Phonology **36** (2019) 91–125. © Cambridge University Press 2019
doi:10.1017/S0952675719000058

*Ternary vowel length in Shilluk**

Bert Remijsen
Otto Gwado Ayoker
Signe Jørgensen
University of Edinburgh

Ternary or three-level vowel length is typologically rare, and supporting evidence is limited. This paper presents the results of an investigation into the hypothesised case of this configuration in Shilluk. We first describe the role of vowel length in Shilluk phonology and morphology, and then report on an acoustic study in which minimal sets for vowel length (short, long, overlong) are measured for vowel duration, coda duration, vowel quality and fundamental frequency. Short, long and overlong vowels differ significantly and substantially in terms of vowel duration: 96% of the items can be classified successfully for vowel length on the basis of this measurement alone. Of the other measurements, only vowel quality is significant, and this effect is considerably smaller. The mean values for vowel duration – 68, 111 and 150 ms for short, long and overlong vowels respectively – are similar to those reported for ternary vowel length in Dinka.

1 Introduction

1.1 Data and theory in relation to ternary vowel length

Ternary or three-level vowel length refers to a three-way phonological distinction between short, long and overlong vowels, whereby duration is the primary phonetic dimension of contrast, distinguishing vocalic nuclei within the syllable. Empirical evidence for this configuration in the world's languages is limited, to the effect that it has been assumed that only binary vowel-length contrasts are found, i.e. short *vs*. long (Kohler 2001, Duanmu 2009). Commenting on the lack of evidence for this configuration, Odden (2011: 466) writes that: 'for the vast majority of [languages for which ternary vowel length has been postulated], so little is known that it is difficult to say that length ternarity is a real phenomenon,

* E-mail: B.REMIJSEN@ED.AC.UK, OTTOAYOKER@GMAIL.COM, SIGNE.JORGENSEN@ED-ALUMNI.NET.

We thank the people who took part in data collection: Augustino Changjwok Otin, Elisabeth Antipas Onak, Francis Boyoumo Opiti, James Olwak Laa, Josephina Romano Oywac, Maria Bocay Onak, Peter Mojwok Yor, Stephen Okwen Agwet, Teresa Akic Awanh and Yoana Nyathaji Kolang. We are also grateful to SIL South Sudan for sponsoring our research visits to South Sudan and for hospitality in Juba. We gratefully acknowledge the Leverhulme Trust, which supported this research financially through the research grant 'A descriptive analysis of the Shilluk language' (RPG-2015-055).

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

rather than an epiphenomenon arising from binary length and some orthogonal dimension'.

Critical evaluations of hypothesised cases can be found in Odden (2011) and Prehn (2012). Both conclude that the analysis of ternary vowel length can be rejected for most languages for which it has been invoked, including Estonian, Low German, Mixe and Applecross Gaelic. However, there is one language for which both find that ternary vowel length is the best phonological analysis of the quantity contrast: Dinka, a Nilo-Saharan language spoken in South Sudan (Odden 2011: 486–487, Prehn 2012: 252–253). For this language, the hypothesis of ternary vowel length is supported by phonological and phonetic investigations of several dialects (Andersen 1987, 1990, 1992–94, Flack 2007, Remijsen & Gilley 2008, Remijsen 2013, 2014).

Dinka content words alternate in vowel length either between short and long grades, or between long and overlong grades (Remijsen & Gilley 2008). This is illustrated in Table I. Note how /lel/ 'isolate' alternates between short and long levels of vowel length. It is in the short grade in the 2nd singular, and in the long grade in the 3rd singular. In contrast, /leel/ 'provoke' alternates between long and overlong levels of vowel length. Again, it is in the short grade in the 2nd singular, and in the long grade in the 3rd singular. As a result, two levels of length are sufficient to represent vowel length in lexical roots, i.e. there is a contrast between roots with a short vowel, such as /lel/, and those with a long vowel, such as /leel/. The third level of vowel length is the result of morphological processes (cf. Dimmendaal 1995), so that, at the level of the surface phonology, lexical length and morphological length combine to create a ternary contrast. This is illustrated in Table I by /ă-lèl/, /ă-lèel/ and /ă-lèeel/. The sound files for the forms in Table I reveal how these forms differ in terms of vowel duration.[1]

A production study on ternary vowel length in Dinka (Remijsen & Gilley 2008) demonstrated on the basis of data from ten speakers that vowel duration is the primary phonetic correlate of the contrast. The mean values of short, long and overlong vowels were 72, 102 and 147 ms respectively in utterance-medial context, and the distributions were well separated. The contrast also had significant effects on coda duration and on vowel quality (F1, F2), but these were considerably smaller in size, thereby discounting alternative hypotheses involving consonantal quantity or the vowel inventory as orthogonal dimensions of contrast.

The controversy surrounding ternary vowel length relates to a more general debate in phonology, concerning the status of binarity in sound

[1] The following morpheme abbreviations are used throughout the paper: AMB = ambitransitive, APPL = applicative voice, ANTIP = antipassive, CS = construct state, DECL = declarative, DEM = demonstrative, EXIST = existential marker, FUG = centrifugal derivation, FUT = future, IMP = imperative, IMPF = imperfective, INV = inversion, ITER = iterative derivation, NEVP = non-evidential derivation, NOM = nominalisation, OV = object voice, PERT = pertensive, PET = centripetal derivation, PL = plural, PREP = preposition, SG = singular.

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

		morphological vowel length (grade)	
		short	long
lexical vowel length	short	ràaan ă-lèl person DECL.SG-isolate.2SG 'You are isolating somebody.'	ràaan ă-lèel person DECL.SG-isolate.3SG 'He is isolating somebody.'
	long	ràaan ă-léel person DECL.SG-provoke.2SG 'You are provoking somebody.'	ràaan ă-lèeel person DECL.SG-provoke.3SG 'He is provoking somebody.'

Table I

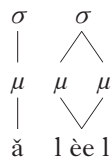
The interaction between lexical and morphological vowel length in Dinka, illustrated by a minimal set example, consisting of /lel/ 'isolate' and /leel/ 'provoke'. The embedded sound examples come from a speaker of the Twic (Rek) dialect of Dinka.

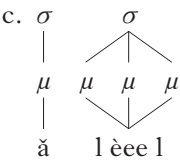
systems. In relation to tone, for example, inventories with three or more level tones have frequently been analysed using a nested combination of two binary features (Yip 1980, Bao 1999). However, such an architecture is ill-suited for representing not just rich inventories but also common contextual tone processes (Clements *et al.* 2010, Hyman 2010). With regards to metrical systems, constraints have been invoked with the specific aim of avoiding the postulation of ternary feet (e.g. Elenbaas & Kager 1999, Hermans & Torres-Tamarit 2014, Torres-Tamarit & Jurgec 2015). The appeal of binarity is also evident from the proposal within Government Phonology that all syllables are CV, i.e. involve binary branching between onset and nucleus (Lowenstamm 1996). In relation to syllables whose surface-phonological structure is more complex, this requires postulating empty nuclei. For example, the structure /takti/, ostensibly disyllabic, is analysed as trisyllabic, with an empty nucleus postulated following /k/ (Lowenstamm 1996). In each of these three cases, phenomena that are ostensibly ternary have been analysed in terms of binary constituency. In relation to vowel length, the same issue arises. It is clear that, across the world's languages, vowel length tends to involve a binary contrast. The question is whether this limit is absolute. The evidence from Dinka, in particular, challenges this restrictive perspective.

In the phonological analysis of Dinka, overlong vowels can be represented by associating three weight units or moras (μ) with the vocalic nucleus of overlong syllables (Hayes 1989, Remijsen & Gilley 2008). That is, vocalic quantity in the stem syllables in the Dinka minimal set for tone in Table I can be represented by association with one, two or three weight units, as in (1).

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

(1) a. 

b. 

c. 

