Open Library of Humanities

Voice quality and tone as independent dimensions of contrast in Dinka

Bert Remijsen, University of Edinburgh, b.remijsen@ed.ac.uk

Mirella L. Blum, Dartmouth College, mirella.l.blum@dartmouth.edu

Umberto Noè, University of Edinburgh, umberto.noe@ed.ac.uk

Languages with independently contrastive Voice Quality and Tone are rare, and the evidence base on them is limited. As a result, key hypotheses on how their phonetic realizations interact with one another and with the vowel system remain tentative. Against the background of this evidence base, it is worthwhile to conduct a production study in which Voice Quality, Tone, and Vowel Quality are orthogonally crossed. This paper presents the results of a comprehensive acoustic study on this configuration in Dinka (Nilo-Saharan, South Sudan). The study is based on 29 four-member minimal sets for Voice Quality (Modal vs. Breathy) and Tone (Low vs. High), across all seven of the Dinka vowels (/i, e, ɛ, a, ɔ, o, u/). These materials were elicited from eight speakers of the Bor South variety of Dinka. The results indicate that Voice Quality, Tone, and Vowel Quality each have their own primary correlate: phonation, F0, and formants, respectively. In addition, each distinction influences other phonetic parameters to a lesser extent. Importantly, the Voice Quality contrast is realized saliently on vowels in Low- and-High-toned syllables alike, and across the vocalic domain.

1 Introduction

1.1 Motivation

Voice quality (or phonation) in speech can be conceived of as a continuum of degree of constriction at the larynx, with phonetic creaky voice being more constricted than phonetic modal voice, which is itself more constricted than phonetic breathy voice (Gordon & Ladefoged 2001).[1] The baseline factor determining the degree of constriction is the speaker-specific range of phonation—that is, a speaker has a particular voice quality to begin with, which may be more or less constricted relative to other people. Within this speaker-specific phonetic range, phonation may also play a role in phonological contrasts. One way phonation may be part of a phonological contrast is as a secondary correlate of a tone contrast. That is, fundamental frequency (F0) and phonation are often part of the realisation of a single dimension of contrast, whereby phonation serves as a secondary correlate to a tonal or intonational category. For example, creaky phonation is part of the phonetic realisation of Mandarin Tone 3 (Belotel-Grenié & Grenié 1997). In Black Miao, the Low tone is realised with creaky phonation, and the Mid tone with breathy phonation (Kuang 2013). And in English, creaky phonation is associated with low intonational boundaries in some varieties (Redi & Shattuck-Hufnagel 2001). Such links are not surprising, given their articulatory connection—F0 and phonation both result from laryngeal settings. Given this connection, it is remarkable that in some languages, Voice Quality and Tone represent independent dimensions of contrast.[2] This set of languages includes Mpi (Maddieson & Ladefoged 1996), Jalapa Mazatec (Garellek & Keating 2011)—and Dinka, the language under investigation in the current study (Andersen 1987; Remijsen & Manyang 2009). To illustrate this configuration, consider the minimal-set example in **Table 1**, which is drawn from the Bor South dialect of Dinka. As seen from this example, there are four distinctive tone categories, and each of them appears with Modal voice and with Breathy voice. We encourage the reader to ascertain the differences in melody and voice quality for themselves by listening to the associated sound examples.

The central question, in relation to independently contrastive Voice Quality and Tone, is how these two contrasts–both of them laryngeal in nature–can be realized simultaneously on the same vocalic domain. We will address this question with respect to Dinka. Prominent in this debate are a number of studies on Jalapa Mazatec, an Oto-Manguean language that presents a

---

[1] In an articulatory sense, constriction is not just a matter of the constriction of the vocal folds, but additionally involves five other valves in the pharyngeal cavity (Esling & Harris 2005). Crucially, these are not active in an independent manner, but rather "operate together in hierarchical fashion" (Edmondson & Esling 2006:187).

[2] The terms voice quality, phonation, modal voice and breathy voice can be used both in a phonetic sense and to refer to a phonological contrast. To disambiguate between these uses, we write concepts with capitals whenever we are referring to them in a phonological sense. We do the same with the other phonological factors under investigation in this study, i.e. Tone, Vowel Quality, and Vowel Length.

Voice Tone	Modal voice		Breathy voice	
Low	cìiir 'big.river'	▶	c̤ìiir 'thorny.tree'	▶
High	cíiir 'pour:NOM'	▶	c̤íiir 'pour:BNF:NOM'	▶
Fall	cîiir 'pour:FUG:3SG'	▶	c̤îiir 'big.river:PL'	▶
Rise	cǐiir 'plot'	▶	c̤ǐiir 'star:PL'	▶

Table 1: A minimal set for Voice Quality and Tone in Dinka. In the sound examples, these words are embedded in an utterance in which they follow the conjunction *kṳ̀* 'and.'

three-way Voice Quality contrast between Modal, Breathy and Laryngealized[3] voices–in addition to a Tone contrast (Silverman 1997; Garellek & Keating 2011). These studies have yielded two important findings. First, the phonetic realization of Jalapa Mazatec's two non-modal levels of Voice Quality (Breathy voice and Laryngealized voice) is significantly more salient on Low-toned vowels than on Mid-toned and High-toned vowels (Garellek & Keating 2011). Second, the non-modal levels of Voice Quality are realized more saliently in the first half of the vowel than in the second half. Silverman (1997) argues that the latter finding can be explained on the basis of the former: on the hypothesis that non-modal levels of Voice Quality are difficult to realize on non-Low tones, languages may evolve to realize Voice Quality and Tone on different parts of the vowel. This hypothesis, hereafter the Sequencing Hypothesis, can be seen as an instantiation of enhancement in the sense of Keyser & Stevens (2006), i.e., the idea that speech systems may evolve to enhance the phonetic difference between contrastive elements. According to the Sequencing Hypothesis, systems with independently contrastive Voice Quality and Tone may respond to the difficulty in producing them simultaneously by realizing them in sequence: the Voice Quality contrast in the first half of the vowel, and the Tone contrast in the second half. However, it could be that, instead, the Voice Quality contrast of Jalapa Mazatec is realized early in the vowel simply because it has its diachronic origin in laryngeal coarticulations in onset consonants: both aspiration and laryngealization have been reconstructed as consonantal phenomena in Proto Mazatec (Kirk 1966: 38).

Importantly, it is also not clear that the Sequencing Hypothesis holds across languages with independently contrastive Voice Quality and Tone. In the Sino-Tibetan language Mpi, which has six categories of Tone in addition to a Voice Quality contrast, there is no evidence that the realisations of the Voice Quality and Tone contrasts are ordered sequentially within the vocalic domain (Ladefoged & Maddieson 1997). Hence, it is difficult to determine whether the findings observed in relation to Jalapa Mazatec represent universal tendencies, or instead result

---

[3] Garellek & Keating (2011) reserve the term creaky voice to refer to a constricted phonation that is more narrowly defined, involving low F0, cf. the notion of 'prototypical creaky voice' in Keating et al. (2015). They use the term laryngealized voice as an overarching term to refer to constricted phonations, irrespective of F0.

from language-specific diachronic circumstances. One reason is that the number of languages that present independently contrastive Voice Quality and Tone is small. A second reason is the available data are limited. In relation to Mpi, for example, the evidence (Ladefoged & Maddieson 1997) consists of just a few examples coupled with auditory impressions. We will come back to the quality of the evidence base below.

With respect to the configuration of independent Voice Quality and Tone contrasts, it is worthwhile to note that the primary phonetic correlates of these phonological phenomena, i.e., phonation and F0, exert some influence on each other. The phonetic phenomenon of 'creaky voice' is in fact not a single phonation, but rather an umbrella term encompassing several specific phonations involving laryngeal constriction, each with their own properties. These phonations include vocal fry, multiply-pulsed voice, and tense voice (Keating et al. 2015). F0 plays a constraining role in determining which specific phonations can be realized. For example, when creaky voice is associated with vowels that have low F0, the vowel is often found to display stretches of aperiodicity. In contrast, aperiodicity is not part of the phonetic realization of creaky voice on vowels with high F0, a configuration that has been referred to as tense voice. This allows for the scenario whereby, in a language with independently contrastive Voice Quality and Tone, a category of Creaky voice is saliently realized on Low-toned and High-toned vowels alike, but through different types of creaky phonation as a function of Tone.

Irrespective of whether a language that has a Voice Quality contrast can also have an independent Tone contrast, we can expect Voice Quality to affect F0 and vowel height. Of particular note here are 'register' systems, in which phonation, F0, and vowel height are correlates of a single dimension of contrast. In such languages, the combination of more constricted phonation, higher F0, and more open vowel height realizes a phonological category that is in contrast with a category marked by less constricted phonation, lower F0, and more closed vowel height (Brunelle 2012). While some languages may have only two of these three correlates, the patterning together of the values is consistent. Moreover, the package of correlates invariably includes phonation as a salient correlate, which is what distinguishes a register contrast from a tone contrast (DiCanio 2009: 162). Examples include Cham (Brunelle 2012), Chong (Thongkum 1988; DiCanio 2009), Sylheti (Gope 2021), Tamang (Mazaudon & Michaud 2008), and Yi (Kuang & Cui 2018). The correlation between phonation and F0 may be due to a shared diachronic origin: the loss of a voicing contrast in the syllable onset (Mazaudon & Michaud 2008). There may also be a physiological explanation for this correlation: constricting/spreading of the vocal folds (which affects phonation type) alters stiffness (which affects f0). As for the correlation between phonation and vowel height, here there may also be a physiological explanation, in the sense that vowel height affects the position of the larynx, and larynx lowering is one of the articulatory mechanisms involved in controlling laryngeal

constriction (Gordon & Ladefoged 2001; Edmondson & Esling 2006). More closed vowel height in Breathy voice than in Modal voice has been reported in several languages, including Jalapa Mazatec (Garellek & Keating 2011), Chong (Thongkum 1988; DiCanio 2009), and Kedang (Samely 1991), as well as in Dinka (Malou 1988).

Aside from phonation contrasts influencing phonetic vowel quality, vowel quality contrasts also determine phonation. In an acoustic analysis of languages with a phonation contrast, Esposito, Sleeper & Schafer (2019: 374) report that, in seven of the eight languages under investigation, "vowels, regardless of phonemic phonation type, with higher formant frequencies were creakier, while vowels with lower formant frequencies were breathier." This finding underscores the importance to include phonemic vowel quality as a factor in investigations into languages with a phonation contrast. The Dinka Voice Quality contrast developed diachronically out of an Advanced Tongue Root (ATR) contrast, which is still attested in most West Nilotic languages (Denning 1989; Dimmendaal & Jakobi 2020): Proto West Nilotic is hypothesized to have had five -ATR vowels /ɪ,ɛ,a,ɔ,ʊ/ and five +ATR vowels /i,e,ʌ,o,u/ (Storch 2005). Given that ATR is realized primarily through phonetic vowel quality, it is worthwhile to consider to what extent phonetic vowel quality plays a role in the phonetic realization of the Voice Quality contrast.

A final consideration motivating this study relates to the evidence base regarding languages with independently contrastive Voice Quality and Tone; these include include Jalapa Mazatec (Ladefoged et al. 1993), Ju|'hoansi (Miller 2007) and Mpi (Ladefoged & Maddieson 1996). Data on how the contrasts interact *phonetically* in these languages is limited. For example, with regards to Jalapa Mazatec, the materials on which the phonetic analyses are based—(Ladefoged et al. 1993)—are not minimal sets, and the recordings from different speakers do not represent independent observations; speakers were recorded together and immediately consecutively (Steriade, personal communication). In the case of Mpi, the acoustic evidence is limited to only a few examples (Ladefoged & Maddieson 1996). In the case of Ju|'hoansi, the quantitative analysis in Miller (2007) finds significant differences in phonation only for half of the speakers in the sample. It is not in question whether languages like Jalapa Mazatec, Mpi, and Ju|'hoansi have independent phonation and tone contrasts; rather, the phonetic data on such systems is limited. This is also true in relation to Dinka—we will come back to this in Section 1.3. Against the background of this evidence base, it is worthwhile to conduct a production study in which Voice Quality, Tone, and Vowel Quality are orthogonally crossed.

The paper is structured as follows. In the remainder of this introduction, we introduce the sound system of Dinka (Section 1.2), survey earlier work on its Voice Quality contrast (Section 1.3), and summarize the goals of this investigation (Section 1.4). We then detail our methodology (Section 2) and present the results (Section 3). The results are discussed further in Section 4, and the paper is concluded in Section 5.

1.2 Language background on Dinka

Within the Dinka language, four dialect clusters are commonly distinguished—Rek, Bor, Padang and Agar— each of which includes several varieties (Tucker & Bryan 1956; Roettger & Roettger 1989).

Dinka is known for the complexity of its suprasegmental phonology. Tone—Dinka dialects have three or four tone categories—Voice Quality on vowels (Breathy, Modal), and Vowel Length (Short, Long, Overlong) are all independently contrastive (Andersen 1987; Remijsen & Manyang 2009). As for the segmentals, Dinka has seven vowels: /i, e, ɛ, a, ɔ, o, u/. There are two limitations on the crossing of these dimensions of contrast: /u/ is invariably Breathy (Andersen 1987; Remijsen & Manyang 2009),[4] and /ɛ/ and /a/ are not in contrast on Short vowels (Andersen 1987). Most of the language's morphology is expressed via the suprasegmentals, as can be seen from the morphological alternations in **Table 1**. Dinka also has twenty consonants. Plosive consonants are found at five places of articulation: bilabial (/p, b, m/), dental (/t̪, d̪, n̪/), alveolar (/t, d, n/), palatal (/c, ɟ, ɲ/), and velar (/k, g, ŋ/). In addition there are the liquids (/l, r/) and the semivowels (/ɰ, j, w/). The Bor South dialect, with which this paper is concerned, displays all of the above phonological elements. Its tonal inventory includes four tonemes: Low (L, v̀); High (H, v́); Fall (HL, v̂); and Rise (LH, v̌). The transcriptions in this article invariably represent the surface phonological form. Given that Bor South presents several contextual tone processes (Remijsen 2013; Blum 2021), this means that the transcription of words in the sentences in which the target words are elicited may not be identical to the underlying specification.

