any perceptual reason for harmony comes solely from the disadvantages of disharmony.

Acknowledgements

We would like to thank Alexander Angsongna for his help with the Dàgáárè language, Phil Monahan for helpful discussion about experimental design, the audience at the 2022 Annual Conference on African Linguistics for feedback on a presentation on our initial results, and the reviewers and editors of this paper for their insightful comments. This paper draws on research supported by the Social Sciences and Humanities Research Council of Canada.

Competing interests

The authors have no competing interests to declare.

References

Ali, M., Grimm, S., & Bodomo, A. (2021). *A dictionary and grammatical sketch of Dagaare* (African Language Grammars and Dictionaries 4). Berlin, Germany: Language Science Press. DOI: https://doi.org/10.5281/zenodo.5154710

Andersen, T. (1999). Vowel harmony and vowel alternation in Mayak. *Studies in African Linguistics*, *28*(1), 1–29. DOI: https://doi.org/10.32473/sal.v28i1.107377

Angsongna, A. (2023). *Aspects of the morphophonology of Dagaare* (Doctoral dissertation, University of British Columbia, Vancouver, Canada). DOI: https://doi.org/10.14288/1.0427267

Angsongna, A., & Akinbo, S. (2022). Dàgáárè (Central). *Journal of the International Phonetic Association*, *52*(2), 341–367. DOI: https://doi.org/10.1017/S0025100320000225

Aralova, N. (2015). *Vowel harmony in two Even dialects: Production and perception* (Doctoral dissertation. University of Amsterdam, Amsterdam, The Netherlands). Retrieved from https://hdl.handle.net/11245/1.484816

Archangeli, D., & Pulleyblank, D. (1994). *Grounded phonology*. Cambridge, MA, USA: MIT Press. DOI: https://doi.org/10.2307/416805

Archangeli, D., & Pulleyblank, D. (2002). Kinande vowel harmony: domains, grounded conditions and one-sided alignment. *Phonology*, *19*(2), 139–188. DOI: https://doi.org/10.1017/S095267570200430X

Aziza, R. O. (2008). Neutralization of contrast in the vowel system of Urhobo. *Studies in African Linguistics*, *37*(1), 1–19. DOI: https://doi.org/10.32473/sal.v37i1.107297

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. DOI: https://doi.org/10.1016/j.jml.2007.12.005

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Beltzung, J.-M., Patin, C., & Clements, G. N. (2015). The feature [ATR]. In A. Rialland, R. Ridouane & H. van der Hulst (Eds.), *Features in phonology and phonetics: Posthumous writings by Nick Clements and coauthors* (pp. 217–246). Berlin, Germany: De Gruyter Mouton. DOI: https://doi.org/10.1515/9783110399981-011

Bladon, A., Clark, C., & Mickey, K. (1987). Production and perception of sibilant fricatives: Shona data. *Journal of the International Phonetic Association*, *17*(1), 39–65. DOI: https://doi.org/10.1017/S0025100300003248

Bodomo, A. (1997). *The structure of Dagaare*. Stanford, CA, USA: CSLI Publications.

Bodomo, A. (2017). Mabia: its etymological genesis, geographical spread and some salient genetic features. Paper read at the Mabia Languages Conferences in Winneba, Ghana and Vienna, Austria.

Bodomo, A. (2020). Mabia: its etymological genesis, geographical spread and some salient genetic features. In A. Bodomo, H. Abubakari & S. A. Issah (Eds.), *Handbook of the Mabia Languages of West Africa* (pp. 5–34). Glienicke, Germany: Galda Verlag.

Boersma, P., & Weenink, D. (2022). *Praat: doing phonetics by computer* (Version 6.2.10) [Computer software]. http://www.praat.org/

Busà, M., & Ohala, J. (1999). In search of the perceptual correlates of vowel harmony. In *Proceedings of the XIVth International Congress of Phonetic Sciences* (Vol. 1, pp. 357–360). San Francisco, CA, USA.

Casali, R. F. (1995). On the reduction of vowel systems in Volta-Congo. *African Languages and cultures*, *8*(2), 109–121. DOI: https://doi.org/10.1080/09544169508717790

Casali, R. F. (2003). [ATR] value asymmetries and underlying vowel inventory structure in Niger-Congo and Nilo-Saharan. *Linguistic Typology*, *7*(3), 307–382. DOI: https://doi.org/10.1515/lity.2003.018

Choi, W., & Tsui, R. K. Y. (2022). Perceptual integrality of foreign segmental and tonal information: Dimensional transfer hypothesis. *Studies in Second Language Acquisition*, 1–18. DOI: https://doi.org/10.1017/S0272263122000511

Connell, B. (2000). The perception of lexical tone in Mambila. *Language and Speech*, *43*(2), 163–182. DOI: https://doi.org/10.1177/00238309000430020201

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12. DOI: https://doi.org/10.3758/s13428-014-0458-y

Edmondson, J. A., Padayodi C. M., Hassan, Z. M., & Esling, J. H. (2007). The laryngeal articulator: Source and resonator. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, (Vol. 16, pp. 2065–2068). Saarbrücken, Germany: Universitat des Saarlandes.