However, just as in relation to tone, metrical structure and syllable constituency, one could postulate alternative representations, to the effect that the hypothesis that vowel length is maximally binary is not violated. Instead of the structures in (1b) and (1c) being distinguished in terms of the mora count of the stem syllable, they could, for example, be distinguished from one another by postulating mora-sharing with the coda consonant, an analysis similar to that postulated in relation to coda quantity in Saami by Baal *et al.* (2012).

In summary, the phenomenon of ternary vowel length raises two questions. First, can a language have a ternary contrast in vocalic quantity in its surface phonology? As seen from the discussion in Odden (2011), this question has not been settled conclusively, in spite of the evidence from Dinka, in particular because so many postulated cases can be analysed instead as a combination of binary vowel length and an orthogonal phonological dimension, such as vowel quality. Second, if a language has a ternary contrast in vowel length in the surface phonology, then what is the most compelling phonological representation? To address these two questions, we need more research on the phonetics, phonology and morphophonology of hypothesised cases. In relation to the first question, our phonetic investigations go beyond transcriptions, which are subjective and therefore cannot serve as a common ground when the facts themselves are in contention, as is the case in relation to ternary vowel length. A well-established way of making the analysis more accountable is to carry out acoustic measurements on a representative sample, which are then subject to statistical analysis. Also, by making the data publicly available, we aim to ensure that our analysis can be replicated and challenged. A further way to increase accountability is to include sound examples (Rice 2014). In this way, the speech phenomenon itself forms the basis of the report, rather than the transcription, which is, in its essence, a hypothesis. To address the second question, the issue of representation, we need detailed descriptive analyses on the role of the quantity contrast in the lexicon and in the morphology of languages that present hypothesised cases. In summary, the comprehensive investigation of ternary vowel length crosses the boundaries between phonetics, phonology and morphology.

Ternary vowel length has been postulated for three languages closely related to Dinka, all belonging to the West Nilotic subbranch within the Nilo-Saharan family: Shilluk (Tucker 1955, Remijsen *et al.* 2011, Remijsen *et al.* 2015), Thok Reel (Reid 2010) and Nuer (Monich 2017).

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

However, the quantity distinctions in these three languages remain understudied, especially in terms of their phonetic realisation.

This paper presents the results of an investigation into the phonetics, phonology and morphophonology of vowel length in Shilluk. The primary goal of the paper is to answer our first question, i.e. whether Shilluk has a ternary contrast in vocalic quantity in the syllable nucleus. With respect to the second question – what is the most compelling phonological representation – we will not evaluate the full range of options here. Instead, our aim is to provide a descriptive analysis of the role of vowel length in Shilluk phonology and morphology. Sound examples are embedded, and the dataset as a whole is publicly available (Remijsen *et al*. 2018).

The paper is structured as follows. In §2, we first describe the syllable-internal quantity contrasts that serve lexical and morphological functions in Shilluk in terms of a hypothesised ternary vowel-length contrast (§2.1). We then introduce an alternative analysis of the same phenomena (Gilley 2004), which maintains the hypothesis that vowel length is maximally binary (§2.2). §3 presents the results of an acoustic analysis of minimal sets, in which the hypothesis of ternary vowel length is put to the test. The conclusion and discussion follow in §4. The remainder of this introduction presents some background information on the Shilluk sound system, except for the hypothesised vowel-length contrast, which will be described in detail in the remainder of the paper.

1.2 Background information on the Shilluk sound system

This section presents a brief summary of the main characteristics of the Shilluk sound system, to facilitate the interpretation of forms throughout the paper. Shilluk content words have at their core a monosyllabic stem that reflects the root. This is illustrated by the examples in (2a).[2] With few exceptions, these stem syllables are closed. The stem can combine with one prefix, typically /a-/ or /ʊ-/, and up to two suffixes, as in (b). Polysyllabic word roots, such as /úgĭik/ in (c), conform to the template of affixed words, suggesting that they are either the diachronic result of derivational processes or formed to match this word-shape template.

[2] Our transcriptions follow IPA conventions. The following should be noted. First, segmentable morphemes or combinations of morphemes are separated by hyphens, in line with Comrie *et al*. (2015). Second, long and overlong vowels are represented by double and triple sequences of vowels, rather than using the [ˑ] and [ː] conventions respectively. Third, tone is transcribed using the conventions in (4) below. Note that various tone categories are represented using multiple tonal diacritics, either on a vowel symbol or distributed over vowel and coda. This is an ad hoc solution to the need to distinguish all of the tone categories; it does not prejudge the phonological analysis. For example, in the case of the High Fall to Mid, as in /kâ̌aaŋ̄/ 'trumpet.PERT.SG', the combination of diacritics denotes a syllable-level tone category realised as a high-to-mid falling contour tone which has the syllable as a whole as its domain of association. Crucially, it should not be interpreted as meaning that High Fall is associated with the vowel and Mid with the coda.

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

(2) a. dɔ́ɔc líŋ
good conflict

b. ɟâaaa̰k̄-ɔ̄ ó-cāaam á-māṯ-ì-á [á-māṯ-áa]
chief-SG FUT-eat.APPL PAST-drink-ITER-1SG

c. ógǐik
buffalo

The inventory includes ten vowel phonemes, five [−ATR] vowels (/ɪ ɛ a ɔ ʊ/), and their [+ATR] counterparts (/i e ʌ o u/) (Gilley 1992). Each [+ATR] vowel differs phonetically from the corresponding [−ATR] vowel primarily in terms of vowel height (F1), and to a lesser extent in terms of voice quality (Remijsen *et al*. 2011). [ATR] does not play a role in vowel harmony. However, there are several morphological processes in which [−ATR] changes to [+ATR] (Remijsen *et al*. 2015). Two of these are illustrated in (3): antipassive formation in (3a), and centripetal spatial deixis formation in (3b).

(3) a. á-mâaṯ á-mÂʌʌṯ
PAST-drink.OV PAST-drink.ANTIP

b. á-máaaṯ̀ á-mÁʌʌṯ̀
PAST-drink.FUG.OV PAST-drink.PET.OV

There are 19 consonantal phonemes. Voiceless stops, voiced stops and nasals are phonemic 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/. The inventory also includes the liquids /l r/ and the semivowels /w j/. There are no fricative phonemes. Consonant clusters within a syllable are limited to the onset of the stem syllable, where a semivowel can follow another consonant, as in /kwáac/ 'fish scales' and /rjɛ̄w/'second'. There are two additional phonotactic restrictions. First, dental and alveolar obstruents do not co-occur within the phonological word (cf. Mackenzie 2016).[3] Second, voicing in plosives is not contrastive in the stem-final consonant. In this position, plosives vary freely in voicing.[4]

The Shilluk tone inventory includes nine tone categories that are contrastive on stem syllables as a result of lexical and morphological specification. These categories are Low, Mid, High, Low Fall, High Fall, High Fall to Mid, Late Fall, Low Rise and High Rise. The contrast between High Fall and Late Fall is remarkable, in that these categories present a rare instance of contrastive alignment in contour tones (Remijsen &

[3] Remijsen *et al*. (2011: 113) note a few forms that violate this consonantal place harmony. However, we have since realised that these forms are actually harmonic, in line with Mackenzie (2016), and we take this opportunity to set the record straight. The correct transcriptions for the nouns in the dental row in (5) of Remijsen *et al*. are /n̪ú̱uu̱t̪̄/, /d̪ûuu̱t̪-ɔ̀/ and /t̪ùuu̱t̪/.

[4] Gemination in stem-final consonants is postulated in some earlier studies (Gilley 1992, Noske 1995). However, Remijsen *et al*. (2011) found no evidence for it in a quantitative study.

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

Ayoker 2014). Eight of the categories are illustrated in (4); see also note 2.[5] On affixes, the inventory is limited to Low, Mid and High.