Our transcriptions broadly follow the IPA, with one exception: Vowel Length is represented by one (Short), two (Long) or three (Overlong) vowel symbols. Tone is represented on the first vowel, and applies to the entire syllable; Voice Quality is also represented on the first vowel character, and applies to the vowel as a whole. With respect to the phonetic realization of the Voice Quality contrast, speakers vary with regards to their speaker-specific voice quality (cf. Garellek 2019: 79–80), so it is possible for a Dinka speaker's *phonologically* Breathy vowels not to be particularly breathy in a phonetic sense. However, because the phonation contrast in Dinka is independently contrastive, Breathy vowels can be expected to be less constricted than the same speaker's Modal vowels. The between-speaker variation in the realization of the Voice Quality contrast is illustrated by the sound clips associated with the examples in this paper, which include minimal sets for Voice Quality by six different speakers.

---

[4] Malou (1988) is the only study that analyzes /u/ as Modal. This analysis may be driven by applied linguistic considerations: /u/ does not contrast for Voice Quality, and hence it can remain unmarked for Breathy Voice in the Dinka orthography. We will present evidence in support of the interpretation that Dinka /u/ is Breathy in Section 3.1.2.

1.3 Earlier work

The independently contrastive nature of Dinka's Voice Quality and Tone contrasts is well established. Most studies postulate a binary voice quality contrast on vowels, between Modal[5] and Breathy, as outlined in Section 1.2 and illustrated in **Table 1**. These include Tucker (1936; 1975); Ayom (1980); Andersen (1987; 1992–1994; 2002, among others); Malou (1988); Denning (1989[Chapter3]);[6] Duerksen (1994); Gilley (2003); Remijsen & Manyang (2009); Remijsen (2013); Blum (2021). This generalization includes studies on all four of the dialect clusters, and several analyses by native-speaker linguists. In addition, there is an alternative analysis of voice quality in Dinka, which postulates a further two voice qualities in addition to Modal and Breathy voice. The first of these has been referenced alternatively as Hard, Creaky, or Harsh; the second as Hollow or Faucalized. This analysis postulating four contrastive voice qualities was first presented in two unpublished conference presentations (Hall, Ayom & Hall 1982; Ayom, Hall & Odden 1985), and subsequently in Ayom (1987), Denning (1989[Chapter4]), Edmondson & Esling (2006), and Kenstowicz (2014). However, none of the published studies provide minimal-set evidence for a four-way voice quality contrast. This is surprising given the high functional load of voice quality alongside vowel length, tone, and vowel quality in Dinka.

In the earliest remaining record of the alternative analysis, Ayom (1987: 171) describes them as follows: "As well as non-breathy voice […] versus breathy voice […] quality contrasts, there are two additional voice qualities in Dinka: hard voice (v̰) and hollow voice (v̰̈), which, linguistically have a rather low functional yield and are only used to show a verbal process indicating direction of the action verb toward or away from the speaker."[7] Morphological marking for movement away from the deictic center is known as *centrifugal* or *itive*, and for movement towards the deictic center as *centripetal* or *ventive*. In (1), we show one of the examples for Hard (1a) and Hollow (1b) voice from Ayom (1987), and alongside our own transcription of the same sentences. In contrast to Ayom, we postulate that the centrifugal and centripetal derivations are not marked by additional Voice Qualities; we postulate Modal voice and Breathy voice where Ayom postulates Hard and Hollow voice, respectively. Relative to the base inflectional paradigm, we postulate that the centrifugal derivation is marked by increased Vowel Length, and

---

[5] Andersen (1987 and subsequent publications) transcribes the contrast as one of Creaky vs. Breathy rather than Modal vs. Breathy.

[6] We qualify Denning's (1989) position with respect to the chapter, as he advocates different analysis in different chapters of his dissertation. He analyzes Dinka as having two voice qualities in Chapter 3 and as having four voice qualities in Chapter 4.

[7] While Ayom hypothesizes these additional voice qualities in relation to Dinka in general in this quote, both Denning 1989[Chapter4] and Edmondson & Esling (2006) postulate them specifically in relation to the variety spoken around the city of Bor, i.e., Bor South, the dialect under investigation in the present study. In this context, it is worthwhile to note that Ayom was almost certainly himself a speaker of the Bor South dialect, as indicated by the word form ḓ̀ùk 'boy' (most Dinka dialects have ḓ̀òk), and by the fact that most of his research deals specifically with Bor Dinka.

the centripetal derivation by increased Vowel Length and Breathy Voice. As we explain further in this section, inherently Breathy roots do not change in Voice Quality in the derivation of the centripetal; Modal roots, however, shift to Breathy (cf. Andersen 1992–94).

(1)   **Ayom (1987)**

a. **dhùk amị́:t yúɛn wéi**   ▶

'The boy is pulling the rope (away from speaker).'

Our transcription

ḓùk   à-míiit   jwɛ́ɛn   wěj

boy   DEC-pull:FUG   rope   away

b. **dhùk amị̀:t yúɛn**   ▶

'The boy is pulling the rope (towards speaker).'

Our transcription

ḓùk   ǎ-mị̀iit   jwɛ́ɛn

boy   DEC-pull:PET   rope

Both analyses attest to the fact that the centripetal and centrifugal are morphologically marked in a fusional manner, as is characteristic of Dinka. All native transitive verbs consist of a single closed syllable, which appears in dozens of distinct morphological forms, diverging in vowel length, tone, voice quality and vowel quality. At issue here, then, is the morphological exponence of the centrifugal and centripetal forms. It is now well-established that all dialects of Dinka have a ternary vowel length contrast (Andersen 1987; Remijsen & Gilley 2008; Remijsen 2013; 2014). In a detailed description of the derivational morphology of Dinka transitive verbs in the Agar dialect, Andersen (1992–1994) describes how lengthening by one level of vowel length distinguishes the two spatial derivations from the base paradigm, in a regular manner across the system of transitive verbs. In Andersen's analysis, the spatial derivations both differ from the base paradigm in terms of increased vowel length. The exponence of the centripetal derivation additionally involves Breathy voice, often in conjunction with vowel raising. Our own investigations on the Bor South dialect concur with Andersen's analysis. Illustration (2) shows Bor South transitive verbs with a short Modal (2a), long Modal (2b), and long Breathy (2c) stem vowel. All items increase by one level of vowel length in the spatial derivations; the Modal roots become Breathy in the centripetal derivation, and the Breathy root (2c) retains its Breathy voice, and here the contrast between centrifugal and centripetal is marked solely through tone.

(2)   a. **ǎ-dòl** 'DEC-roll.up'   ▶ **à-dóol** 'DEC-roll.up:FUG'   ▶ **ǎ-dṳ̀ul** 'DEC-roll.up:PET'   ▶

      b. **ǎ-mìit** 'DEC-pull'   ▶ **à-míiit** 'DEC-pull:FUG'   ▶ **ǎ-mị̀iit** 'DEC-pull:PET'   ▶

      c. **ǎ-ḓèeṱ** 'DEC-transport'   ▶ **ǎ-ḓěeṱ** 'DEC-transport:FUG'   ▶ **ǎ-ḓèeṱ** 'DEC-transport:PET'   ▶

In support of an analysis that does not postulate additional contrasts of Voice Quality for the centrifugal and centripetal derivations, illustration (3) presents comparisons with other forms

in the morphology of transitive verbs that have the same specifications for voice quality, vowel length and tone according to the analysis of Andersen (1992–1994). Example (3a) includes the centrifugal verb form from (1a) above; in (3b) we offer the comparison with the infinitive form of the base paradigm of the same verb, which has the same specifications for voice quality, tone, and vowel length. According to Ayom's analysis, the two forms differ in voice quality, with the centrifugal forms having Hard voice. Crucially, there is no audible difference in phonation, in line with the two-voice quality analysis.

(3) a. **dèeŋ à-míiit   tìim**   ▶
         Deng DEC-pull:FUG tree
         'Deng is pulling the wood (away).'

     b. **dèeŋ ĕe   tìim míiit**   ▶
         Deng DEC:PST tree pull:NF
         'Deng has pulled the wood.'

Then we consider voice quality in the centripetal derivation. Example (4a) shows the 3rd singular form of the transitive verb 'carry (on shoulder)', and (4b) presents the same inflection in the centripetal derivation. Because this verb is lexically Breathy, the change to Breathy deriving the centripetal derivation should therefore apply vacuously in (4b) according to Andersen's analysis. According to Ayom (1987), in contrast, the centripetal should be Hollow, differentiating this derivation from the base paradigm for lexical roots with Modal and Breathy vowels alike. Here as well, we cannot perceive a difference in phonation: both sound breathy. This again supports the two-voice quality analysis. The reader is invited to make their own assessment by listening to the associated sound clips.

(4) a. **dèe̤ŋ ǎ-kè̤eet   é̤mà̤nè̤**   ▶
         Deng DEC-carry:3SG ADV
         'They are carrying Deng now.'

     b. **dèeŋ ǎ-kè̤eet   tì̤ːn**   ▶
         Deng DEC-carry:PET:3SG ADV
         'They are carrying Deng this way.'

More recently, Edmondson & Esling (2006) adopt the four Voice Quality analysis, interpreting Dinka as a register tone system, in which phonation and F0 represent phonetic correlates of a single dimension of contrast. According to this register tone analysis, each of the voice qualities combines with a specific melodic shape, e.g. low F0 combines with Breathy Voice and high F0 with Faucalized/Hollow Voice (Edmondson & Esling 2006:188). This interpretation is contradicted by independent role of Tone and Voice Quality in phonological contrasts in Dinka, illustrated in **Table 1**. In the present study, we will examine the role of F0 in the realization of the Voice quality contrast.

Acoustic evidence for the four-voice-qualities analysis has been presented in Denning (1989) and in Edmondson and Esling (2006). The energy distribution profiles of the spectra for the four voice qualities displayed by Denning (1989) do not fit with what is known about the relation between perceived voice quality and spectral tilt.[8] As for Edmondson & Esling, the key laryngeal difference between Breathy and Faucalized is supposed to be glottal fold tension (Edmondson and Esling 2006:188). But the fibroscopic images do not show a clear difference.

Overall, we find that the hypothesis that some dialects of Dinka have four phonological voice qualities is not adequately supported by evidence in the studies that postulate it. Therefore, it is difficult, if not impossible, to test it; in spite of having investigated Dinka phonology and phonetics for two decades ourselves, we have not come across any phenomena that warrant postulating levels of Voice Quality beyond Modal vs. Breathy. Hence, Voice Quality represents a binary factor in our study, alongside Tone and Vowel Quality. As we will see in Section 3, this matches with the acoustic evidence.

1.4 Goals

Even though phonation and F0 result from the same organ, the larynx, they can represent orthogonal distinctions in human languages. Because phonetic evidence on this configuration is limited, our primary goal is to examine the realisation of this constellation of phonological contrasts. We are doing so through a production study, on the basis of an optimally-structured dataset, in which Voice Quality, Tone, and Vowel Quality are orthogonally crossed.

The results will enable us to determine whether several patterns observed and hypotheses postulated in the study of other languages with independently contrastive Voice Quality and Tone hold up in Dinka as well. Against the background of Silverman (1997), it will be interesting to find out whether Dinka realizes its independent voice quality and tone contrasts sequentially to optimize phonetic realization. Moreover, is Breathy phonation in Dinka realized more saliently on Low-toned syllables than on High-toned syllables? And, with reference to Brunelle (2012), do phonation, F0 and vowel height correlate in the marking of Dinka Voice Quality in the same way as in a register system? It will be worthwhile to find out whether these patterns are borne out in an optimally-controlled study.

2 Methods

In this section we lay out our methodology. We describe the materials (Section 2.1), the speakers from whom the materials were elicited (Section 2.2), the elicitation procedure (Section 2.3)

---

[8] In footnote 52 (Denning 1989:133), in the [tem] set, the hypothesized 'Hollow' voice has the least downward tilt. This does not make sense because Hollow voice would be a lax phonation alongside Breathy voice. And in the [cɔl] set, the Modal voice form has the steepest downward slope. Again this does not make sense, as Modal voice should certainly have less of a downward slope than Breathy voice.

and finally the way the recordings were processed (Section 2.4) and analysed (Section 2.5). The dataset is publicly available as an electronic resource (Remijsen, Blum & Pen de Ngong 2022).

2.1 Materials

The materials consist of minimal sets in which Voice Quality (Breathy, Modal) and Tone (Low, High) are crossed. The crossing of these factors yields four segmentally identical forms for each minimal set. Apart from Voice Quality and Tone, the third phonological factor that is experimentally manipulated in the dataset is Vowel Quality: all seven of the Dinka vowels /i,e,ɛ,a,ɔ,o,u/ are included in the dataset. The crossing of these three factors is illustrated in **Table 2**, which presents one minimal set for Voice Quality and Tone for each of the seven vowel phonemes. The minimal sets for the vowel /u/ have just two members, because this vowel does not contrast for Voice Quality – it is analysed as Breathy (Section 1.2); this is why the minimal set for /u/ in **Table 2** contrasts for Tone (Low, High) only. For each of the vowels /i,e,ɛ,a,ɔ,u/, the dataset includes four minimal sets, including the example sets in **Table 2**; and for the vowel /o/, there are five minimal sets in total.[9]

	Modal		Breathy	
	Low	**High**	**Low**	**High**
/i/	mìiit 'pull:3SG' ▶	míiit 'pull:NF' ▶	m̤ìiit 'rainbow' ▶	m̤íiit 'pull:BNF:NOM' ▶
/e/	dèeem 'tie:3SG' ▶	déeem 'tie:NF' ▶	d̤èeem 'tie:PET:3SG' ▶	d̤éeem 'tie:BNF:NOM' ▶
/ɛ/	rèɛɛr 'split:NF' ▶	réɛɛr 'arid.land' ▶	r̤èɛɛr 'go.berserk:NF' ▶	r̤éɛɛr 'sit' ▶
/a/	làaak 'wash:1SG' ▶	láaak 'downstream' ▶	l̤àaak 'belt' ▶	l̤áaak 'saliva:GEN' ▶
/ɔ/	ɥɔ̀ɔɔc 'buy:3SG' ▶	ɥɔ́ɔɔc 'buy:NF' ▶	ɥ̤ɔ̀ɔɔc 'drag:PET:1SG' ▶	ɥ̤ɔ́ɔɔc 'carry:NF' ▶
/o/	tòook 'light:3SG' ▶	tóook 'light:NF' ▶	t̤òook 'crack:3SG' ▶	t̤óook 'crack:NF' ▶
/u/			b̤ùuut 'soak:3SG' ▶	b̤úuut 'soak:NF' ▶

Table 2: Example minimal sets for Voice Quality and Tone from the dataset, one for each of Dinka's seven vowels.