Elugbe, B. O. (1983). The vowels of proto-Edoid. *Journal of West African Languages*, *13*(1), 79–89. Retrieved from https://journalofwestafricanlanguages.org/index.php/downloads/summary/36-volume1301/149-the-vowels-of-proto-edoid

Esling, J. H. (2014). The articulatory function of the larynx and the origins of speech. In K. Carpenter, O. David, F. Lionnet, C. Sheil, T. Stark & V. Wauters (Eds.), *Proceedings of the 38th Annual Meeting of the Berkeley Linguistics Society* (Vol. 38, pp. 121–149). Berkley, CA, USA: Berkeley Linguistics Society. DOI: https://doi.org/10.3765/bls.v38i0.3325

Finley, S. (2012). Typological asymmetries in round vowel harmony: Support from artificial grammar learning. *Language and Cognitive Processes*, *27*(10), 1550–1562. DOI: https://doi.org/10.1080/01690965.2012.660168

Fulop, S. A., Kari, E., & Ladefoged, P. (1998). An acoustic study of the tongue root contrast in Degema vowels. *Phonetica*, *55*(1–2), 80–98. DOI: https://doi.org/10.1159/000028425

Gansonre, C., Højlund, A., Leminen, A., Bailey, C., & Shtyrov, Y. (2018). Task-free auditory EEG paradigm for probing multiple levels of speech processing in the brain. *Psychophysiology*, *55*(11), e13216. DOI: https://doi.org/10.1111/psyp.13216

Genzel, S., & Kügler, F. (2020). Production and perception of question prosody in Akan. *Journal of the International Phonetic Association*, *50*(1), 61–92. DOI: https://doi.org/10.1017/S0025100318000191

Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review*, *19*, 847–857. DOI: https://doi.org/10.3758/s13423-012-0296-9

Gerrits, E., & Schouten, M. E. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, *66*(3), 363–376. DOI: https://doi.org/10.3758/BF03194885

Gick, B., Wilson, I., Koch, K., & Cook, C. (2004). Language-specific articulatory settings: Evidence from inter-utterance rest position. *Phonetica*, *61*, 220–233. DOI: https://doi.org/10.1159/000084159

Grosvald, M. (2009). Interspeaker variation in the extent and perception of long-distance vowel-to-vowel coarticulation. *Journal of Phonetics*, *37*(2), 173–188. DOI: https://doi.org/10.1016/j.wocn.2009.01.002

Han, J. I. (2009). Task effects on the perception of phonological contrasts by Korean monolingual adults. *언어 [Language]*, *34*(3), 737–759. DOI: https://doi.org/10.18855/lisoko.2009.34.3.013

Harrison, P. (1996). An experiment with tone. *UCL Working Papers in Linguistics*, *8*, 575–593.

Harrison, P. (2000). Acquiring the phonology of lexical tone in infancy. *Lingua*, *110*(8), 581–616. DOI: https://doi.org/10.1016/S0024-3841(00)00003-6

Hombert, J. M. (1976). Perception of tones of bisyllabic nouns in Yoruba. *Studies in African Linguistics*, *7* (Suppl. 6), 109–121.

Hyman, L. M. (2002). Is there a right-to-left bias in vowel harmony? In *9th International Phonology Meeting* (Vol. 1). Vienna, Austria.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. DOI: https://doi.org/10.1016/j.jml.2007.11.007

Junge, C., Boumeester, M., Mills, D. L., Paul, M., & Cosper, S. H. (2021). Development of the N400 for word learning in the first 2 years of life: a systematic review. *Frontiers in Psychology*, *12*, Article 689534. DOI: https://doi.org/10.3389/fpsyg.2021.689534

Kadyamusuma, M. R. (2012). Effect of linguistic experience on the discrimination of Shona lexical tone. *Southern African Linguistics and Applied Language Studies*, *30*(4), 469–485. DOI: https://doi.org/10.2989/16073614.2012.750821

Kaun, A. R. (1995). *The typology of rounding harmony: An optimality theoretic approach* (Doctoral dissertation, University of California, CA, USA). DOI: https://doi.org/10.7282/T3R49PM2