(4) á-kɔ̀l (Low) PAST-disturb.2SG
á-kɔ̄l (Mid) PAST-disturb.APPL
á-kól (High) PAST-take out.APPL.2SG
á-kɔ̂l (Low Fall) PAST-take out.2SG
á-kɔ́l (High Fall) PAST-disturb.OV
á-kɔ́l̄ (High Fall to Mid) PAST-take out.APPL
á-kɔ́l̀ (Late Fall) PAST-take out.FUG
á-kɔ̌l (High Rise) PAST-disturb.APPL.2SG

2 The phonology of syllable-internal quantity in Shilluk

§2.1 presents a descriptive analysis of Shilluk's syllable-internal quantity contrasts at the lexical and morphological levels. Here we use the ternary vowel-length analysis to represent the patterns of quantity. In §2.2, we evaluate the competing hypotheses that have been put forward to analyse these phenomena.

2.1 Syllable-internal quantity in Shilluk phonology and morphophonology

The great majority of Shilluk content words display one of three different patterns of alternation with respect to vowel length in the stem syllable. These patterns are illustrated in (5) with transitive verb inflections. Here the lexical length of the root vowel is found in the stem in the past tense object voice form, whereas the past tense 2nd singular is an inflection that displays morphological lengthening of the stem vowel. First, some verbs have a short root vowel, but appear with increased vowel length in many inflections, including the past tense 2nd singular. This pattern of stem-internal alternation, which we refer to as short-with-grade, is illustrated in (5) by /càm/ 'eat'. In inflections such as the past tense object voice form, the stem vowel is in the short morphological grade, and in inflections such as the past tense 2nd singular it is in the long morphological grade. For short-with-grade verbs, the short and long morphological grades correspond to short and overlong levels of vowel length respectively. Second, some verbs have a long root vowel, and also undergo morphological lengthening. Verbs displaying this pattern of stem-internal alternation are the long-with-grade verbs, such as /lɛ̀ɛŋ/ 'throw'. Here again, the stem vowel is in the short grade in inflections

[5] The contrast between Low Rise and High Rise was not postulated in Remijsen *et al.* (2011); we have discovered it only recently. An example of a minimal pair is /pa̰ʌʌt̪/ 'twist rope using hands.ANTIP.IMP' *vs.* /pʌ̰ʌʌt̪/ 'fall.IMP'. In the base forms of nouns, Low Rise is found in e.g. /ŋo̰om/ 'awl' and /t̪a̰ak/ 'hat', and High Rise in /ɲa̰aaw/ 'domestic cat' and /mwɔ̰ɔɔl/ 'morning'.

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

such as the past tense object voice form, and in the long grade in inflections such as the past tense 2nd singular. For long-with-grade verbs, the short and long morphological grades correspond to long and overlong levels of vowel length respectively. Note that morphological lengthening, i.e. the distinction between the morphological short and long grades of the stem vowel, is 'overlengthening' in Shilluk: if a verb displays morphological lengthening, then it lengthens to a third level, i.e. overlong, irrespective of whether the root vowel is short or long, i.e. CVC, CVVC → CVVVC. In this respect, Shilluk is different from Dinka, where short and long stem vowels both lengthen morphologically by one level of vowel length, i.e. CVC → CVVC and CVVC → CVVVC (Andersen 1990). Third, certain transitive verbs have a short root vowel and do not display morphological lengthening of the stem vowel. The verb /ŋɔ̀l/ 'cut', for example, has a short vowel throughout its morphological paradigm. We refer to these as fixed short roots (cf. Remijsen & Ayoker 2014).

(5) *Transitive verbs*

	PAST OBJ VOICE	PAST 2SG	
short-with-grade	á-cə́m	á-càaam	'eat'
long-with-grade	á-lɛ́ɛŋ	á-lèɛɛŋ	'throw'
fixed short	á-ŋɔ́l	á-ŋɔ̀l	'cut'

The same quantity alternations appear elsewhere in the verb paradigm. Consider for example the antipassive. This derivation is marked by a package of stem-internal markers: quantity alternation, vowel raising and a change of coda /l r/ to /t/ (cf. Remijsen *et al*. 2016). The antipassives of the three verbs in (5) are /á-ŋùt/, /á-cʌ̀ʌʌm/ and /á-lèeeŋ/ respectively. Note that vowel length in the stem syllable in these antipassives is identical to the pattern in the past tense 2nd singular.

The same three patterns of quantity alternation are found in nouns. This is illustrated in (6) for suffixless singular nouns. In such nouns, the lexical length of the root vowel is found in the base form, which is the form used without any modifiers. As seen from (6), these base forms either have a short or a long vowel. Noun inflections such as the pertensive and the construct state display morphological vowel lengthening.[6] Similarly to the transitive verb paradigms, we find short-with-grade nouns, which alternate between short and overlong, long-with-grade nouns, which alternate between long and overlong, and fixed short nouns, which have a short stem vowel throughout.

[6] The pertensive (Dixon 2010: 268) – is used when a noun such as /káŋ/ 'trumpet' is modified by a possessor, as in /kə́aaŋ̄ kùl/ 'Kul's trumpet', and /káaaŋ ɟìɪ/ 'people's trumpet'. If the head noun is singular, the pertensive also marks the number of the possessor modifier. As seen in (6), with a singular possessor the pertensive never ends in a high-tone target, while with a plural possessor it invariably does.

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

(6) *Suffixless singular nouns*

	BASE	PERT SG	PERT PL	
short-with-grade	káŋ	kɑ́̂aaɪ̄ŋ	káaaŋ	'trumpet'
long-with-grade	gʌ̀ʌ̀t	gʌ̀ʌʌt	gɑ̌ʌʌt	'river bank'
fixed short	d̪ɔ́k	d̪ɔ́k̄	d̪ɔ́k	'mouth'

We also find singular nouns with a suffixed base form. Almost invariably, the suffix at issue is singular /-ɔ/. As seen in (7), we find the same three patterns of alternation here, albeit with a twist. Note that, in short-with-grade and long-with-grade paradigms, the overlong stem vowel is found in the base form, and the underlying length of the lexical root in the inflected form. That is, as compared to suffixless nouns, the length alternation in suffixed singular nouns applies in the opposite direction.

(7) *Suffixed singular nouns*

	BASE	PERT SG	PERT PL	
short-with-grade	pâaal-ɔ̀	pâl-ì	pâl-í	'knife'
long-with-grade	bɔ̀ɔɔ̀t̪-ɔ̀	bɔ̀ɔ̀t̪-ì	bɔ̀ɔ̀t̪-í	'craftsman'
fixed short	pîc-ɔ̀	pîc-ì	pîc-í	'tethering pole'

The patterns of length alternation interact in part with vowel quality. The short-with-grade verbs either have a low vowel (/ʌ a/) or a complex onset that includes a semivowel, as in /gwêt/ 'write' and /tjêk/ 'surround'. The fixed short verbs have monophthongal vowels other than /ʌ a/.[7] The long-with-grade vowels, finally, are found with any vowel quality.

In the patterns of length alternation surveyed so far, lexical items appear with either a short or a long vowel somewhere in their paradigm. In Dinka, this state of affairs generalises to the lexicon as a whole. On this basis, Dimmendaal (1995) postulates a binary vowel-length distinction for Dinka – i.e. short *vs.* long – at the level of lexical roots, and derives the third level of length in the morphology, through the addition of a prosodic morpheme that yields overlength. In contrast, Shilluk also has content-word morphemes that are only found with an overlong vowel, i.e. roots that are fixed overlong. (8) presents four examples of intransitive verbs that display this pattern. Because these intransitive verbs do not appear with a short or a long stem vowel anywhere in their morphological paradigm, there is no alternative to interpreting the overlong vowel as part of the lexical representation of the root, i.e. /jûuut̪/, /dʌ̂ʌʌr/, /bûuut̪/, /tûuul/. Most of these verbs are telic in a semantic sense, and this shared characteristic suggests that, diachronically, they are the result of a morphological process. Crucially, however, their paradigms offer no evidence for a morphological root with either a short or a long vowel.

[7] We know of one exception: /bác/ 'amniotic sac' is fixed short, even though the vowel is /a/. This may be a case of homophony avoidance; cf. /páac/ 'amniotic fluid'.