The minimal sets in **Table 2** show that the four combinations of Voice Quality and Tone may distinguish lexical meaning, as in the /laaak/ set, where each combination of Voice Quality and Tone involves a different lexical root: 'wash', 'downstream', 'belt' and 'saliva.' Alternatively, the combinations can express morphological meaning, as in the case of the /deeem/ set, where all four forms are drawn from the morphological paradigm of the verb 'tie'. In most cases, however, the sets represent a combination of lexical and morphological contrast, as in the /toook/ set, which involves two inflections for each of two lexical roots.

---

[9] Originally, the dataset included a fifth minimal set for the vowel /e/. In the course of data collection, we realised that most participants were unfamiliar with one of the lexical-morphological forms involved, and for this reason we stopped eliciting it.

Each member of each minimal set appears in the dataset in two utterance-final contexts: at the end of a sentence, and in isolation, i.e., as a one-word utterance. In this position – at the end of a declarative utterance – no intonational boundary tone is added to the lexical-morphological specification (Remijsen 2013). This is illustrated in **Table 3**. The composition of the sentence is not fixed, but instead chosen for the target word form to fit grammatically and felicitously in a semantic sense. The complete set of materials is included in the archived dataset, which is publicly available (Remijsen et al. 2022).

Voice quality, Tone	Sentence		Isolation	
Modal, Low	**dèeŋ à-bǐ màac jǫ̀ok kṳ̀ tòook** Deng DEC.SG-FUT fire find:NF CONJ light:3SG 'Deng will find the fire and then he will light it.'	▶	**tòook** light:3SG	▶
Modal, High	**dèeŋ à-bǐ màac tóook** Deng DEC.SG-FUT fire light:NF 'Deng will light the fire.'	▶	**tóook** light:NF	▶
Breathy, Low	**dèeŋ à-bǐ ròooŋ jǫ̀ok kṳ̀ tǫ̀ook** Deng DEC.SG-FUT nutshell find:NF CONJ crack:3SG 'Deng will find the nutshell and then he will crack it.'	▶	**tǫ̀ook** crack:3SG	▶
Breathy, High	**dèeŋ à-bǐ ròooŋ tǫ́ook** Deng DEC.SG-FUT nutshell crack:NF 'Deng will crack the nutshell.'	▶	**tǫ́ook** crack:NF	▶

Table 3: An example of the sentence and isolation forms, for each of the four members of the /toook/ set.

The total number of types per speaker is 216. For the vowels /i,e,ɛ,a,ɔ/, we have four sets for each of four combinations of Tone and Voice quality (5*4*4 = 80); for the vowel /o/, there are five sets for the same four combinations (5*4 = 20); for /u/ we have four sets for two levels of Tone, as Voice quality is not contrastive for this vowel (4*2 = 8). Hence the total number of target words is 108. These were elicited in two contexts (isolation, sentence), yielding 216 types. These were collected from each of the 8 speakers (Section 2.2), resulting in an intended total of 1728.

2.2 Speakers

The participants in our study come from a single variety within one of four Dinka dialect clusters: the Bor South variety within the Bor cluster. Within the Bor cluster, Bor South is the southernmost variety, and it is spoken in and around the Bor municipality, on the east side of the White Nile.

The choice to focus on this variety is motivated by the earlier work on voice quality in Dinka; Ayom (1987), Denning (1989), and Edmondson & Esling (2006) all worked on Bor dialect data. Within Bor South, the speech community itself makes a further division between two groups: Athööc and Gok. We recorded four speakers from Athööc, and four from Gok. We did not find any differences as a function of this division; for this reason, it is not a factor in the analysis. The speakers range in age from 24 to 56, with an average age of 38. They are all native speakers of Dinka. While they were living in Nairobi, Kenya at the time of the recordings, they continued to use Dinka on a day-to-day basis with family and friends.

2.3 Recording procedure

The speakers were recorded individually in Nairobi using a solid-state recorder (Marantz PMD 661) and a headset-mounted directional microphone (Shure SM10A), at a sampling frequency of 48kHz and a bit depth of 16. The recording location was either a professional studio or a quiet room. In relation to five of the speakers, all those involved in the session were in Nairobi. In the case of the other three speakers, the elicitation happened remotely from Edinburgh through a video call. An assistant who is himself a native-speaker of Dinka was invariably present with the speakers in Nairobi.

For each speaker, we started out with a practice session, in which we went through the complete set of materials, so as to enable the speaker to fully familiarize themselves with the task and the materials. If this went well,[10] we made the actual recordings on a different day. Because of the size of the dataset, the practice session typically took three to four hours, and as a result it would stretch over one or two sessions. The actual recording session took between two and three hours, and the forms were elicited as follows.

The data were elicited through an interview rather than as a reading task, because many of our target words are homographs in their orthographic form, as the Dinka orthography does not distinguish Tone at all and does not distinguish Vowel Length accurately. In each case, we started with asking for a translation of the target word from English, working our way towards the intended sentence construction. If the speaker offered a synonym, we asked the speaker to express the same meaning using a different word, or we explained the crucial semantic difference in English. If that was unsuccessful, the Dinka-speaking assistant described the intended meaning in Dinka. As a last resort, we would show the speaker the orthographic representation of the word. Importantly, the target word was never uttered as an example for the speaker to copy in the recording session. After the speaker had uttered the target word in the sentence context, we asked them to repeat by itself the final word in the utterance they had just produced (cf. **Table 3**).

---

[10] In the case of six people, we started with the practice session, but did not move on to the actual recording. This happened when the speaker was not familiar with several of the target words.

This is how we recorded the target word in isolation. We included two or more breaks, depending on how long the elicitation session lasted.

2.4 Data processing

We processed the data using Praat (Boersma & Weenink 2021). In a first step, the realizations of the target words in sentence context and in isolation were all stored as separate sound files. If there was more than one realization, then all were stored, distinguished by a repetition number. Segmentation was carried out in a subsequent phase, and at this point only a single repetition was retained, both for the isolation context and for the sentence context. In case there was more than one repetition, the one with the greatest intensity at the midpoint of the vowel of the target word was retained. We segmented the vocalic nucleus of the target word, on the basis of acoustic landmarks of changes in constriction in the vocal tract (Turk et al. 2006). Differing from Turk et al., however, we excluded the burst phase of onset plosives from the interval segmented for the vowel.

As noted above, the total number of intended realizations is 1728 (216 for each of 8 speakers). Following segmentation, we had only 1700; there were 28 missing target words. These are divided fairly evenly across the dataset, with no strong weighting to a particular speaker or a particular vowel.[11]

Ahead of the acoustic analyses, the traces for fundamental frequency (F0) and the first two formants were generated and checked individually in Praat, before using them in subsequent acoustic analyses. In the checks, we paid particular attention to the values over the vowel of the target word. In relation to F1 and F2, errors were corrected by rerunning the analysis with reference values appropriate to the perceived vowel quality. This was done in Praat using the Track function that is available for Formant objects (Boersma & Weenink 2021). In this function, the detection of F1 and F2 is informed by reference values, which are set by default at the center of the typical range of the formant in question. In our attempt to correct problematic cases, we would set it to fit with the typical value of the perceived vowel, e.g. 300 (F1) and 800 (F2) in the case of a vowel that sounds like [u]. This has the effect of making candidates that are closer to this reference value more attractive to the formant detecting algorithm.

In relation to F0, tracking errors were corrected by selecting the correct candidate in the Pitch object. In a small number of cases, no F0 measurement could be made because the vowel was aperiodic, resulting in missing values. All of these cases involve Speaker 1, and all but

---

[11] Analysing the gaps, the number of missing target words varies between speakers from zero to eight (out of 216) per speaker. With respect to the minimal sets, the number of missing values varies between zero and four (out of 64, unless the vowel is /u/, in which case this is out of 32). And considering the combinations of speakers and vowels, any gaps are invariably limited to one of the four—or, in the case of /o/, five—minimals sets. That is, all speakers have, for any of the seven vowels, all combinations of tone, phonation and context for at least three minimal sets. Some gaps were due to a speaker not knowing a word form; in other cases, we failed to elicit a target word by accident.

one involve the combination of Modal voice with Low tone. Speaker 1 realises Modal Voice with aperiodicity or with irregular periodicity on Low-toned vowels in about half of the cases.[12] In most cases this is limited to the final third of the vowel and the coda, as in the example in **Figure 1**. In terms of Keating et al.'s (2015) overview of phonation types associated with creaky voice, these allophonic variants can be characterized as prototypical creaky voice (with irregular periodicity) and aperiodic voice (without periodicity). Elsewhere, in other realisations by speaker 1, and in the material from the other seven speakers, the Modal voice category can be characterized as vocal fry or modal voice on Low-toned vowels, and as pressed voice or modal voice on High-toned vowels (cf. Keating et al. 2015).
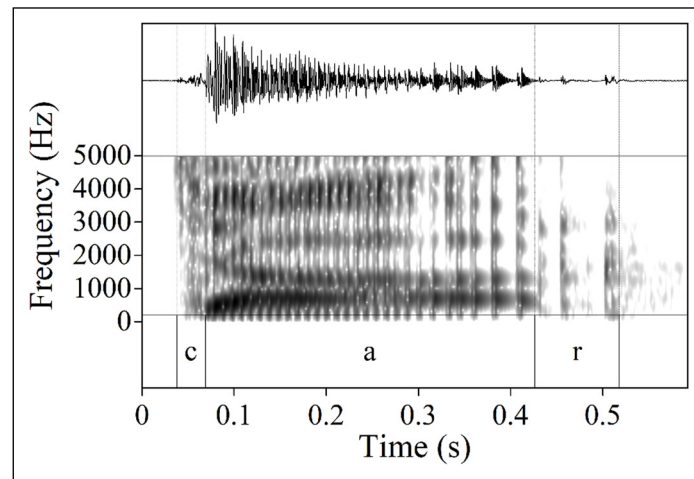

Figure 1: Waveform and spectrogram of an isolation-form realization of the word /càaar/ 'examine:1SG', as produced by Speaker 1.

2.5 Analysis

The acoustic measurements were made in Praat using scripts, on the basis of the checked formant and F0 traces. The primary package used was PraatSauce (Kirby 2018), a package to extract a wide range of measurements on phonation, formants, F0 and duration. PraatSauce is an implementation in Praat of the VoiceSauce package (Shue et al. 2011), and it is based on the architecture of Mills' (2010) suite of scripts for spectral measurements. We used PraatSauce to extract the following measurements on energy distribution: H1*-H2*, H1*-A1*, H1*-A2*, H1*-A3*, where the asterisks refer to the fact that the measurement of the phonetic parameter has been corrected for the effects of formant values, using the formula of Iseli et al. (2007)., and also to extract the following harmonics-to-noise measurements: CPP, HNR05, HNR15, HNR25,

---

[12] Specifically, aperiodicity resulted in 4, 3, and 27 missing values at 25%, 50%, and 75% of the vowel's duration. Given that there are 50 types with a Modal-voiced Low-toned vowel (25 items × 2 contexts), this means that in about half of these, Speaker 1 has aperiodicity, in the majority of cases limited to the final third of the vowel's duration.

HNR35. Of all of these measurements, H1*-H2* is typically the most successful measurement in distinguishing between modal and non-modal phonation in earlier studies (Gordon & Ladefoged 2001; Garellek 2019). To these we added Spectral Emphasis, a measure of energy distribution developed by Traunmüller & Eriksson (2000) to measure vocal effort. The addition of Spectral Emphasis to the standard set of acoustic measurements is motivated by the fact that it performed better than H1*-H2* in an acoustic investigation of the Advanced Tongue Root contrast in Shilluk (Remijsen et al. 2011). Shilluk's Advanced Tongue Root contrast involves a perceptible difference in phonation, which appears to be detected well by the Spectral Emphasis measurement. It is calculated by low-pass filtering the signal at 1.5 times the F0, and then subtracting the intensity value of the resulting signal from the intensity of the original signal. As a result, the Spectral Emphasis measure has higher values if there is more high-frequency intensity in the signal and lower values if there less high-frequency intensity; this is the opposite of better-established measures such as H1*-H2* and H1*-A3*. Spectral Emphasis is a measure of relative intensity akin to the above-mentioned energy distribution measures.

These measurements were subjected first to descriptive and then to inferential statistical analyses, with the aim to determine the primary and secondary acoustic correlates of the categorical contrasts under investigation (Voice Quality, Tone, Vowel Quality). In the case of the phonation measurements, we started out with using Linear Discriminant Analysis (LDA) to determine which measure is most sensitive to the Voice Quality contrast (cf. Garellek 2020). Subsequently, we used the best phonation measure as determined through LDA in the descriptive and inferential statistics. The inferential tests use linear mixed effects (LME) modeling. The fixed effects structure is the same across the LME models reported for phonation, F0, and vowel formants: it includes the main effects of the three factors (Voice Quality, Tone, and Vowel Quality), their two- and three-way interactions, and the additive effect of Context. All factors used sum-to-zero contrasts for the ANOVA table computations. In determining the random effects structure for the LME models, we start out, for phonation, F0, and vowel formants alike, from the same maximal model (Barr et al. 2013). This maximal model has random effects by Speaker, which include a random intercept and random slopes for Voice Quality, Tone, and Vowel, as well as their two- and three-way interactions. It also models random effects by Item, which include a random intercept and random slopes for Voice Quality and Tone, as well as their two-way interaction. The factor Vowel is not part of the maximal random effects structure for Item, because each item occurs with a single vowel only. Starting from this maximal model, we simplified the random effects structure, taking out interactions and factors which led to convergence issues or singularity of the model matrices due to being highly correlated with other random effects or having a standard deviation very close to zero. The resulting final models will be reported in the sections reporting the LME results for phonation, F0, and vowel formants. In assessing statistical significance on the basis of said models, we set the significance level (α') by adjusting the overall significance level ($\alpha = 0.05$) using the

Bonferroni method. With eight comparisons, this yields a significance threshold of $\alpha' = 0.00625$ (0.05/8) and only p-values smaller than the adjusted significance level α' will be considered significant. This was done in a conservative spirit, to prevent reporting a significant result due to minor deviations from normality, as indicated by some of the QQ-plots in the Supplementary Material. For Spectral Emphasis, transforming the outcome variable helped alleviate the normality concerns, but for the other DVs, neither log transformations or Box-Cox transformations led to a noticeable improvement, hence the more conservative approach used in this paper.