Kaun, A. R. (2004). The typology of rounding harmony. In B. Hayes, R. Kirchner & D. Steriade (Eds.), *Phonetically based phonology* (pp. 87–116). Cambridge, MA, USA: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511486401.004

Kimper, W. A. (2011). Competing triggers: Transparency and opacity in vowel harmony (Doctoral dissertation, University of Massachusetts, MA). Retrieved from https://scholarworks.umass.edu/dissertations/AAI3482711

Kimper, W. A. (2017). Not crazy after all these years? Perceptual grounding for long-distance vowel harmony. *Laboratory Phonology*, *8*(1), 19. DOI: https://doi.org/10.5334/labphon.47

Kula, N. C., & Braun, B. (2015). Mental representation of tonal spreading in Bemba: Evidence from elicited production and perception. *Southern African Linguistics and Applied Language Studies*, *33*(3), 307–323. DOI: https://doi.org/10.2989/16073614.2015.1108768

Kuroda, S.-Y. (1967). *Yawelmani phonology*. Cambridge, MA, USA: MIT Press.

Kuubezelle, N. (2013). *An autosegmental analysis of phonological processes in Dagara* (Doctoral dissertation, University of Ghana, Ghana). Retrieved from http://197.255.68.203/handle/123456789/5360

Kuubezelle, N., & Akanlig-Pare, G. (2017). Dagara tongue-root vowel harmony. *Ghana Journal of Linguistics*, *6*(2), 1–14. DOI: https://doi.org/10.4314/gjl.v6i2.1

La Velle, C. R. (1974). An experimental study of Yoruba tone. In I. Maddieson (Ed.), *UCLA Working Papers in Phonetics 27: Tone project* (pp. 160–170). Los Angeles, CA: University of California, Los Angeles. Retrieved from https://escholarship.org/uc/item/3st6f4rg

Ladefoged, P. (1968). *A phonetic study of West African languages*. Cambridge, MA, USA: Cambridge University Press.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford, England: Blackwell Publishers Ltd.

Lindau, M. (1975). [Features] for vowels. *UCLA Working Papers in Phonetics*, *30*, 16–21. Retrieved from https://escholarship.org/uc/item/7gv6z0vq

Lindau, M. (1978). Vowel features. *Language*, *54*(3), 541–563. DOI: https://doi.org/10.2307/412786

Liu, L., Chen, A., & Kager, R. (2017). Perception of tones in Mandarin and Dutch adult listeners. *Language and Linguistics*, *18*(4), 622–646. DOI: https://doi.org/10.1075/lali.18.4.03liu

Lloy, A., Akinbo, S., Angsonga, A., & Pulleyblank, D. (2019). *The Dagaare [Dàgáárè] low vowels: A tenth vowel in the phonological inventory of an understudied language*. Poster presentation at the Multidisciplinary Undergraduate Research Conference, University of British Columbia, Vancouver, Canada.

Mailhot, F., & Reiss, C. (2007). Computing long-distance dependencies in vowel harmony. *Biolinguistics*, *1*, 28–48. DOI: https://doi.org/10.5964/bioling.8587

Ndinga-Koumba-Binza, H. S., & Roux, J. C. (2009). Perceived duration in vowel-length based Civili minimal pairs. *South African Journal of African Languages*, *29*(2), 216–226. DOI: https://doi.org/10.1080/02572117.2009.10587330

Neumann, E., & Nkrumah, I. K. (2019). Reversal of typical processing dynamics in positive and negative priming using a non-dominant to dominant cross-language lexical manipulation, *Memory*, *27*(6), 829–840. DOI: https://doi.org/10.1080/09658211.2019.1573902

Nevins, A. (2010). Locality in vowel harmony. Cambridge, MA, USA: MIT Press. Retrieved from https://www.jstor.org/stable/41475377. DOI: https://doi.org/10.7551/mitpress/9780262140973.001.0001

Ohala, J. J. (1994). Towards a universal, phonetically-based, theory of vowel harmony. In *Proceedings of the 3rd international conference on spoken language processing* (Vol. 2, pp. 491–494). Yokohama, Japan: Acoustical Society of Japan. DOI: https://doi.org/10.21437/ICSLP.1994-113

Omamor, A. P. (1988). Okpẹ and Uvwiẹ: A case of vowel harmony galore. *Journal of West African Languages*, *181*, 47–64.