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

(8) *Intransitive verbs with an overlong root vowel throughout their paradigm*

PAST	NEVP	IMPF	FUT	INV	
á-jûuu̱t	jûuu̱t-ɔ̀	ɔ́-jûuu̱t-ɔ̀	ɔ́-júuu̱t	jú́uu̱t̄	‘finish’
á-dɅ̂ʌʌr	dɅ̂ʌʌr-ɔ̀	ɔ́-dɅ̂ʌʌr-ɔ̀	ɔ́-dɅ́ʌʌr	dɅ́ʌʌr̄	‘become tired’
á-bûuu̱t	bûuu̱t-ɔ̀	ɔ́-bûuu̱t-ɔ̀	ɔ́-búuu̱t	bú́uu̱t̄	‘become bored’
á-tûuul	tûuul-ɔ̀	ɔ́-tûuul-ɔ̀	ɔ́-túuul	tú́uul̄	‘come up’

The morphological role of vowel length is not limited to overlengthening. There is also evidence of morphological shortening, both in verbs and in nouns. But whereas overlengthening is inflectional, vowel shortening is derivational. We will show this first for transitive verbs, and then for nouns. In transitive verbs, morphological shortening is found in the formation of the antipassive derivation of certain long-with-grade verbs that have a Low Fall as their lexical tone.[8] This is illustrated in (9): the past tense subject voice form displays the lexical specifications for tone and vowel length of the root, which is CV̂VC (cf. Remijsen & Ayoker 2018). These long-with-grade verbs alternate inflectionally between long and overlong; this is illustrated by the past tense 2nd singular forms in (9). In the antipassive derivation, however, the vowel is short if the vowel of the root is high (closed), i.e. /ɪ i ʊ u/. If the vowel height of the root is not high, then antipassive formation involves increased quantity, as it does for all other verb classes. As a result of these inflectional and derivational processes, verbs such /bûuk/ ‘cover with powder’ and /cʊ̂ʊl/ ‘pay for’ appear with short, long and overlong stem vowels within their paradigm. The change of /ʊ/ to /u/ in the latter verb is predictable. The vowel qualities /ɪ ʊ e o/ are avoided in morphologically derived short vowels. That is, when the vowel is underlyingly either /ɪ/ or /e/, it shifts most often to /i/, and sometimes to /ɛ/. If it is either /ʊ/ or /o/, it usually shifts to /u/, and sometimes to /ɔ/.

(9) *Lengthening and shortening in long-with-grade Low Fall verbs*

PAST SUBJ VOICE	PAST 2SG	ANTIPASS PAST	
á-bûuk	á-bûuuk	á-bûk	‘cover with powder’
á-cʊ̂ʊl	á-cʊ̂ʊʊl	á-cût	‘pay for’
á-gûur	á-gûuur	á-gûr	‘grind’
á-mîɪn	á-mîɪɪn	á-mîn	‘pierce’
á-mêɛɲ	á-mêɛɛɲ	á-mèeeɲ	‘illuminate’
á-mâaa̱t	á-mâaaa̱t	á-mɅ̂ʌʌ̱t	‘drink’

Vowel shortening is regularly found in other morphological operations on transitive verbs as well, such as the derivations for benefactive and iterative (Remijsen *et al.* 2016). For example, long-with-grade /mâaa̱t/ has a short stem vowel in both – the past tense object voice form of the iterative

[8] The key motivation for treating operations such as antipassive, benefactive and iterative as derivations rather than inflections is that, in each of these derivations, we find inflectional paradigms for tense, aspect and modality, and for voice and subject marking, parallel to the corresponding inflectional paradigm of the transitive base.

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

derivation is /á-máṯ-ì/, and the past tense object voice form of the benefactive derivation is /á-mʌ́ṯ-ì/.

Outside the system of transitive verbs, morphological shortening of the vowel is also found in nouns, in particular in number marking. We provide some background information on this part of the grammar before discussing morphological shortening. Number marking is neither productive nor regular in Shilluk. As far as productivity is concerned, roughly 26% of singular nouns in our lexicographical data do not have a corresponding plural (our data includes 917 singular nouns and 678 plural nouns). As for regularity, number can be marked through suffixation and a variety of stem-internal changes (involving tone, vowel length, the stem-final consonant and vowel quality/breaking); which combination of these is involved is not predictable (Xu 2017). (10) presents a set of singular nouns that are morphologically unmarked, all of which contain a Low Fall and a long vowel, i.e. the same specification for tone and vowel length as the transitive verbs in (9).[9] The irregularity of number marking is illustrated by the fact that the nouns in (10b) display five different specifications for tone in the marking of plural, in additional to overlengthening. Because number marking is neither productive nor regular, it is best interpreted as a derivational rather than an inflectional phenomenon (cf. Haspelmath 1996: 47).

(10) *Vowel length and tone alternations marking number in* CV̂VC *nouns*

a. *shortening in plural*

SG	PL	
mîɪŋ	mìŋ	'deaf person'
pîɪl	pɛ̀l	'grindstone'
lûum	lùm	'grass'
tûuŋ	tùŋ	'horn'
kîic	kìc	'orphan'
lôʊl	lùl	'steep bank'

b. *lengthening in plural*

SG	PL	
dâak	dǎaak	'herd'
kɛ̂ɛl	kɛ̌ɛɛl	'cheetah'
mÂʌṯ	mÂ̌ʌʌṯ	'friend'
jîɪṯ	jìiiṯ	'well'
jêel	jêeel	'bracelet'
pôʊl	pǒʊʊl	'widest part of lower leg'

The importance of the data in (10) to the present discussion on quantity is that they show that the stem vowel can either lengthen or shorten in the

[9] It is not invariably the case that the singular is morphologically unmarked and the plural marked. Two other patterns are attested: (a) the plural can be unmarked with the singular being marked, and (b) both singular and plural can be marked. In the data in (10), it is unambiguously clear that the singular is unmarked and the plural marked, because only the latter displays floating quantity, a phenomenon only found in morphologically marked forms. Floating quantity lengthens the duration of a following vocalic prefix. For example, the marked plural noun /lùm/ 'grasses' has floating quantity. When it is followed by a past tense prefix, as in /lùm á-lîɪṯ-à/ 'I saw the grasses', the prefix has increased duration, i.e. [áalîɪɪɲ̣à] rather than [álîɪɪɲ̣à]. Presumably, floating quantity in these forms is a diachronic reflex of a lost plural suffix. An unmarked noun does not trigger such lengthening, even if its phonological form is identical in terms of vowel length and tone. This is illustrated by /tìm/ 'forest' in /tìm á-lîɪṯ-à/ 'I saw the forest'. The unmarked singular /tìm/ does not have floating quantity: there is no increased duration of the following past tense prefix, which is realised as [álîɪɪɲ̣à].

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

morphological marking of number. Shortening is illustrated by the nouns in (a): these illustrate a decrease in vowel length in the marking of plural number. (As explained in relation to vowel shortening in transitive verbs, morphological shortening of the root vowel is accompanied by raising or lowering if the root vowel is one of /ɪ ʊ e o/.) Importantly, shortening of the stem vowel is derivational rather than inflectional. That is, the inflectional paradigm of nouns only includes the quantity-increasing alternations illustrated in (6) and (7), i.e. fixed short, short-with-grade and long-with-grade. By implication, all of the singular nouns in (10) display overlengthening in the formation of the pertensive, construct state and demonstrative inflections, as they have a long vowel in the base. For example, the demonstrative inflection of /lûum/ is /lúuum̀/. Just as in transitive verbs, then, morphological shortening in nouns is derivational, not inflectional. Also, in nouns and verbs alike, morphological shortening of the stem vowel is found only in stems that have a high (closed) or a mid (half-open) vowel underlyingly, and such roots can appear with all three levels of vowel length in their paradigm.