Statistical analyses were performed in R (R Core Team 2021, version 4.4.0). In addition to the base package, we used the packages gplots (Warnes et al. 2024), phonR (McCloy 2016), MASS (Venables & Ripley 2002), tidyverse (Wickham et al. 2019), lme4 (Bates et al. 2015), lmerTest (Kuznetsova et al. 2017), forecast (Hyndman et al. 2024), performance (Lüdecke et al. 2021), and sjPlot (Lüdecke et al. 2024).

3 Results

In this section we present the analysis of the measurements on phonation (Section 3.1), F0 (Section 3.2), and formants (Section 3.3).

We do not report in detail on the duration measurements, because there is not much going on here. Vowel length was kept constant in the materials, i.e., the vowel is invariably overlong. The values for vowel duration do not vary greatly as a function of the phonological factors under investigation. In relation to Voice Quality, we find mean values of 307 ms for Modal vowels vs. 310 for Breathy vowels; in relation to Tone, the mean values are 314 for Low tones vs. 303 for High tones.

3.1 Phonation

To analyze phonation in the data set, we first assess the relative ability of the full set of phonation measures to discriminate between the levels of the Voice Quality contrast using Linear Discriminant Analysis (LDA) (Section 3.1.1). The results of the LDA then inform the descriptive (Section 3.1.2) and inferential (Section 3.1.3) analyses of phonation as a function of the three phonological contrasts at issue – Voice Quality, Tone, and Vowel Quality. Finally, we consider whether phonation varies across the time course of the vowel, as hypothesized by the Sequencing Hypothesis (Section 3.1.4).

3.1.1 Determining the best acoustic measure

We carried out an LDA to determine which phonation measure distinguishes best between the levels of the Voice Quality contrast (Modal vs. Breathy). A total of eleven acoustic measures of phonation were taken into consideration: H1*-H2*, H2*-H4*, H1*-A1*, H1*-A2*, H1*-A3*,

CPP, HNR05, HNR15, HNR25, HNR35, and Spectral Emphasis. These measures are all first z-transformed by speaker, to normalize for the speaker's inherent phonation. As Voice Quality has two levels (Modal vs. Breathy), the LDA generates one discriminant function. Spectral emphasis is the measure that is correlated most strongly with the linear discriminant function, with correlation coefficient $r = 1.39$. It is followed by H1*-A2* ($r = 0.54$), HNR15 ($r = 0.16$), and HNR25 ($r = 0.11$). The correlation coefficients of the other measurements are smaller than 0.1. Overall, the accuracy of this model is very high, with 92.7 percent of the data correctly classified based on this discriminant function. Spectral Emphasis is well ahead of H1*-A2* and the other measures in its ability to distinguish between Modal and Breathy vowel: in a separate LDA with Spectral Emphasis as the sole measure, the correct classification score is 89.2 percent, i.e., just 3.5% percent less accurate than with the full set of 11 measurements.

Further analyses of phonation are based on Spectral Emphasis and H1*-H2*. Spectral Emphasis is included because the LDA showed that it is the measure that best distinguishes Modal vs. Breathy vowels in this dataset. H1*-H2* is included because it is the best acoustic correlate of breathy voice quality in many studies, and as such its inclusion allows for straightforward comparison with earlier work on phonation. Moreover, H1*-H2* is the second-best phonation measure for the current dataset: in an LDA based on the acoustic measurements except for Spectral Emphasis, it has the highest correlation coefficient, with $r = -1.08$. This LDA yields a correct classification of 88.4 percent, as compared to 92.7 correct classification for the LDA with Spectral Emphasis included.

3.1.2 Phonation as a correlate of Voice Quality, Tone and Vowel Quality: descriptive statistics

In this section we explore the role of phonation as a correlate of the phonological contrasts under investigation through descriptive statistics. We start with Voice Quality, for which we would expect phonation to represent the primary correlate. As seen from **Figure 2A**, on average Modal vowels have considerably higher values for Spectral Emphasis than Breathy vowels; this indicates that Modal vowels are realized with relatively more high-frequency energy. **Figure 2A** also shows that the distributions of Modal and Breathy vowels are well separated. This can be seen from the error bars, which cover one standard deviation, i.e. 68% of the variability around the mean; the fact that they do not overlap is indicative of a clear difference in the phonetic realisation of the phonological categories.

For the sake of comparison, the corresponding results for H1*-H2* are included in **Figure 2B**. Here the Breathy vowels have higher values. This is to be expected, because a higher value for H1*-H2* indicates that H2 is small relative to H1. In line with the LDA results reported in Section 3.1.1, Modal vowels and Breathy vowels are less well separated in terms of H1*-H2* than in terms of Spectral Emphasis.
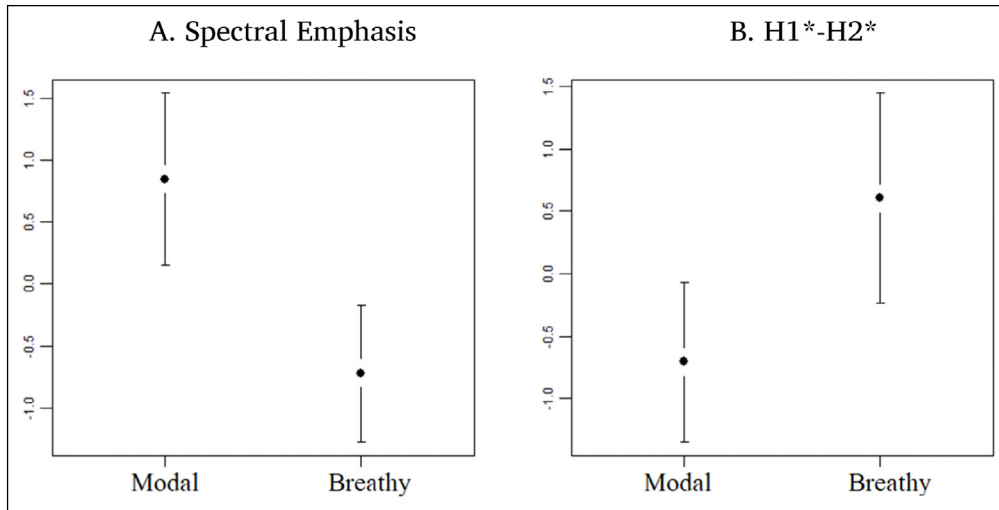
Figure 2: Means ± one standard deviation for A. Spectral Emphasis and B. H1*-H2* measured at the temporal midpoint of the vowel, as a function of the Voice Quality contrast (Modal vs. Breathy), over all speakers and items. The values are z-transformed by speaker.

Then we examine the extent to which phonation varies as a function of Tone and Vowel Quality in addition to Voice Quality. **Figure 3A** reveals that phonation does not distinguish Low tones from High tones, neither on Modal nor on Breathy vowels. Considered from the perspective of Voice Quality, this graph shows that the separation in phonation between Modal and Breathy is salient on both Low- and High-toned vowels. In summary, phonation is a phonetic correlate of Voice Quality, but not of Tone. This is contrary to Edmondson and Esling (2006:188), who state that the categories of the Voice Quality contrast involve different F0 patterns. Instead, the lack of a substantial difference between Low-toned and High-toned Breathy vowels supports an analysis in which Voice Quality and Tone are orthogonal dimensions of contrast.

Finally, **Figure 3B** shows that Spectral Emphasis is sensitive to Vowel Quality: more open vowels have higher values for Spectral Emphasis than more closed vowels. The fact that vowels which have higher F1 values have more high-frequency energy reflects a widely attested cross-linguistic pattern (Esposito, Sleeper & Schäfer 2019:374). This influence of Vowel Quality on Spectral Emphasis accounts for some of the variability within the distributions for Modal and Breathy vowels in **Figure 2**.

It is worthwhile to point out that the vowel /u̯/ clearly patterns with the Breathy vowels rather than with the Modal in terms of energy distribution (**Figure 3B**). As noted above (Section 1.2), the high back vowel does not contrast for voice quality and, with the exception of Malou (1988), all studies have interpreted it as Breathy. Its patterning with the Breathy vowels in **Figure 3B** corroborates the latter interpretation.
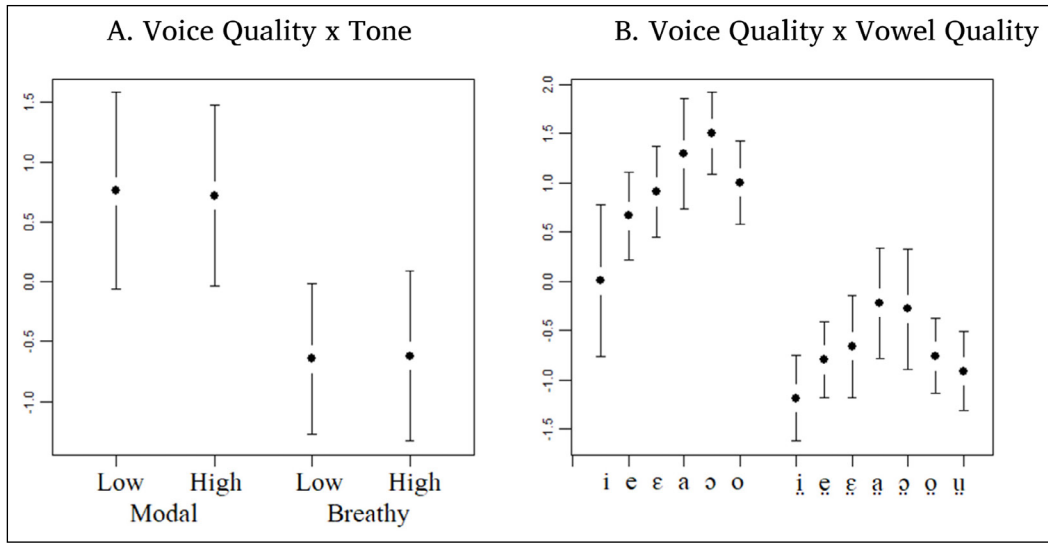
Figure 3: Means ± one standard deviation for Spectral Emphasis, measured at the temporal midpoint of the vowel, as a function of A. Voice Quality (Modal vs. Breathy) and Tone (Low vs. High); and B. Voice Quality and Vowel Quality, in both cases over all speakers and items. The values are z-transformed by speaker.

3.1.3 Phonation as a correlate of Voice Quality, Tone and Vowel Quality: inferential test

We carried out a linear mixed effects (LME) analysis with as the dependent variable Spectral Emphasis, the phonation measure that conditioned the highest correlation coefficient in the LDA (Section 3.1.1). Items with the vowel /ɯ̤/ were removed from the dataset before running the LME model, because they are invariably Breathy; failure to do so results in rank deficiency of the model matrix. Using Spectral Emphasis as the dependent variable resulted in a departure from normality according to the residuals QQ-plot. To improve this, the values for the dependent variable (Spectral Emphasis) were added 1, log-transformed, and subsequently z-transformed. This resulted in an improved QQ-plot of normality of the residuals (see Figure S2.1 in the Supplementary Material). The fixed effects structure is the same across the LME models reported for phonation, F0, and vowel formants: it includes the main effects of the three factors (Voice Quality, Tone, and Vowel Quality), their two- and three-way interactions, and the additive effect of Context. After pruning of the maximal model for the random factors as described in Section 2.5, we arrive at a random effects structure that includes a random intercept by Item and, as random effects by Speaker, a random intercept, random slopes of Voice Quality, Tone, and their interaction. In **Table 4** we report the result of performing a Type III Anova of the final model, with denominator degrees of freedom computed according to the Satterthwaite approximation. Diagnostic plots for the model assumptions are reported in Figure S2.2 of the Supplementary Material.

	SSq	MnSq	DF	DenDF	F	Pr(>F)
Context	0.01	0.01	1	1512.92	0.03	0.874
VoiceQ	**12.54**	**12.54**	**1**	**8.00**	**56.49**	**<0.001**
Tone	0.002	0.002	1	7.99	0.01	0.919
VowelQ	**173.99**	**34.80**	**5**	**24.19**	**156.73**	**<0.001**
VoiceQ*Tone	**0.33**	0.33	1	8.05	1.50	0.256
VoiceQ*VowelQ	**15.24**	**3.05**	**5**	**1513.99**	**13.73**	**<0.001**
Tone*VowelQ	**8.80**	**1.76**	**5**	**1514.40**	**7.93**	**<0.001**
VoiceQ*Tone*VowelQ	2.00	0.40	5	1514.03	1.80	0.11

Table 4: Type III Analysis of variance of the fixed effects in a linear mixed effects model of Spectral Emphasis. Results significant at the Bonferroni-corrected alpha (0.00625) are in bold. The R formula of this model is: lmer(zScore(log(SpectralEmphasis+1)) ~ Context + VoiceQuality*Tone*VowelQuality + (1 | Item) + (1 + VoiceQuality*Tone | Speaker)).

The linear mixed effects model yields significant main effects of both Vowel Quality and Voice Quality. These are the significant effects with the biggest F values. These results indicate that, in addition to the inherent effect of vowel height on energy distribution, Voice Quality conditions a sizeable effect on this phonetic parameter. Tone does not condition a significant effect on Spectral Emphasis. In addition, two of the interactions present smaller F values. First, the more substantial of these is the interaction between Vowel Quality and Voice Quality. As seen in **Figure 3B**, the difference in Spectral Emphasis as a function of Vowel is greater in Modal Voice than in Breathy Voice. Second, the interaction between Tone and Vowel Quality also registers a small significant effect.