Ozburn, A., Akinbo, S., Angsongna, A., Schellenberg, M., & Pulleyblank, D. (2018). Dagaare [a] is not neutral to ATR harmony. *The Journal of the Acoustical Society of America*, *144*(3), 1938–1938. DOI: https://doi.org/10.1121/1.5068478

Przezdziecki, M. A. (2005). *Vowel harmony and coarticulation in three dialects of Yoruba: phonetics determining phonology* (Doctoral dissertation, University of Ithaca, Ithaca, NY, USA). DOI: https://doi.org/10.5281/zenodo.3967375

Pulleyblank, D. (1996). Neutral vowels in Optimality Theory: a comparison of Yoruba and Wolof. *Canadian Journal of Linguistics*, *41*(4), 295–347. DOI: https://doi.org/10.1017/S0008413100016601

R Code Team. (2022). *R: a language and environment for statistical computing* [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Repp, B. H. (1981). Two strategies in fricative discrimination. *Perception & Psychophysics*, *30*(3), 217–227. DOI: https://doi.org/10.3758/BF03214276

Remijsen, B. (2013). Tonal alignment is contrastive in falling contours in Dinka. *Language*, *89*(2), 297–327. DOI: https://doi.org/10.1353/lan.2013.0023

Ringen, C. O., & Vago, R. M. (1998). Hungarian vowel harmony in Optimality Theory. *Phonology*, *15*(3), 393–416. Retrieved from https://www.jstor.org/stable/4420136. DOI: https://doi.org/10.1017/S0952675799003632

Rose, S. (2018). ATR vowel harmony: New patterns and diagnostics. In G. Gallagher, M. Gouskova & S. H. Yin (Eds.), *Proceedings of the 2017 Annual Meeting on Phonology*. Washington, DC, USA: Linguistic Society of America. DOI: https://doi.org/10.3765/amp.v5i0.4254

Rose, S., Obiri-Yeboah, M., & Creel, S. (2023). Perception of ATR contrasts by Akan speakers: a case of perceptual near-merger, *Laboratory Phonology*, *14*(1). DOI: https://doi.org/10.16995/labphon.8948

RStudio Team. (2022). *RStudio: integrated development for R* [Computer software]. Boston, MA, USA: RStudio, PBC. https://www.rstudio.com/

Schouten, B., Gerrits, E., & Van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication, 41*(1), 71–80. DOI: https://doi.org/10.1016/S0167-6393(02)00094-8

Snider, K. L. (1984). Vowel harmony and the consonant l in Chumburung. *Studies in African Linguistics, 15*(1), 47–57. DOI: https://doi.org/10.32473/sal.v15i1.107521

Somé, P. (1982). Systématique du signifiant en Dagara: variété Wúlé [System of meaning in Dagara: Wúlé variety]. Paris, France: L'Harmattan.

Starwalt, C. (2008). *The acoustic correlates of ATR harmony in seven- and nine-vowel African languages: a phonetic inquiry into phonological structure* (Doctoral dissertation, Unviersity of Texas Arlington, Arlington, TX, USA). Retrieved from http://hdl.handle.net/10106/1015

Stevenson, S. D. (2015). The strength of segmental contrasts: a study on Laurentian French (Doctoral dissertation, University of Ottawa, Ottawa, Canada). University of Ottawa. DOI: https://doi.org/10.20381/ruor-2860

Szeredi, D. (2016). Exceptionality in vowel harmony (Doctoral dissertation, New York University, New York City, NY, USA). Retrieved from http://ling.auf.net/lingbuzz/003148

Tuomainen, J. (2001). *Language specific cues to segmentation of spoken words in Finnish: Behavioral and event-related brain potential studies* (Doctoral dissertation, de Katholieke Universiteit Brabant, Tilburg, The Netherlands).

Turco, G., & Zerbian, S. (2021). Processing of prosody and semantics in Sepedi and L2 English. *Journal of Psycholinguistic Research, 50*(3), 681–706. DOI: https://doi.org/10.1007/s10936-020-09746-z

Van Hessen, A. J., & Schouten, M. E. H. (1999). Categorical perception as a function of stimulus quality. *Phonetica, 56*(1–2), 56–72. DOI: https://doi.org/10.1159/000028441

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software, 40*(1), 1–29. Retrieved from https://www.jstatsoft.org/v40/i01/. DOI: https://doi.org/10.18637/jss.v040.i01

Wickham, H. (2016). *ggplot2: elegant graphics for data analysis* [Computer software]. New York City, NY, USA: Springer-Verlag New York. https://ggplot2.tidyverse.org

Woods, A. T., Velasco, C., Levitan, C. A., Wan, X., & Spence, C. (2015). Conducting perception research over the internet: a tutorial review. *PeerJ, 3*, e1058. DOI: https://doi.org/10.7717/peerj.1058