In summary, this descriptive analysis of Shilluk quantity shows that the hypothesised ternary vowel-length contrast has both lexical and morphological functions. Lexically, short and long vowels are commonly found in roots. Roots with an overlong vowel are also attested, but they are few in number (see (8)). Presumably, such roots represent the diachronic outcome of valency-increasing processes whereby the base has subsequently been lost. Morphologically, we find the following three alternations: CVC ~ CVVC, CVC ~ CVVVC, CVVC ~ CVVVC. The latter two patterns have the highest functional load, and are inflectional in nature. That is, short and long root vowels both alternate between their lexically specified vowel length on the one hand and overlength on the other, in various inflections and derivations (CVC ~ CVVVC, CVVC ~ CVVVC). In addition, long lexical roots with a Low Fall tone display shortening of the stem vowel (CVC ~ CVVC). Across word classes, vocalic shortening is invariably derivational in nature. In conclusion, the hypothesised ternary vowel-length contrast serves to distinguish unrelated lexical forms, and also to mark a wide range of morphological processes.

2.2 Competing hypotheses

In §2.1, we described Shilluk syllable-internal quantity in terms of the ternary vowel-length hypothesis (3VL). In this section, we provide some context on the study of syllable-internal quantity in Shilluk, and we evaluate an alternative hypothesis, which maintains the axiom that vowel length is maximally binary.

To the best of our knowledge, Shilluk is actually the first West Nilotic language for which 3VL was proposed. Three decades ahead of Andersen's (1987) paper postulating this configuration for Dinka, Tucker (1955: 427) makes the following observations:

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

> Vowel length is an extremely elusive phenomenon in Shilluk. In some words the vowels are pronounced very short indeed; in others they are pronounced very long (and are here doubled). In many more words, however, length seems to lie between the very short and the very long, but the number of degrees of significant variation cannot yet be determined.

Crucially, and in spite of the tentative nature of these initial comments, Tucker distinguishes three levels of vowel length throughout the paper, using the IPA half-long character to transcribe long vowels. Later in the same paper, he emphasises the importance of this three-way distinction, i.e. 'very short, medium, very long' in morphophonological alternations (1955: 432).

The 3VL hypothesis for Shilluk represents lexical and morphological contrast in terms of the same phonological dimension, i.e. vowel length (cf. §2.1). This is crucially different from the analysis of syllable-internal quantity in Shilluk presented in Miller & Gilley (2001, 2007) and Gilley (2004) (the latter extends the analysis to Dinka). In their analysis, syllable-internal quantity in Shilluk is represented with a combination of (a) a binary vowel-length contrast, marking lexical contrast, and (b) a stress contrast, marking morphological contrast. We refer to their hypothesis as '2VL + stress'. The stress contrast is invoked to represent morphological quantity alternations, in such a way that stress is postulated in some morphologically unmarked forms, whereas some morphologically marked forms are unstressed (Gilley 2004: 118). For example, the past tense object voice form, which we hypothesise to display the quantity of the root (cf. (5)), has a stressed stem syllable according to the 2VL + stress hypothesis (Miller & Gilley 2001). The realisation of stress is described as follows:

> When a vowel is stressed, the syllable shortens to about half of its original length. Thus, a long stressed vowel will be about the same length as a short unstressed vowel. The short stressed vowel will become extra short. In this way, vowels in Shilluk and Dinka sound as though they exhibit three different vowel lengths. (Gilley 2004: 104)

The different lengths and stress patterns are shown schematically in (11) (from Gilley 2004: 104).

(11) short stressed '|__|
short unstressed |____|
long stressed '|____|
long unstressed |_____|

But this apparent ternary contrast is represented by the combination of a binary vowel-length contrast and another phonological dimension (cf. Odden 2011: 466). The primary phonetic realisation of the hypothesised stress feature is anomalous: across the world's languages, greater duration of segments in stressed syllables than in unstressed syllables is the most

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

reliable acoustic correlate of stress (see e.g. Ortega-Llebaria & Prieto 2011, van Heuven 2014). The primary correlate of the stress feature postulated in Gilley (2004) is the opposite: stressed syllables display reduced vowel duration.

Apart from vowel duration, Gilley mentions several other phonetic characteristics of the hypothesised stress contrast. She describes stressed syllables as being followed by a pause, and as being 'fortis'. The phonetic meaning of the fortis–lenis dimension is not explained, except for the statement that stressed syllables are not perceptibly louder (Gilley 2004: 104). She also states that stress affects tone, suggesting that this may be due to time pressure. Elsewhere, she describes unstressed syllables as more peripheral (2004: 102). In §3 we report on a detailed investigation of the phonetic realisation of the quantity contrast.

We can compare the competing hypotheses – 3VL *vs.* 2VL + stress – on the basis of evidence from antipassives. Antipassives are derived from transitive verbs through a range of suprasegmental changes, of which change in quantity is one. In §2.1, we represented the changes in terms of 3VL. In the 2VL + stress model, antipassive verbs are invariably unstressed, whereas the corresponding transitive verb forms are stressed (Miller & Gilley 2001: 43). (12a) presents a lexical minimal set for vowel length, consisting of the transitive verbs /nàŋ/ 'lick' *vs.* /nâaŋ/ 'seal, fill'.[10] Both of these verbs display increased quantity in the antipassive. If the change in quantity is realised by a stress feature, in line with the 2VL + stress hypothesis, then the antipassive forms of /nàŋ/ and /nâaŋ/ should remain distinct, as each would increase in vowel length by one level. That is, if antipassive formation is marked by the loss of the hypothesised stress feature, a short unstressed vowel would have the same duration as a long stressed vowel. In turn, a long unstressed vowel should display a vowel that has a saliently greater duration. According to the 3VL hypothesis, morphological lengthening is overlengthening (cf. §2.1). This means that both verbs with a lexically short vowel and those with a lexically long vowel display an overlong vowel in the antipassive, thereby neutralising the lexical contrast.

[10] Shilluk displays topic drop, such that preverbal internal arguments can be omitted if they have a 3rd person singular referent and can be inferred from the preceding context. As a consequence, the utterances in (12) are grammatically well-formed clauses.

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

(12)		3VL	2VL + stress		
a.	{nàŋ}	á-nā̋ŋ	á-ˈnāŋ	PAST-lick.OV	'it was licked'
	(short-with-grade)	á-nʌ̀ʌʌŋ	á-nʌ̄ŋ	PAST-lick.ANTIP	's/he was licking'
	{nâaŋ}	á-nâaŋ	á-ˈnāaŋ	PAST-seal.OV	'it was sealed'
	(long-with-grade)	á-nʌ̀ʌʌŋ	á-nʌ̄ʌŋ	PAST-seal.ANTIP	's/he was sealing'
b.	{càm}	á-cā́m	á-ˈcām	PAST-eat.OV	'it was eaten'
	(short-with-grade)	á-cʌ̀ʌʌm	á-cʌ̄m	PAST-eat.ANTIP	's/he was eating'
	{cā́am}	á-cā́am	á-ˈcāam	PAST-feed.OV	'it was fed'
	(long-with-grade)	á-cʌ̀ʌʌm	á-cʌ̄ʌm	PAST-feed.ANTIP	's/he was feeding'

The sound files for the forms in (12a) reveal that the two antipassive forms do not sound different in terms of duration. Also, the stem vowels of both antipassive forms are longer than the stem vowels of the corresponding transitive verb forms. A second example illustrating the same point is presented in (12b).[11]

The 2VL + stress hypothesis runs into further difficulties. If the morphological marking of antipassive through quantity is represented as loss of stress, we should find greater vowel duration in the antipassive for all transitive verbs. The data contradict this in two ways. First, fixed short verbs do not display any quantity alternation: e.g. /á-ŋɔ́l/ 'PAST-cut' → /á-ŋùt/ 'PAST-cut.ANTIP'. Second, as shown in (9) above, the members of the long-with-grade Low Fall class which have a closed vowel, i.e. /ɪ i u ʊ/, have a stem vowel with reduced quantity in the antipassive. (13) presents transcriptions of an example in terms of the competing hypotheses. As heard in the sound files, the antipassive of /gûur/ 'grind' has a shorter vowel than the corresponding transitive.