3.1.4 Phonation across the vocalic domain

In this section, we address the question as to whether the Voice Quality contrast is realized to a greater extent in the first or the second half of the vowel. This question relates to the Sequencing Hypothesis (Section 1.1), which holds that Voice Quality and Tone may be prone to be phonetically realized on different parts of the vowel (Silverman 1997). To this end, we examine the most successful measures, Spectral Emphasis and H1*-H2* at 25%, 50%, and 75% into the vowel's duration. The descriptive statistics for both of these measures are shown in **Figure 4**. Both for Spectral Emphasis and H1*-H2*, the values diverge substantially between Modal and Breathy vowels, at each of the three time points in the vowel. In contrast, there is not much divergence between the values of Modal vowels across the three time points, and the same goes for the values of the Breathy vowels. Scrutinized in detail, we see for Breathy

vowels a slight downward drift in the values for Spectral Emphasis across the time course of the vowel, and an upward drift across the time course for H1*-H2*. Both of these findings indicate that vowels are realized with slightly less laryngeal constriction further along in the vowel. While this phenomenon affects Modal and Breathy vowels alike, Breathy vowels are affected to a greater extent, resulting in a greater separation between Modal and Breathy vowels further along in the vowel's duration. The lack of a salient difference in the realization of the Voice Quality contrast in different parts of the vowels is in line with our impressionistic observations and earlier studies (e.g. Andersen 1987; Denning 1989; Malou 1988; Remijsen & Manyang 2009), none of which report a difference in the realization of the Voice Quality contrast across the vocalic domain in Dinka.
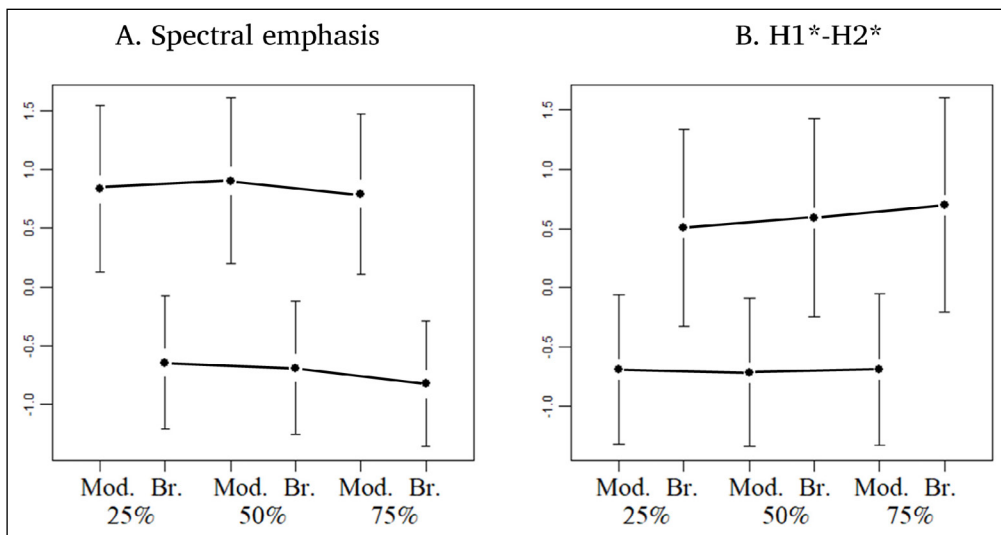

Figure 4: Means $\pm$ one standard deviation for Modal and Breathy vowels at 25%, 50% and 75% into the vowel's duration, for A. Spectral Emphasis and B. H1*-H2* (both z-transformed by speaker), over all speakers and items.

3.1.5 Summary

Through an LDA, we found that Spectral Emphasis is the measurement that distinguishes best between Modal and Breathy vowels, followed by H1*-H2*. As seen from **Figure 4**, these measures vary little over the time domain of the vowel. Considered as a function of the various factors under investigation, we find that Spectral Emphasis differs substantially as a function of both Voice Quality and Vowel Quality, and both of these differences translate into sizeable effects in the LME model. In contrast, Tone does not condition a significant effect. These findings are as expected. Spectral Emphasis is a measure of energy distribution; and while this particular energy distribution has been little-used in the study of phonation, other energy distribution measurements, and H1*-H2* in particular, are effective in the measurement of phonation

contrasts across a wide range of languages (Gordon & Ladefoged 2001; Garellek 2019). As for the effect of Vowel Quality on phonation, this is also in line with a cross-linguistically attested pattern, whereby vowels with higher F1 values are creakier and vowels with lower F1 values are breathier (Esposito et al. 2019). The lack of an effect of Tone underscores the independence of Tone and Voice Quality as dimensions of contrast in Dinka.

3.2 Fundamental frequency (F0)

3.2.1 F0 as a correlate of Voice Quality, Tone and Voice Quality: descriptive statistics

The F0 measure used in this section is F0 averaged over the central 50% of the vowel's duration, i.e., over the time domain from ¼ to ¾ of the vowel's time domain. As seen from **Figure 5A**, the distributions for Low and High tones are well separated: the mean values are 4.4 semitones apart (128 Hz for Low tones and 165 Hz for High tones), which is salient in perceptual terms. In addition, the standard deviations do not overlap. **Figure 5B** shows the influence of Context. In the Sentence context, where the target word is embedded and sentence-finally and therefore affected by declination across the utterance domain, F0 is considerably lower, but this effect is limited to the Low tones. This is in line with findings that declination may affect Low tones more than High tones (cf. Connell & Ladd 1990). Specifically, Low tones are 10 Hz lower in sentence-final context than in isolation – 123 Hz vs. 133 Hz, respectively. For the High tone, the mean values are just 1Hz apart – 165 Hz vs. 166 Hz, respectively.
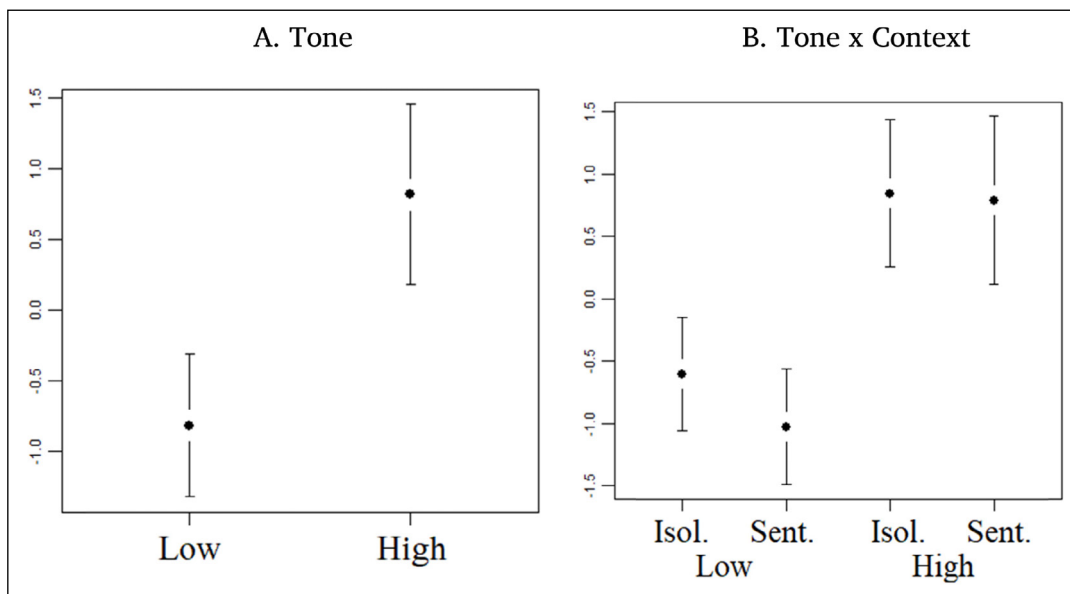

Figure 5: Means ± one standard deviation for F0 measured over the central 50% of the vowel's duration, by A. Tone and B. Tone crossed with Context, over all items and speakers, after z-transformation by speaker.

Having established that F0 distinguishes Tone, we now move on to examine the influence of Voice Quality and Vowel Quality on F0, using the same measure as in **Figure 5**. The results are presented in **Figure 6**. Starting with the influence of Voice Quality on F0, **Figure 6A** shows that Breathy vowels on average have higher F0 than Modal vowels, and this difference is slightly greater when the tone is High, where the F0 for Breathy vowels is 10 Hz higher on average than that of Modal vowels; in contrast, when the tone is Low, the difference is just 7 Hz. Across levels of Tone, Breathy vowels are 9 Hz higher in F0 than Modal vowels, on average – the mean values are 151 and 142 Hz, respectively. This is a difference of 1 semitone; while this may be audible, it is considerably smaller than the difference conditioned by Tone (4.4 semitones). Interestingly, the direction of the difference in F0 as a function of phonation is the opposite of the pattern typically observed in Southeast Asian register languages, where a more lax phonation goes hand-in-hand with lower F0 (Brunelle 2012).
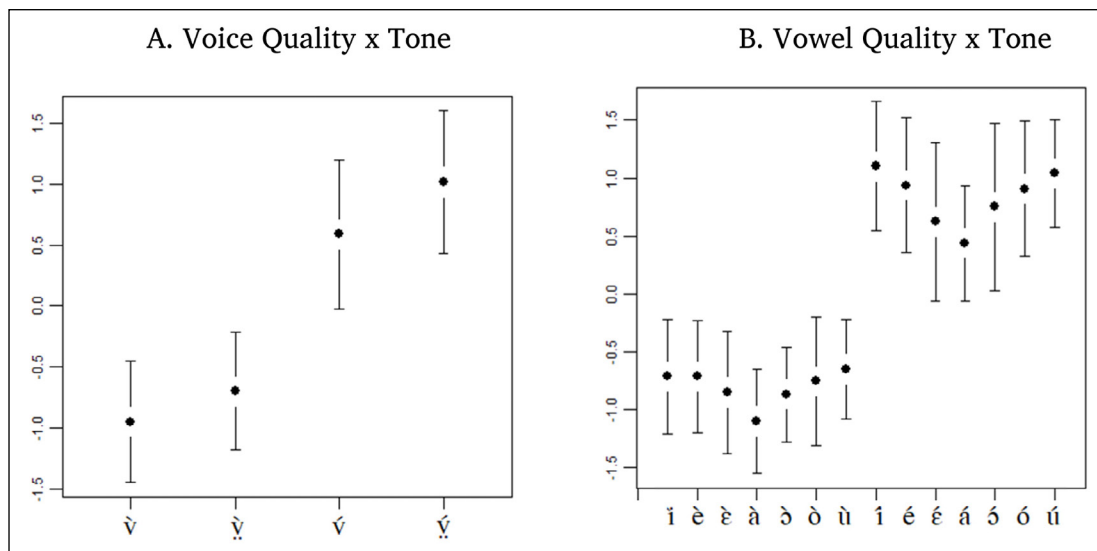

Figure 6: Means ± one standard deviation for F0 over the central 50% of the vowel's duration, z-transformed by speaker, by A. Tone (ṽ, v́) crossed with Voice Quality (v, v̤); and B. Tone (ṽ, v́) crossed with Vowel Quality (i,e,ɛ,a,ɔ,o,u), across all speakers and items.

As for the effect of Vowel Quality in addition to that of Tone, **Figure 6B** reveals the influence of vowel height: the F0 values are lower for open vowels than for closed vowels. This correlation between F0 and vowel height reflects a cross-linguistic universal (Hombert 1978; Whalen & Levitt 1995). As seen from **Figure 6B**, this influence is present both when the Tone is Low and when it is High. Across levels of Tone, the difference between the closed vowel /i/ and the open vowel /a/ is 12 Hz on average.

3.2.2 F0 as a correlate of Voice Quality, Tone and Voice Quality: inferential statistics

We carried out a linear mixed effects (LME) analysis with F0 as the dependent variable. For this analysis, we used the F0 measurement at the temporal mid point of the vowel, i.e., the same time value used in the linear mixed effects analysis of phonation (Section 3.1.3). Similarly to the inferential test for phonation, the dependent, here F0, was z-transformed. Also as in that analysis, items with the vowel /ʉ/ were removed in advance, to forestall rank deficiency in the model matrix. After simplifying the random effects structure as described in Section 2.5, the resulting model has a random intercept and random slope for Tone by Item, as well as a random intercept and slopes for Voice Quality and Tone by Speaker. The fixed effects structure in the model of F0 was kept the same as in the analysis for phonation. A Type III Anova of the final model is reported in **Table 5** below, where the denominator degrees of freedom are computed according to the Satterthwaite approximation, while diagnostic plots for the model assumptions are reported in Figure S3.1 of the Supplementary Material.

	SSq	MnSq	DF	DenDF	F	Pr($>$F)
Context	**4.27**	**4.27**	**1**	**1495.86**	**106.12**	**<0.001**
VoiceQ	0.25	0.25	1	8.00	6.27	0.037
Tone	**2.39**	**2.39**	**1**	**8.58**	**59.26**	**<0.001**
VowelQ	**1.05**	**0.21**	**5**	**24.07**	**5.21**	**0.002**
VoiceQ*Tone	**0.33**	**0.33**	**1**	**1496.41**	**8.27**	**0.004**
VoiceQ*VowelQ	**2.23**	**0.45**	**5**	**1496.33**	**11.10**	**<0.001**
Tone*VowelQ	0.15	0.03	5	24.28	0.76	0.59
VoiceQ*Tone*VowelQ	**1.85**	**0.37**	**5**	**1496.29**	**9.17**	**<0.001**

Table 5: Type III Analysis of variance of the fixed effects in a linear mixed effects model of F0. Results significant at the Bonferroni-corrected alpha (0.00625) are in bold. The R formula of this model is: lmer(F0_ZTransformed ~ Context + VoiceQuality*Tone*VowelQuality + (1 + Tone | Item) + (1 + VoiceQuality + Tone | Speaker)).

Among the main effects, the biggest effects are those of Context and Tone, both of which register sizeable F values. In addition, Vowel Quality registers a smaller effect, and Voice Quality approaches significance. These results are in line with the descriptive statistics reported above: High tones have substantially higher F0 values than Low tones; the Isolation context has higher F0 values than the Sentence context; more closed vowels have higher F0 values than more open vowels; and Breathy vowels have higher F0 than Modal vowels. As for the interactions, most of these are significant as well. The significant interaction between Voice Quality and Tone

reflects the fact that the difference in F0 between Modal and Breathy vowels is greater in High-toned than in Low-toned syllables. The significant three-way interaction between Voice Quality, Tone and Vowel Quality indicates that the above-mentioned interaction between Tone and Voice Quality is substantially stronger for some vowels than for others.