(13)	3VL	2VL + stress		
{gûur}	á-gū́ur	á-ˈgūr	PAST-grind.OV	'it was ground'
(long-with-grade)	á-gûr	á-gūr	PAST-grind.ANTIP	's/he was grinding'

The problem for the 2VL + stress approach in this context is that morphological quantity alternations in the stem vowel are both increasing and decreasing in quantity. For example, relative to the past tense subject voice form /á-gûur/ 'PAST-grind', we find an increase in stem vowel length in the 2nd singular inflection, /á-gûuur/, but a decrease in the antipassive derivation, /á-gûr/. This means that a binary stress contrast is by its nature insufficient to represent the morphological quantity alternations in Shilluk.

[11] The verbs /càm/ 'eat' and /cā́am/ 'feed' are undoubtedly related; this causative derivation is sporadic and rare.

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

In summary, the discussion of the role of quantity in antipassive formation demonstrates that the representation of morphological quantity as stress fails to correctly predict the morphological marking of antipassives for fixed short verbs and short-with-grade verbs, and for long-with-grade Low Fall verbs. This set of cases that 2VL + stress cannot account for makes up the majority of the transitive verbs.

An unrelated problem with the 2VL + stress hypothesis is that stress has been used to represent alternations that do not involve a change in quantity. (14) displays two examples which show the role of stress in a syntactic voice alternation, from Gilley (2004: 107). The preverbal topic slot accommodates the thematic subject in (14a), but the beneficiary in (14b). Stress is hypothesised to be the sole dimension of morphophonological contrast. Given that 2VL + stress postulates that stress is associated with decreased vowel duration, Gilley's transcription predicts that the stem vowel of the verb has an audibly shorter duration in (14b) than in (14a). In our own transcriptions, also given in (14), we represent this as an alternation of tone rather than quantity.[12]

(14) a. 2VL + stress ānéj á-cwōttì kī ḏ̱áṉ-ɔ̀
3VL ànéj á-cwòtì kí ḏ̱ɑ́ɑaṉ-ɔ̀
Aney PAST-call.AMB PREP person-SG
'Aney called a person.'

b. 2VL + stress ānéj á-ˈcwòttì kī ḏ̱áṉ-ɔ̀
3VL ànéj á-cwôtì kí ḏ̱ɑ̂ɑaṉ-ɔ̀
Aney PAST-call.AMB.OV PREP person-SG
'Somebody called a person for Aney.'

The sound files demonstrate that the stem syllables of the verb sound different in terms of melody, i.e. tone: the stem syllable has a falling contour tone in (14b), whereas it has a level low pitch in (14a). In contrast, the verbs are not saliently different in terms of quantity.

In addition to this language-internal evidence, the 2VL + stress hypothesis is also problematic in terms of the structural characteristics of the stress feature. In a widely accepted interpretation of the concept, stress is (a) culminative and (b) obligatory (Hyman 2006). As a culminative feature, stress singles out one syllable in a string. As an obligatory feature, there should be no phonological words that are not marked for stress. The hypothesised stress feature of Miller & Gilley (2001, 2007) and Gilley (2004) displays neither of these properties. Intransitive verbs, for example, are hypothesised not to carry stress (Miller & Gilley 2001), while monosyllabic words can carry stress: e.g. /ˈlwák/ 'cattle byre' (Gilley 2004: 108). The hypothesised stress feature is equally anomalous

[12] When the ambitransitive form of the verb has a Low stem syllable in the morphologically unmarked form, i.e. subject voice, then the voice alternation in which a thematic role other than the thematic subject appears in the topic slot is marked by a Low Fall on the stem syllable of the verb.

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

in terms of its phonetic realisation: it is typologically unexpected for stress to be marked by a reduction in vowel duration.

In summary, the 2VL + stress hypothesis can be rejected on language-internal and on theoretical grounds. First, it does not make correct predictions with respect to the morphological alternations in which it is hypothesised to be involved, such as antipassive and syntactic voice alternations. Second, it is anomalous.

3 An acoustic study of ternary vowel length in Shilluk

We put the 3VL hypothesis to the test in a production study in which the hypothesised vowel-length contrast was manipulated, while other factors were controlled. The 3VL hypothesis predicts that vowel duration is the primary correlate in three-way minimal sets, and that any other correlate will play a lesser role. The methodology is described in §3.1, and the results are reported in §3.2.

Aside from vowel duration, we examine coda duration, first and second formants (F1, F2) and fundamental frequency (F0). We include these additional measurements because, if any measurement other than vowel duration were to reliably distinguish any pairwise combination of the levels of the contrast under investigation, this would offer support for an alternative interpretation involving a phonological dimension orthogonal to vowel length, as suggested in Odden (2011: 466). For example, if the hypothesised levels of short *vs.* long vowels can be differentiated reliably by F1 or F2, then this half of the contrast can be represented in terms of phonemic vowel quality, and a binary vowel-length contrast is adequate.[13] The orthogonal dimension postulated in Gilley (2004) is stress, and she mentions centralisation of vowels, i.e. vowel quality, and tone as phonetic characteristics of the stress contrast (Gilley 2004: 102).[14] Measuring F0, F1 and F2 allows us to evaluate this hypothesis. We also measure coda duration, because gemination of stem-final consonants is postulated in earlier work on Shilluk (Gilley 1992, Noske 1995), and also because coda duration may show a complementary interaction with a vowel-length contrast, with consonant duration being greater after short vowels than after long vowels (Elert 1964, Schaeffler *et al.* 2002 on Swedish; Letterman 1994 on Sinhala; Yu 2008 on Washo).

3.1 Methods

3.1.1 *Target words and contexts.* The target words are listed in Table II. Two three-way minimal sets for vowel length are included for each of four vowels: /ɛ a ɔ u/. In each set, vowel length marks lexical and/or

[13] This is the analysis proposed in relation to the controversy of ternary vowel length in Low German by Kohler & Tödter (1984) and Kohler (2001).

[14] Gilley (2004) also notes that there may be a pause following a word that ends in a stressed syllable. We have not found this to be the case; in our dataset there are no pauses following the target words.

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

morphological contrast. All of the targets are nouns. Many are agent-oriented nominalisations derived from transitive verbs; these display the same inflectional paradigm as non-derived nouns. The third level of vowel length is invariably represented through the pertensive inflection, an instance of head-marking in possessive noun phrases (cf. note 6).

	short	long	overlong
/ɛ/	lɛ́ŋ 'beat drum.NOM' kɛ́l 'peel.NOM'	lɛ́ɛŋ 'throw.NOM' kɛ́ɛl 'separate.NOM'	lɛ́ɛɛŋ 'throw.NOM.PERT.PL' kɛ́ɛɛl 'separate.NOM.PERT.PL'
/a/	pál 'dodge.NOM' lám 'pray.NOM'	páal 'surgery knife' láam 'prayer'	páaal 'surgery knife.PERT.PL' láaam 'pray.NOM.PERT.PL'
/ɔ/	kɔ́l 'take out.NOM' t̪ɔ́l 'eat prematurely.NOM'	kɔ́ɔl 'herd.NOM' t̪ɔ́ɔl 'rope'	kɔ́ɔɔl 'herd.NOM.PERT.PL' t̪ɔ́ɔɔl 'rope.PERT.PL'
/u/	dúp 'mess up.NOM' gút 'stab.NOM'	dúup 'mouse' gúut 'navel'	dúuup 'mouse.PERT.PL' gúuut 'navel.PERT.PL'

Table II
The target words used in the production study.