3.2.3 F0 across the vocalic domain

Now we consider whether the Tone contrast is realized to a greater extent in the first or the second half of the vowel. **Figure 7** shows F0 values at 25%, 50% and 75% of the vowel's duration by Tone. The measurements reveal that F0 increases over the time course of High-toned vowels, while it decreases over the time course of Low-toned vowels. This is in line with the cross-linguistic universal for tone targets to be reached relatively late in the syllable (Xu 1999). Accordingly, the distributions of Low vs. High tones, as indicated by the standard deviations, are further apart further along in the vowel. This means that the realization of the tone contrast is more salient in the second half of the vowel. We will return to this in the Discussion (Section 4).
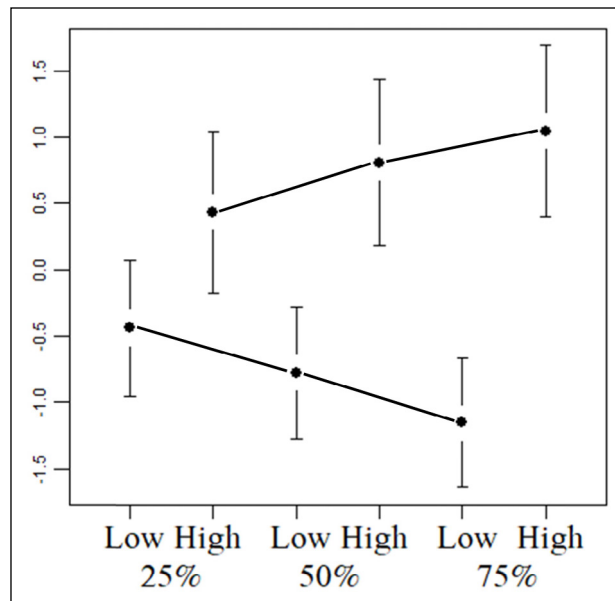

Figure 7: Means ± one standard deviation for F0, z-transformed by speaker, by Tone (Low, High) at 25%, 50% and 75% into the vowel's duration, over all speakers and items.

3.3 Formants (F1, F2)

3.3.1 F1 and F2 as a correlate of Voice Quality, Tone and Voice Quality: descriptive statistics

The acoustic realization of the Dinka vowels is illustrated in **Figure 8**, which again shows means and standard deviations. Because F1 and F2 define a two-dimensional space, the standard deviations are now marked by ellipses rather than by error bars. The distributions of the seven

vowel categories are fairly well separated, with only /i/ vs. /e/ and /u/ vs. /o/ showing a limited amount of overlap at the level of one standard deviation. This fits with the expectation for the first and second formants (F1 and F2) to be the primary correlates of the Vowel Quality contrast.
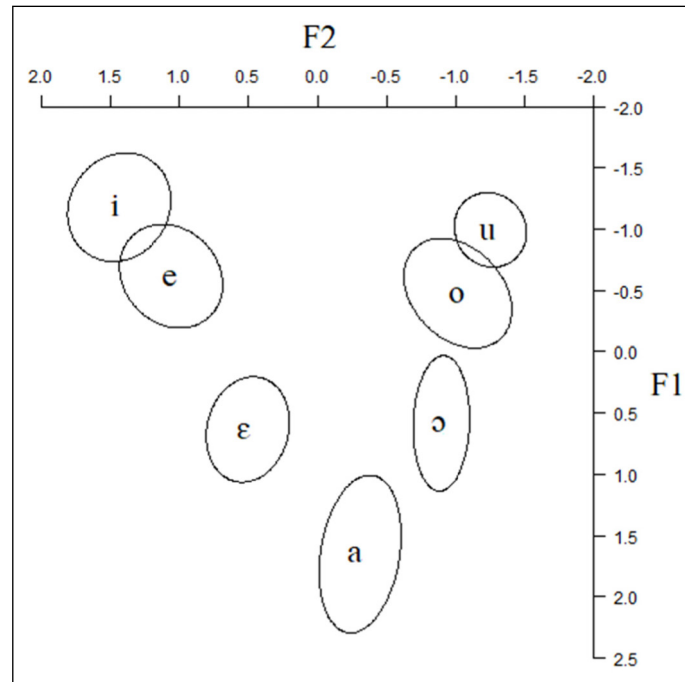

Figure 8: Means and standard deviations for the first and second formants (F1, F2), z-transformed by speaker, for each of the seven Dinka vowel categories across the eight speakers. The F1 and F2 measurements are made at the temporal midpoint.

Next we consider the potential for phonetic vowel quality to play an additional role as a secondary correlate to the other two phonological distinctions that were experimentally manipulated: Voice Quality and Tone contrasts. The results are shown in **Figure 9**. As seen in **Figure 9A**, Voice Quality has a small but consistent effect on F1, the formant determined primarily by vowel height: for each of the six vowel phonemes that appear both in Modal Voice and in Breathy Voice (/i, e, ɛ, a, ɔ, o/), the Breathy vowel has a lower F1 value, meaning that it is more closed. Across the dataset, the average difference in F1 between the Breathy vs. Modal vowels is 80 Hz. The same phenomenon—lower F1 values for Breathy vowels—is reported in Malou (1988). With respect to F2, which is associated with backness, only /o/ shows a substantial difference: the F2 value of Breathy /o̤/ is considerably higher than that of Modal /o/. The average values are 1076 Hz for Breathy /o̤/ vs. 882 Hz for Modal /o/ – a difference of 192 Hz. Moreover, the lack of overlap of ellipses (at 1 standard deviation) indicates that the realisations of /o/ and /o̤/ are well separated in terms of phonetic vowel quality. This finding that Breathy /o̤/ is centralised fits with several earlier studies, which report centralisation as a
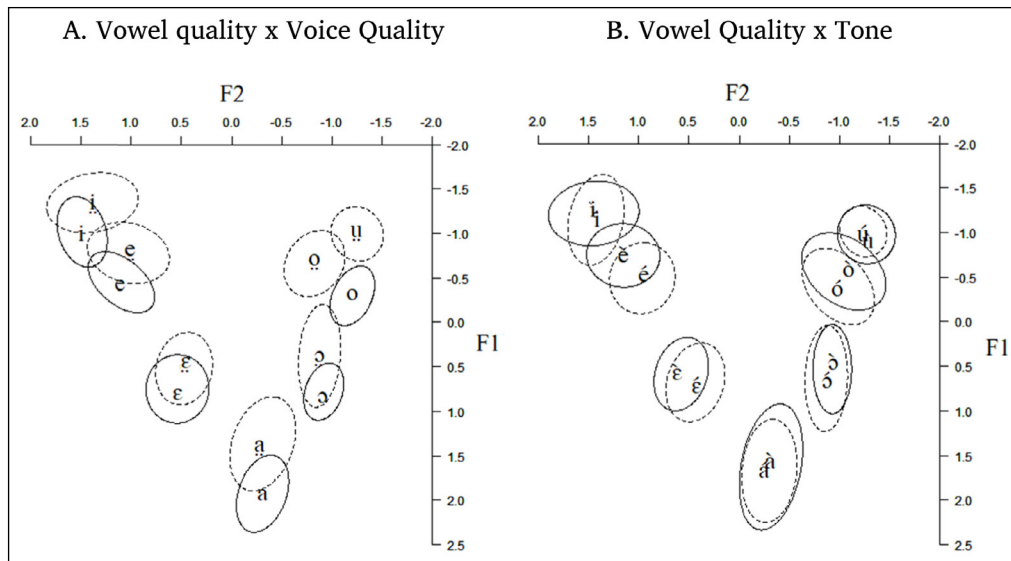
Figure 9: Means and standard deviations for the first and second formants (F1, F2) for each of the seven Dinka vowel categories across the eight speakers, by Voice Quality (9A) and by Tone (9B). The F1 and F2 measurements are made at the temporal mid point, and z-transformed by speaker.

feature specifically of some Breathy vowels, but not of any of the Modal vowels (Tucker 1936; Ayom 1980; Remijsen 2013).

Based on our investigations into a wide variety of Dinka dialects, it appears that the centralisation of Breathy /o̤/ is specific to the Bor South dialect. This is illustrated in **Table 6**, which presents transcriptions with associated sound examples of a minimal pair for Voice Quality involving Modal /o/ vs. Breathy /o̤/, as produced on the one hand by one of the speakers drawn from our dataset, and, for the sake of comparison, from a speaker of Luanyjang, a dialect within the Rek cluster. While the Breathy voice involves breathy phonation for both speakers, it is only the speaker of Bor South who pronounces Breathy /o̤/ with centralisation, i.e., [ɵ̈].

	Bor South Dinka		**Luanyjang Dinka**	
/ròoor/ 'forest'	[ròoor]	▶	[ròoor]	▶
/rò̤oor/ 'men'	[rè̤ɵɵr]	▶	[rò̤oor]	▶

Table 6: The phonetic realisation of the vowel /o/ as a function of Voice Quality and dialect. Each of the forms has a sound example associated with it, elicited in isolation.

As for the effect of Tone on phonetic vowel quality, **Figure 9B** shows that High-toned vowels have slightly higher F1 values for all vowel phonemes except for /u/. The difference is very small: the mean F1 values of High-toned vowels are only 21 Hz higher than those of Low-toned vowels (501 vs. 521 Hz, respectively), which is negligible in perceptual terms. Moreover, the standard deviations largely overlap.

3.3.2 F1 and F2 as a correlate of Voice Quality, Tone and Voice Quality: inferential statistics

We carried out linear mixed effects (LME) analyses with F1 and F2 as the dependent variables. The measurements were made at the temporal midpoint of the vowel, i.e., the same time value used in the LME analyses of phonation and F0. Again, the dependents were z-transformed, and items with the vowel /ᵾ/ were removed in advance to forestall rank deficiency of the model matrix. The fixed effects structure in the models for F1 and F2 was kept the same as the earlier analyses. After simplification of the maximal model as described in Section 2.5, the final model for F1 includes a random intercept and slope for Voice Quality by Speaker. Models for F1 with Item included as a random factor did not converge. The final model for F2 includes a random intercept by Item, as well as a random intercept, random slopes for Voice Quality and Tone and their interaction by Speaker. A Type III Anova with denominator degrees of freedom approximated via the Satterthwaite method is shown in **Table 7** for F1 and in **Table 8** for F2. Their diagnostic plots are shown in Figure S4.1 and S4.2 of the Supplementary Material, respectively.

	SSq	MnSq	DF	DenDF	F	Pr($>$F)
Context	**0.72**	**0.72**	**1**	**1552.04**	**9.39**	**0.002**
VoiceQ	**9.50**	**9.50**	**1**	**8.02**	**123.42**	**$<$0.001**
Tone	**6.72**	**6.72**	**1**	**1552.11**	**87.29**	**$<$0.001**
VowelQ	**1249.65**	**249.93**	**5**	**1552.09**	**3246.69**	**$<$0.001**
VoiceQ*Tone	0.13	0.13	1	1552.11	1.67	0.196
VoiceQ*VowelQ	**3.48**	**0.70**	**5**	**1552.12**	**9.05**	**$<$0.001**
Tone*VowelQ	**1.92**	**0.38**	**5**	**1552.10**	**4.99**	**$<$0.001**
VoiceQ*Tone*VowelQ	0.90	0.18	5	1552.06	2.33	0.04

Table 7: Type III Analysis of variance of the fixed effects in a linear mixed effects model of F1. Results significant at the Bonferroni-corrected alpha (0.00625) are in bold. The R formula of this model is: lmer(F1_ZTransformed ~ Context + VoiceQuality*Tone*VowelQuality + (1 + VoiceQuality | Speaker)).

The analysis for F1 yields significant main effects for all of the factors, and especially the factors under investigation. The very large F value of Vowel Quality is in line with its primary role in distinguishing vowels of different height. The second biggest F value is that of Voice Quality, which reflects the fact that, for all six of the vowel phonemes that present a distinction in Voice Quality, the Breathy vowels have a higher F1, i.e., is more closed than the corresponding Modal vowels. The significant effect of Tone results from the fact that Low-toned vowels have a lower

F1 value than High-toned vowels. The F value for Tone is smaller than that for Voice Quality, in line with the degree of separation between the factor levels in **Figure 9A** vs. **Figure 9B**, respectively. There are also significant interactions between Vowel Quality on the one hand and both Voice Quality and Tone on the other, indicating that influence on F1 of Voice Quality and of Tone is greater for some vowels than for others.

	SSq	MnSq	DF	DenDF	F	Pr(>F)
Context	0.02	0.02	1	1512.07	0.34	0.557
VoiceQ	0.08	0.08	1	8.05	1.70	0.228
Tone	0.25	0.25	1	7.99	5.49	0.047
VowelQ	**280.86**	**56.17**	**5**	**24.13**	**1224.15**	**<0.001**
VoiceQ*Tone	0.01	0.01	1	8.03	0.21	0.661
VoiceQ*VowelQ	**10.90**	**2.18**	**5**	**1512.78**	**47.51**	**<0.001**
Tone*VowelQ	**3.75**	**0.75**	**5**	**1513.00**	**16.32**	**<0.001**
VoiceQ*Tone*VowelQ	0.68	0.14	5	1512.89	2.96	0.011

Table 8: Type III Analysis of variance of the fixed effects in a linear mixed effects model of F2. Results significant at the Bonferroni-corrected alpha (0.00625) are in bold. The R formula of this model is: lmer(F2_ZTransformed ~ Context + VoiceQuality*Tone*VowelQuality + (1 | Item) + (1 + VoiceQuality*Tone | Speaker)).

In relation to F2, the only factor that registers a significant main effect, and a very substantial one at that, is Vowel Quality. This factor is also significant in interaction with a) Voice Quality and b) with Tone. The significant main effect of Vowel Quality is to be expected given the role of F2 in distinguishing between front and back vowels. The significant interaction between Vowel Quality and Voice Quality reflects the fact that the Breathy vowels and Modal vowels diverge substantially in F2 for the vowel /o/, but not for the other vowels (cf. **Figure 9A**). The significant interaction between Vowel Quality and Tone reflects the fact that F2 is lower for High-toned vowels than for Low-toned vowels specifically for some vowels, in particular /e,ɛ/, whereas the relation is reversed for the /o,ɔ/ (cf. **Figure 9B**).

4 Discussion

We begin this discussion by summarizing the phonetic realization of Voice Quality, Tone and Vowel Quality in Section 4.1. We then examine the findings in a cross-linguistic context, considering tendencies observed and hypotheses postulated in relation to other languages with independently contrastive Voice Quality and Tone (Section 4.2).

4.1 The phonetic realizations of Voice Quality, Tone and Vowel Quality

4.1.1 Voice Quality

The Voice Quality contrast is phonetically realized primarily through energy distribution. **Figure 2** shows a clear separation between the distributions of Modal vs. Breathy vowels in terms of both Spectral Emphasis (**Figure 2A**) and H1*-H2* (**Figure 2B**). Moreover, **Figure 4** shows that this difference in energy distribution as a function of Voice Quality is consistent across the time domain of the vowel. In addition, the Voice Quality contrast has vowel height (F1) as a secondary correlate. **Figure 9A** shows that, for each of the six vowel phonemes that contrast for Voice Quality, the Modal vowels are more open (higher F1) than the corresponding Breathy vowels. Also, Breathy /o̤/ is centralized, as seen from the higher F2 values, relative to Modal /o/.

The interpretation that phonation is the primary correlate and F1 the secondary one is based on the difference in the degree of overlap between the distributions. There is no overlap between the distributions of Modal vs. Breathy vowels at the level of one standard deviation, as represented through the whiskers in **Figure 2A**. In the case of the role of F1 and F2 in marking Voice Quality, the standard deviations in **Figure 9A** (marked by ellipses) overlap for all but one of the vowels that participate in the Voice Quality contrast: /i,e,ɛ,a,ɔ/. It is only in the case of the vowel /o/ that the ellipses do not overlap; this is primarily due to Breathy /o̤/ having substantially higher F2 values than Modal /o/, indicating centralisation of Breathy /o̤/. But even for this vowel, the difference in formant values is smaller than the difference in phonation: note that the ellipses for /o/ and /o̤/ are contiguous in **Figure 9A** (formants), whereas the whiskers for these vowels are well separated in **Figure 3B** (phonation).

As for F0, Breathy vowels have higher F0 values than Modal vowels in both Low- and High-toned syllables (**Figure 6A**). However, the difference is small, and there is considerable overlap between the distributions.

Allophonic variation in the realization of the Voice Quality contrast is very limited. One of the speakers realizes Modal voice sporadically with irregular periodicity or aperiodicity. Additionally, all speakers realize Breathy vowel /o̤/ with centralisation. This centralisation is specific to the Bor South dialect. For example, it is not present in the acoustic measurements of vowel quality as a function of Voice Quality in Malou (1988).

4.1.2 Tone

Tone is realized primarily through F0. As seen from **Figure 5**, Low and High tones are characterized by a substantial difference in F0, and their distributions are well separated from one another. Moreover, inspection of F0 across equidistant time points (**Figure 7**) indicates that the separation increases over the course of the vowel's duration. There is also a small difference in formant values: six of the seven vowels have higher F1 values in a High-toned syllable than in a Low-

toned syllable (**Figure 9B**). However, the overlap between the standard deviations indicates that F1 does not play much of a role in the marking of the Tone contrast. As for the influence of Tone on energy distribution, **Figure 3A** reveals that Low and High tones do not diverge much on this measure at all, and it is not significant in the inferential test.

4.1.3 Vowel Quality

The primary phonetic correlates of Vowel Quality are the first and second formants (F1, F2). The seven vowels display a clear separation. This can be seen in **Figure 8**: overlap between the standard deviations is limited to /i/ vs. /e/ and /u/ vs. /o/, and limited in extent for these vowels. As seen from **Figure 9A**, this overlap is smaller still if we take into account the role of Voice Quality. Vowel Quality also affects F0 and phonation, in both cases as a function of vowel height: closed vowels have higher F0 (**Figure 6B**) and proportionally less high-frequency energy (**Figure 3B**) than open vowels. While the effects of Vowel Quality on F0 and Spectral Emphasis are significant, these phonetic parameters represent secondary correlates, characterized by overlap between the distributions.

4.1.4 Summary of primary and secondary correlates of Voice Quality, Tone, and Vowel Quality

In summary, our descriptive and inferential analyses indicate that Voice Quality, Tone and Vowel Quality each have their own primary correlate: phonation in the case of the Voice Quality contrast; F0 for the Tone contrast, and formants (F1, F2) for the Vowel Quality contrast. In addition to their primary correlate, they each influence other phonetic parameters to a lesser extent. This state of affairs as summarized in **Table 9**. As seen from this Table, formant values are the great multi-tasker, serving as secondary correlates for both Tone and Voice Quality in addition to their role as the primary correlate of the Voice Quality contrast.

Phonological contrast	Primary correlate	Secondary correlate(s)
Voice Quality	energy distribution	formants
Tone	F0	formants (weak)
Vowel Quality	formants	energy distribution, F0

Table 9: Correlates of Voice Quality, Tone and Vowel Quality in Bor South Dinka.

The observed state of affairs goes against Edmondson & Esling's (2006:188) interpretation that Dinka's phonation contrast represents a register system, in which the phonological categories are realized through patterns of F0 in addition to patterns of energy distribution. Contrary to this interpretation, the effect of Voice Quality on F0 is not significant in the present study.

4.2 The findings in cross-linguistic and typological perspective

4.2.1 The relation between phonation and F0

Considered from a cross-linguistic angle, the Dinka configuration is of interest in various ways. First we consider the vantage point of phonological register. In languages that present this configuration, F0, phonation, and vowel height represent the multiple correlates of a single phonological dimension of contrast: "[t]hese languages typically oppose a high register, which has a relatively high pitch, a tense voice quality, and a low vowel quality, to a low register, which has a comparatively lower pitch, a laxer voice quality, and a higher vowel quality" (Brunelle 2012). Considering the phonetic realization of the Dinka Voice Quality contrast, the patterning of vowel height with phonation (**Figure 9A**) is the same as in Southeast Asian register systems. This is in line with typological (Gordon & Ladefoged 2001:400) and articulatory evidence (Esling & Harris 2005) showing that larynx lowering is part of the articulation of a less constricted laryngeal setting. Such lowering increases the size of the supralaryngeal vocal tract, thereby lowering the F1 resonance frequency. As for the relation between Voice Quality and F0 in Dinka, the descriptive statistics do not show the pattern reported for Southeast Asian Register systems, but rather a tendency in opposite direction, with Modal vowels on average having lower F0 than Breathy vowels; this result approaches significance ($p = 0.037$). The significant interaction between Voice Quality and Tone, in conjunction with the descriptive statistics in **Figure 6A**, indicates that Breathy vowels have higher F0 than Modal vowels primarily on High-toned vowels.

In this context, it is worthwhile to note that Nuer, a closely-related language, presents a related phenomenon. Nuer also presents a contrast between Modal and Breathy vowels in addition to a tone contrast, and Modal voice conditions the High tone to be realised as a falling allotone, whereas it is realized as a high level allotone if the vowel is Breathy (Monich 2020). Hence the pitch percept of the High tone is lower in Modal voice than in Breathy voice in Nuer, just as it is in Dinka.

4.2.2 The Sequencing Hypothesis

According to the Sequencing Hypothesis (Silverman 1997), Voice Quality and Tone contrasts cannot be realized optimally simultaneously. This hypothesis can explain why non-modal phonations are realised more saliently early in the vowel in Jalapa Mazatec, whereas the Tone contrast is most salient in the second half of the vowel. Silverman acknowledges that the empirical facts do not support this hypothesis unequivocally: in Mpi, another language with independent tone and phonation contrasts, there is no evidence of sequencing of the phonetic realization of Voice Quality and Tone contrasts.

Our own results also do not align with the Sequencing Hypothesis. The key results are repeated in **Figure 10**. **Figure 10A** shows that Tone is realized more saliently in the second half of the vowel. This is to be expected, as tone targets are realized late across languages (Xu 1999). As for Voice Quality, **Figure 10B** shows that phonetic difference between Modal and Breathy

vowels is fairly constant between the first and the second half of the vowel. If anything, the difference is realized slightly more clearly in the second half, as indicated by the separation between the whiskers, which is greater at the 75% time point than at the 25% time point. This is contrary to the Sequencing Hypothesis: given that Tone is realized more clearly in the second half of the vowel, the Sequencing Hypothesis predicts that Voice Quality would be realized more clearly in the first half.
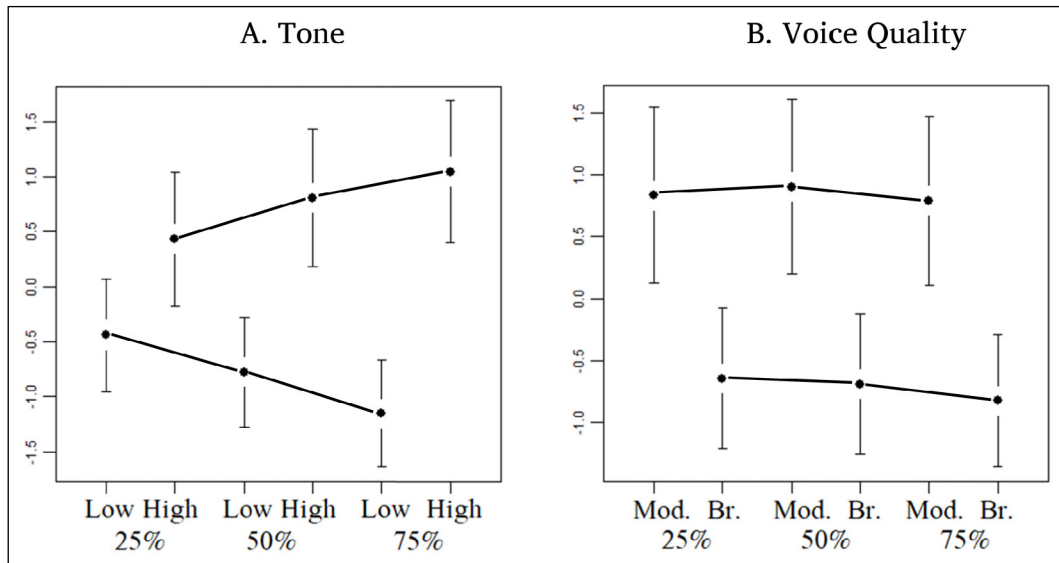

Figure 10: Means and standard deviation for A. F0 of Low-toned and High-toned vowels, and B. Spectral Emphasis of Modal and Breathy vowels, both at 25%, 50% and 75% into the vowel's duration, z-transformed by speaker, over all speakers and items.

In summary, there is no evidence that the phonetic realisations of tone and voice quality are sequenced in Dinka as they are in Jalapa Mazatec. It appears that the sequencing of the realization of tone and voice quality contrasts in Jalapa Mazatec is specific to that language, and can be attributed to the diachronic origin of its voice quality contrast in laryngeal coarticulations in onset consonants in Proto Mazatec (Kirk 1966). This bears out the value of expanding the evidence base on languages that present tone and voice quality as independent dimensions of contrast: if this evidence base is limited, a pattern that is specific to a particular language can more easily be mistaken for a cross-linguistic tendency.

5 Conclusion

This production study reveals that the Dinka contrasts of Voice Quality, Tone and Vowel Quality each have their own primary correlate, namely phonation, F0, and vowel formants, respectively. Aside from these primary correlates, these contrasts exert smaller influences on other phonetic

parameters (cf. **Table 9**). Voice Quality additionally affects vowel height (F1), and in the case of the vowel /o/, also degree of centralisation (F2). Tone additionally influences vowel height (F1), but its influence on this phonetic parameter is much more subtle than that of Voice Quality. Vowel Quality additionally influences both F0 and energy distribution.

These results are significant in the typological context of the limited number of languages with orthogonally contrastive Voice Quality and Tone: they corroborate that human language can contrast Voice Quality and Tone within the same vocalic domain. In particular, we find that the Voice Quality contrast is marked saliently on High-toned vowels, just as it is on Low-toned vowels. We also did not find evidence of sequencing of the realization of these distinctions in different parts of the vowel.

The results also have a bearing on the disagreement on the number of phonological voice qualities in Dinka. Our study focuses on the Bor South dialect—precisely the dialect for which some studies have postulated four phonological voice qualities (Modal, Breathy, Hard, Hollow), rather than the two suggested by the majority of studies (Modal, Breathy). We have argued that the morphological values hypothesized to be expressed by the two additional voice qualities can be analyzed with reference to Modal and Breathy in conjunction with phonological specifications for Vowel Length and Tone. Our study also calls into question the hypothesis that the Dinka Voice Quality contrast represents a register system, in which categories are realized through F0 in addition to phonation: there is no significant effect of Voice Quality on F0 in the linear mixed effects analysis. In conclusion, while it is impossible to prove a negative, i.e., that the additional levels of Hard and Hollow do not exist, our study does provide support for the binary interpretation of the Dinka voice quality contrast, even for the Bor South dialect.

Abbreviations

The following abbreviations are used in the glosses.

ADV	Adverb	NF	Infinitive
BNF	Benefactive	NOM	Nominalisation
CONJ	Conjunction	PST	Past
DEC	Declarative prefix	PET	Centripetal
FUG	Centrifugal	PL	Plural
FUT	Future	SG	Singular
GEN	Genitive		

Data availability

The Dinka dataset on which the analyses in this paper are based is publicly available in Edinburgh DataShare, an electronic archive (Remijsen et al. 2022). As supplementary materials, the paper includes the following. First, in relation to the inferential statistics, we provide additional diagnostic plots for the linear mixed effects (LME) models of the dependent variables considered in this paper. Second, we include sound clips for the transcribed examples in this paper (both in Tables and numbered illustrations). All of these materials, and also those of the Praat scripts used we created ourselves for the acoustic analysis, are additionally available in the OSF record associated with this study (https://osf.io/je9np).

Ethics and consent

The research presented in this paper was approved by the research ethics panel of the School of Philosophy, Psychology and Language Sciences at the University of Edinburgh as part of the project 'Suprasegmentals in three West Nilotic languages'. In line with the approved plan, all participants went through an informed consent procedure before data collection.

Funding information

Acknowledgements

Competing interests

References

Andersen, Torben. 1987. The phonemic system of Agar Dinka. *Journal of African Languages and Linguistics* 9. 1–27. DOI: https://doi.org/10.1515/jall.1987.9.1.1

Andersen, Torben. 1992–1994. Morphological stratification in Dinka: on the alternations of voice quality, vowel length and tone in the morphology of transitive verbal roots in a monosyllabic language. *Studies in African Linguistics* 23. 1–63. DOI: https://doi.org/10.32473/sal.v23i1.107416

Andersen, Torben. 2002. Case inflection and nominal head marking in Dinka. *Journal of African Languages and Linguistics* 23. 1–30. DOI: https://doi.org/10.1515/jall.2002.002

Ayom, Edward. 1980. Some aspects of Dinka Bor phonology and nominal plural formation. Khartoum: University of Khartoum MA thesis.