The target words were elicited in utterance-medial position, using the frames illustrated in (15) for forms with short, long and overlong vowels. For all targets, the clause structure used is that of the existential predicate, in which the target is preceded by the existential marker /dâa/. In the short and long conditions (15a, b), the target nouns are in the base form and followed by a prepositional phrase. In the overlong condition (15c), the following constituent is a possessor modifier, which expresses the agent of the referent event of the source verb.[15]

(15) a. dâa lám kì kɛ̂ŋ 'There is praying here.'
EXIST pray.NOM PREP place.CS.DEM

b. dâa láam kì kɛ̂ŋ 'There is a prayer here.'
EXIST prayer PREP place.CS.DEM

c. dâa láaam ɟìɪ 'There is praying by people.'
EXIST pray.NOM.PERT.PL people

[15] The carrier sentence for the overlong test word in (15c) is one syllable shorter than the carrier sentences in (a) and (b). This is potentially relevant, because, cross-linguistically, segmental durations are shorter in longer constituents than in short ones (e.g. Nooteboom 1972). However, the difference is only salient between one-word and two-word sentences (Windmann *et al.* 2015). To the best of our knowledge, there is no evidence suggesting that the difference between three and four syllables has a significant influence on segment durations.

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

We also included in the design three items with an overlong vowel /i/ in the stem syllable: /kíiic/ 'bee.PERT.PL', /àbíiik/ 'sorghum flour.PERT.PL' and /míiilɔ̀/ (name). These items are not taken into account in the quantitative analyses. Instead, we use them to display the formant measurements on the target vowels in the context of the vowel space as a whole. Doing so requires measurements of the corner points of the vowel space, i.e. /i a u/. Of these three, /i/ is not one of the vowels under investigation.

3.1.2 *Speakers.* The materials were elicited from eight native speakers of Shilluk (four men, four women), ranging in age between 36 and 65 (the second author was one of the speakers). Six of them grew up in the southern part of the Shilluk region, known as Lwak, and the other two in the northern part, Gar. To the best of our knowledge, dialect is irrelevant to the phenomenon under investigation. All were resident in Juba, the capital of South Sudan, where the recordings were made. The speakers use Shilluk on a daily basis. Three other speakers from whom we started eliciting the dataset did not produce the utterances fluently, typically pausing between words. We attribute this to the controlled nature of the task, which is unlike normal speech communication. In those cases, data collection was ended prematurely, and these incomplete data were not processed or analysed.

3.1.3 *Elicitation procedure.* The carrier sentences with the target words embedded in them were divided into two groups: those with a target word based on noun roots, and those with a target word based on verb roots. Within each of these groups, the target words were ordered in such a way that any minimal sets – e.g. /kɔ́l/ 'take out.NOM' *vs.* /kɔ́ɔl/ 'herd.NOM', /kɔ́ɔɔl/ 'herd.NOM.PERT.PL' – were well separated from one another, with forms based on two or more lexical roots intervening. For each lexical noun or verb, the elicitation was built up from the base form, elicited in utterance-final position in an existential predicate, e.g. /dâa kɔ́ɔl/ 'there is herding'. From this starting point, elicitation moved towards the intended form and context, by asking the speaker to (a) add the location 'in this place', or (b) add a possessor. In the latter case, we started with a singular possessor, and from there onwards to the intended plural possessor, which would trigger a High tone on the target. A few fillers involving targets with Low and Low Fall tones were mixed in, so that the speakers did not invariably have a High tone in the target slot of the carrier sentence. The order in which the sentences were collected was reversed for half of the speakers, for both the noun and the verb blocks, yielding in all four different orders in which the data were collected (nouns forward and verbs forward; nouns forward and verbs backward; nouns backward and verbs forward; nouns backward and verbs backward). Both the noun and verb blocks were preceded by a practice item.

We elicited the data through spoken interaction, rather than using written stimuli, because the great majority of Shilluk native speakers are not literate in their first language. In these interactions, we did not

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

produce the target word, in order not to bias the speakers' pronunciation. Instead, we elicited the target words by asking the speakers to translate from English, in which all speakers had some proficiency. If necessary, we described the target meaning in Shilluk using synonyms. There were only two following contexts (cf. (15) above), so that the speakers were quickly familiar with them. We also used gestures: a circular movement of the index to ask for a repetition, and pointing downwards to elicit the /kì kɛ̂ŋ/ 'here' context. Elicitation was carried out by the first and second authors jointly, and both monitored the responses for hesitations and errors. For each type, we collected two realisations, so as to ensure that we would have at least one usable token, even if one of the realisations was affected by, for example, background noise.

3.1.4 *Processing*. Praat (Boersma & Weenink 2015) was used both for data processing and for scripted measurements. In the first step of processing, we extracted the realisations (i.e. carrier utterances with embedded targets) from the raw recording files, and stored these as separate sound files. In these sound files, we segmented the onset, nucleus and coda of the target syllable. Our primary heuristic in locating the boundary between a vowel and a sonorant consonant (nasal or [l]) was F2. Concretely, we looked for a sharp decrease in darkness in the spectrogram in the frequency range of the F2 for the vowel in question. This change was typically matched by a discontinuity in the waveform envelope, which was also displayed. The same heuristic enabled us to determine the location of the boundary between a vowel and the closure phase of a following plosive. In establishing the boundary between a vowel and a preceding plosive, the voice onset time (VOT) lag was segmented with the plosive, rather than with the vowel.

We collected the following measurements: duration of the nucleus, duration of the coda, vowel quality (F1, F2) and fundamental frequency. In relation to F1 and F2, we created the formant candidates using the Burg algorithm (Childers 1978), applied Praat's tracking algorithm to find optimal formant paths and then visually checked the resulting F1 and F2 traces, particularly at the temporal midpoint of the vowel, where the formant measurements were taken. In 13 cases, the resulting measurements for F1 or F2 were not realistic, given our auditory impression of the vowel quality. Nine of these came from the same speaker, and ten involved the vowel /u/. Nine of these cases were corrected by re-running the measurement procedure with vowel-specific formant tracking values for F1 and F2, e.g. 300 Hz and 800 Hz if the vowel was /u/. In four cases, this still did not yield realistic values, and these tokens were discarded. For F0, we created the traces, and checked these visually. Just as for the formants, the F0 measurement was taken at the temporal midpoint. There were no tracking errors for F0.

Ahead of the descriptive and inferential statistics, we aggregated the measurements over repetitions, so as not to treat repetitions as

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

independent observations. This yielded 192 types from the original 361 tokens, the expected total (eight minimal sets × eight speakers × three members).

3.1.5 *Statistical analyses*. If the 3VL hypothesis is correct, then short, long and overlong vowels differ primarily and substantially in terms of vowel duration. The other measures (coda duration, F1/F2 and F0 at the temporal midpoint of the vowel) should only play a secondary role, or no role at all. In the analysis of the results, we can evaluate whether this is the case, in the first place by examining the descriptive statistics for each of these four measures. We use the mean as a measure of central tendency, and the standard deviation as a measure of dispersion around the mean.

In addition, we evaluate this hypothesis through an inferential test procedure: a linear mixed-effects analysis combined with likelihood ratio testing (Bates 2005, Baayen *et al*. 2008, Winter 2013). The crucial advantage of using a linear mixed-effects analysis is that we can treat both subjects and items as random factors.

For each measure, we start out from a model in which Vowel length and any other factor that is likely to affect that measure are included, as well as any interaction between the main effects. Then we take out interactions and factors from the model, one at a time. At every step, we examine whether the more richly specified model is significantly different from the simpler model. If it is, then the factor or interaction at issue is significant. If a factor with more than two levels is significant, then we examine the pairwise comparisons for significance, using *t*-tests with Bonferroni correction.

In the linear mixed-effects models, both subjects and items were modelled as random slopes, expressing the interaction between each of these two random factors and Vowel length, which is the main factor in question in this production study. Modelling the random factors in this manner is conservative, allowing for more of the variance in the dependent to be explained by the random factors (Barr *et al*. 2013, Winter 2013). We also take a conservative approach in setting the significance threshold (α) at 0.01, i.e. lower than the usual level of 0.05 (see Cohen 1994 for a critical evaluation of this threshold value). Both in the modelling of the random factors and in setting the significance threshold, we approach the inferential statistics in a conservative manner, because we are looking for evidence of a phonemic distinction: the primary correlate of a phonemic contrast should condition a highly significant effect in inferential testing.