Ayom, Edward & Hall, R. M. R. & Odden, David. 1985. The phonetics of voice qualities in Dinka. Paper presented at the *16th Conference on African Linguistics*. Yale University.

Ayom, Edward B. G. 1987. A linguistic analysis of Dinka tongue twisters. *Anthropological linguistics* 29. 170–180.

Barr, Dale J., Levy, Roger & Scheepers, Christoph & Tilly, Harry J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68. 255–278. DOI: https://doi.org/10.1016/j.jml.2012.11.001

Bates, Douglas & Mächler, Martin & Bolker, Ben & Walker, Steve. 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

Belotel-Grenié, Agnès & Grenié, Michel. 1997. Types de phonation et tons en chinois standard. *Cahiers de Linguistique – Asie Orientale* 26. 249–279. DOI: https://doi.org/10.3406/clao.1997.1516

Blum, Mirella L. 2021. On the nature of adjectives: evidence from Dinka. *Glossa* 6(1). 98. DOI: https://doi.org/10.16995/glossa.5765

Boersma, Paul & Weenink, David. 2021. *Praat: doing phonetics by computer* (version 6.1.55). http://www.praat.org.

Brunelle, Marc. 2012. Dialect experience and perceptual integrality in phonological registers: fundamental frequency, voice quality and the first formant in Cham. *Journal of the Acoustical Society of America* 131. 3088–3102. DOI: https://doi.org/10.1121/1.3693651

Connell, Bruce & Ladd, D. Robert. 1990. Aspects of pitch realisation in Yoruba. *Phonology* 7. 1–30. DOI: https://doi.org/10.1017/S095267570000110X

Denning, Keith. 1989. *The diachronic development of phonological voice quality, with special reference to Dinka and the other Nilotic languages*. Stanford (CA): Stanford University PhD dissertation.

DiCanio, Christian T. 2009. The phonetics of register in Takhian Thong Chong. *Journal of the International Phonetic Association* 39(2). 162–188. DOI: https://doi.org/10.1017/S0025100309003879

Dimmendaal, Gerrit J. & Jakobi, Angelika. 2020. Eastern Sudanic. In Vossen, Rainer & Dimmendaal, Gerrit J. (eds.), *Oxford Handbook of African Languages*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780199609895.013.74

Duerksen, John. 1994. *A review of Dinka orthography*. Unpublished manuscript, SIL.

Edmondson, Jerold A. & Esling, John H. 2006. The valves of the throat and their functioning in tone, vocal register and stress. *Phonology* 23(2). 157–193. DOI: https://doi.org/10.1017/S095267570600087X

Esling, John H. & Harris, Jimmy G. 2005. States of the glottis: an articulatory phonetics model based on laryngoscopic observations. In Hardcastle, William J. & Beck, Janet Mackenzie (eds.), *A figure of speech: a festschrift for John Laver*, 347–383. Mahwah (NJ): Lawrence Erlbaum.

Esposito, Christina M. & Sleeper, Morgan & Schäfer, Kevin. 2019. Examining the relationship between vowel quality and voice quality. *Journal of the International Phonetic Association* 51. 361–392. DOI: https://doi.org/10.1017/S0025100319000094

Garellek, Marc. 2019. The phonetics of voice. In Katz, William F. & Assmann, Peter F. (eds.), *The Routledge Handbook of Phonetics*, 75–106. Abingdon: Routledge. DOI: https://doi.org/10.4324/9780429056253-5

Garellek, Marc. 2020. Acoustic Discriminability of the Complex Phonation System in !Xóõ. *Phonetica* 77(2). 131–160. DOI: https://doi.org/10.1159/000494301

Garellek, Marc & Keating, Patricia. 2011. The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *Journal of the International Phonetic Association* 41. 185–205. DOI: https://doi.org/10.1017/S0025100311000193

Gope, Amalesh. 2021. The phonetics of tone and voice quality interactions in Sylheti. *Languages* 6. 1–18. DOI: https://doi.org/10.3390/languages6040154

Gordon, Matthew & Ladefoged, Peter. 2001. Phonation types: a cross-linguistic overview. *Journal of Phonetics* 29. 383–406. DOI: https://doi.org/10.1006/jpho.2001.0147

Hall, Beatrice L. & Ayom, Edward B. G. & Hall, R. M. R. 1982. Features for the vowels and voice qualities of Bor Dinka. Paper presented at the *Conference on Phonological/Distinctive Features*. State University of New York, Stony Brook.

Hombert, Jean-Marie. 1978. Consonant types, vowel quality and tone. In Fromkin, Victoria (ed.), *Tone: a linguistic survey*, 77–111. New York: Academic Press.

Hyndman, Rob & Athanasopoulos, George & Bergmeir, Christoph & Caceres Gabriel & Chhay, Leanne & Kuroptev, Kirill & O'Hara-Wild, Mitchell & Petropoulos, Fotios & Razbash, Slava & Wang, Earo & Yasmeen, Farah. 2024. forecast: Forecasting functions for time series and linear models. R package version 8.23.0.9000, https://pkg.robjhyndman.com/forecast/.

Iseli, Markus & Shue, Yen-Liang & Alwan, Abeer. 2007. Age, sex, and vowel dependencies of acoustic measures related to the voice source. *Journal of the Acoustical Society of America* 121(4). 2283–2295. DOI: https://doi.org/10.1121/1.2697522

Samely, Ursula. 1991. *Kedang (Eastern Indonesia), some aspects of its grammar*. Hamburg: Helmut Buske Verlag.

Keating, Patricia & Garellek, Marc & Kreiman, Jody. 2015. Acoustic properties of different kinds of creaky voice. *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, August 10–14 2015.

Kenstowicz, Michael J. 2014. Ablaut in Dinka nouns: a preliminary study. *Fleur de Ling, Tulane University Working Papers in Linguistics* 1. 83–91.

Keyser, Samuel Jay & Stevens, Kenneth N. 2006. Enhancement and overlap in the speech chain. *Language* 82. 33–63. DOI: https://doi.org/10.1353/lan.2006.0051

Kirby, James. 2018. Praat-based tools for spectral analysis. Version 0.2.4. https://github.com/kirbyj/praatsauce.

Kirk, Paul L. 1966. *Proto Mazatec phonology*. Seattle (WA): University of Washington PhD dissertation.

Kuznetsova, Alexandra & Brockhoff, Per B. & Christensen, Rune H. B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82(13). 1–26. DOI: https://doi.org/10.18637/jss.v082.i13

Ladefoged, Peter & Kirk, Paul & Blankenship, Barbara & Steriade, Donca. 1993. Mazatec (Jalapa de Diaz). UCLA Phonetics Archive. http://archive.phonetics.ucla.edu/Language/MAJ/MAJ.html.

Ladefoged, Peter & Maddieson, Ian. 1996. *The sounds of the world's languages*. Oxford: Blackwell.

Lüdecke, Daniel & Bartel, Alexander & Schwemmer, Carsten & Powell, Chuck & Djalovski, Amir & Titz, Johannes. 2024. sjPlot: Data Visualization for Statistics in Social Science. R package version 2.8.16. https://CRAN.R-project.org/package=sjPlot.

Lüdecke, Daniel & Ben-Shachar, Mattan S. & Patil, Indrajeet & Waggoner, Philip & Makowski, Dominique. 2021. performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software* 6(60). 3139. DOI: https://doi.org/10.21105/joss.03139

Malou, Job. 1988. *Dinka vowel system.* Dallas (TX): Summer Institute of Linguistics (SIL) and University of Texas at Arlington.

Mazaudon, Martine & Michaud, Alexis. 2008. Tonal contrasts and initial consonants: a case study of Tamang, a 'missing link' in tonogenesis. *Phonetica* 65(4). 231–256. DOI: https://doi.org/10.1159/000192794

McCloy, Daniel R. 2016. phonR: tools for phoneticians and phonologists. R package version 1.0-7. https://cran.r-project.org/web/packages/phonR.

Miller, Amanda L. 2007. Guttural vowels and guttural co-articulation in Ju|'hoansi. *Journal of Phonetics* 35. 56–84. DOI: https://doi.org/10.1016/j.wocn.2005.11.001

Mills, Timothy Ian Pandachuk. 2010. SpectralTilt-0.0.5. https://sites.google.com/a/ualberta.ca/timothy-mills/downloads.

Monich, Irina. 2020. Nuer tonal inventory. *Studies in African Linguistics* 49. 1–42.

R Core Team. 2021. *R: A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Redi, Laura & Shattuck-Hufnagel, Stefanie. 2001. Variation in the realization of glottalization in normal speakers. *Journal of Phonetics* 29. 407–429. DOI: https://doi.org/10.1006/jpho.2001.0145

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

Remijsen, Bert. 2014. Evidence for three-level vowel length in Ageer Dinka. In Caspers, Johanneke & Chen, Yiya & Heeren, Willemijn & Pacilly, Jos & Schiller Niels O. & van Zanten, Ellen (eds.), *Above and Beyond the Segments: Experimental linguistics and phonetics*, 246–260. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/z.189.20rem

Remijsen, Bert & Ayoker, Otto G. & Mills, Timothy. 2011. Shilluk. *Journal of the International Phonetic Association* 41(1). 111–125. DOI: https://doi.org/10.1017/S0025100310000289

Remijsen, Bert & Blum, Mirella L. & Pen de Ngong, Jon. 2022. A dataset on voice quality, tone and vowel quality in the Bor South dialect of Dinka. *Edinburgh DataShare*. DOI: https://doi.org/10.7488/ds/3493

Remijsen, Bert & Gilley, Leoma. 2008. Why are three-level vowel length systems rare? Insights from Dinka (Luanyjang dialect). *Journal of Phonetics* 36. 318–344. DOI: https://doi.org/10.1016/j.wocn.2007.09.002

Remijsen, Bert & Manyang, Caguor A. 2009. Luanyjang Dinka. *Journal of the International Phonetic Association* 39. 113–124. DOI: https://doi.org/10.1017/S0025100308003605

Roettger, Larry & Roettger, Lisa. 1989. A Dinka dialect study. *Occasional papers in the study of Sudanese languages* 6. 1–64.

Shue, Yen-Liang & Keating, Patricia & Vicenik, Chad & Yu, Kristine. 2011. VoiceSauce: a program for voice analysis. *Proceedings of the 17th International Congress of Phonetic Sciences*, 1846–1849.

Silverman, Daniel. 1997. Laryngeal complexity in Otomanguean. *Phonology* 14. 235–261. DOI: https://doi.org/10.1017/S0952675797003412

Storch, Anne. 2005. *The noun morphology of Western Nilotic*. Cologne: Rüdiger Köppe Verlag.

Thongkum, Theraphan L. 1988. Phonation types in Mon-Khmer languages. In Fujimura, Osamu (ed.), *Vocal fold physiology: voice productions, mechanisms and functions* (volume 2), 319–333. New York: Raven Press.

Traunmüller, Hartmut & Eriksson, Anders. 2000. Acoustic effects of variation in vocal effort by men, women, and children. *Journal of the Acoustical Society of America* 107(6). 3438–3451. DOI: https://doi.org/10.1121/1.429414

Tucker, Archibald N. 1936. The function of voice quality in the Nilotic languages. In Jones, Daniel & Fry, Dennis B. (eds.), *Proceedings of the 2nd International Congress of Phonetic Sciences*, 125–128. Cambridge: Cambridge University Press.

Tucker, Archibald N. 1975. Voice quality in African languages. In Ḥurreiz, Sayyid Ḥāmid & Bell, Herman (eds.), *Directions in Sudanese linguistics and folklore*, 44–57. Khartoum: University of Khartoum.

Tucker, Archibald N. & Bryan, Margaret A. 1956. *The Non-Bantu languages of North-Eastern Africa*. Oxford: Oxford University Press.

Turk, Alice E. & Nakai, Satsuki & Sugahara, Mariko. 2006. Acoustic segment durations in prosodic research: a practical guide. In Sudhoff, Stefan, Lenertová, Denisa & Meyer, Roland & Pappert, Sandra & Augurzky, Petra & Mleinek, Ina & Richter, Nicole & Schliesser, Johannes (eds.), *Methods in empirical prosody research*, 1–28. Berlin: De Gruyter. DOI: https://doi.org/10.1515/9783110914641.1

Venables, W. N. & Ripley, Brian D. 2002. *Modern Applied Statistics with S* (4th edition). New York: Springer. https://www.stats.ox.ac.uk/pub/MASS4/.

Warnes, Gregory R., & Bolker, Ben & Bonebakker, Lodewijk & Gentleman, Robert & Huber, Wolfgang & Liaw, Andy & Lumley, Thomas & Maechler, Martin & Magnusson, Arni & Moeller, Steffen & Schwartz, Marc & Venables, Bill & Galili, Tal. 2024. gplots: Various R programming tools for plotting data. https://cran.r-project.org/web/packages/gplots/.

Wickham, Hadley & Averick, Mara & Bryan, Jennifer & Chang, Winston & McGowan, Lucy D. & François, Romain & Grolemund, Garrett & Hayes, Alex & Henry, Lionel & Hester, Jim & Kuhn, Max & Pedersen, Thomas L. & Miller, Evan & Bache, Stephan M. & Müller, Kirill & Ooms, Jeroen & Robinson, David & Seidel, Dana P. & Spinu, Vitalie & Takahashi, Kohske & Vaughan, Davis & Wilke, Claus & Woo, Kara & Yutani, Hiroaki. 2019. Welcome to the Tidyverse. *Journal of Open Source Software* 4(43). 1686. DOI: https://doi.org/10.21105/joss.01686

Whalen, D. H. & Levitt, Andrea G. 1995. The universality of intrinsic F0 of vowels. *Journal of Phonetics* 23. 349–366. DOI: https://doi.org/10.1016/S0095-4470(95)80165-0

Xu, Yi. 1999. Effects of tone and focus on the formation and alignment of f_0 contours. *Journal of Phonetics* 27. 55–105. DOI: https://doi.org/10.1006/jpho.1999.0086