Finally, we examined the value of the measurements as potential correlates to the quantity contrast under investigation by means of Linear Discriminant Analysis (LDA). LDA is insightful here because statistical significance is a insufficient as a heuristic for phonemic contrast. That is, inferential tests such as the likelihood ratio test yield a significant result if a given factor or interaction merely influences a measure beyond chance. However, when a measure is the primary correlate of a given

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

phonological contrast, we can expect it to reliably predict category membership. LDA allows us to find out for each item whether its factor level in terms of the hypothesised 3VL contrast (short, long or overlong) can be predicted on the basis of a given measurement or combination of measurements. Therefore, a high correct classification result in an LDA implies that the factor levels can be distinguished reliably on the basis of the measurement(s) in question.

We used R (R Core Team 2016) to carry out all of the descriptive and inferential statistics, and also to produce the graphs. The specific packages we used were *gplots* (Warnes *et al*. 2016) for making graphs showing means and standard deviations, *lme4* (Bates *et al*. 2015) for linear mixed-effects analyses, *phonR* (McCloy 2016) for *z*-transformation and plotting Fig. 3, and *MASS* (Venables & Ripley 2002) for LDA.

3.2 Results

3.2.1 *Vowel duration*. Figure 1a presents the descriptive statistics for vowel duration by Vowel length, over speakers and items. The mean durations for short, long and overlong vowels are 68, 111 and 150 ms respectively. In this figure, the whiskers encompass one standard deviation above and below the mean, i.e. 68% of the distribution. Importantly, these standard deviations around the mean do not overlap either between short and long or between long and overlong. This suggests that the three levels of Vowel length are produced with little overlap in terms of vowel duration.

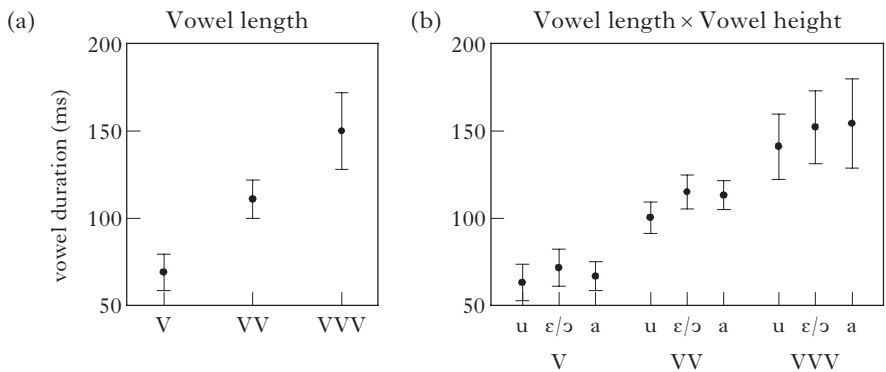


Figure 1

Means and standard deviations for vowel duration (ms) over speakers and items; (a) by Vowel length (short (V) *vs*. long (VV) *vs*. overlong (VVV)), and (b) by Vowel length × Vowel height (/u/ *vs*. /ɛ ɔ/ *vs*. /a/).

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

The variability around the mean in Fig. 1a contains variability as a function of various factors. The most important of these are the rate of speech of the eight speakers and the height of the vowel: across languages, low (more open) vowels tend to have greater duration than high (more closed) ones (e.g. House 1961). In line with this cross-linguistic tendency, Fig. 1b shows that the high vowel /u/ has the shortest duration, across all three levels of Vowel length. The half-open vowels /ɛ ɔ/ and the low vowel /a/ do not show a consistent pattern in terms of vowel duration. As for rate of speech, the slowest and fastest individuals among the speakers, in terms of the mean duration of the syllable nucleus across levels of Vowel length, differ by 21 ms.

We examined the significance of Vowel length (short *vs.* long *vs.* overlong) and Vowel height (/u/ *vs.* /ɛ ɔ/ *vs.* /a/) as fixed factors determining vowel duration in a linear mixed-effects model with random factors Speaker and Item. The effect of Vowel length is highly significant in the likelihood ratio test ($\chi^2(1) = 41.69$, $p < 0.0001$). Because the factor Vowel length has three levels, we explored this factor further through pairwise comparisons between factor levels. We found that all combinations (short *vs.* long, short *vs.* overlong, long *vs.* overlong) are highly significant ($p < 0.0001$). In contrast, the effect of Vowel height is not significant ($\chi^2(1) = 1.05$, $p = 0.31$), and the interaction between Vowel length and Vowel height does not have a significant effect on vowel duration ($\chi^2(1) = 3.86$, $p = 0.0494$) (α is set at 0.01; cf. §3.1.5).

3.2.2 *Coda duration.* The descriptive statistics for coda duration are presented in Fig. 2. The mean values for short, long and overlong vowels as a

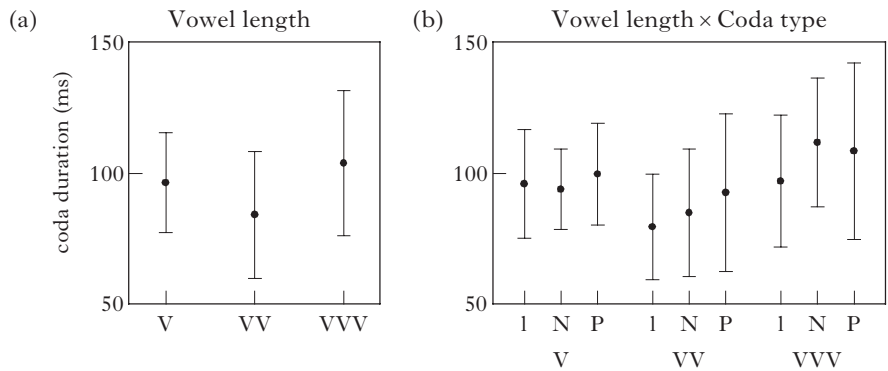


Figure 2
Means and standard deviations for coda duration (ms) over speakers and items; (a) by Vowel length, and (b) by Vowel length × Coda type (/l/ *vs.* nasal (N) *vs.* plosive (P)).

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

function of Vowel length are 96, 84 and 104 ms respectively, as shown in Fig. 2a. Figure 2b shows the results broken down by type of coda consonant, i.e. /l/, nasal or plosive. It is apparent that there is not much difference in the mean duration of the different types of consonants, for each of the three levels of Vowel length. This figure also shows that the variability around the mean is larger in relation to the plosive codas, the release of which is often realised weakly. Both overall (Fig. 2a) and by coda consonant (Fig. 2b), the duration of the coda is shorter for long vowels than for short and overlong vowels. However, there is considerable overlap between the distributions: the mean values of short, long and overlong vowels are well within one standard deviation from one another.

From a cross-linguistic perspective, there is evidence that coda duration is inversely correlated with vowel duration (Lehtonen 1970, Maddieson 1985). In this context, the difference in coda duration between target words that have short *vs*. long vowels is to be expected, but the difference in coda duration between target words that have long *vs*. overlong vowels is not.

We examined the significance of Vowel length as a fixed factor determining coda duration in a linear mixed-effects model with the random factors Speaker and Item. The effect of Vowel length on coda duration is not significant in a likelihood ratio test ($\chi^2(1) = 3.95$, $p = 0.047$).

3.2.3 *Vowel quality*. Time pressure exerts a centralising pressure on the realisation of vowels, so that longer vowel categories have more peripheral positions in the F1 × F2 space than shorter vowels (Lehtonen 1970, Remijsen & Gilley 2008). Hence the 3VL hypothesis predicts that short vowels may be centralised relative to long and overlong vowels. In contrast, if the formant values of levels of Vowel length differ in a way that cannot be interpreted in terms of centralisation, then this lends support to the interpretation that another phonological dimension is involved.

Figure 3 displays the descriptive statistics for F1 and F2. Because the formant values for the different speakers vary considerably, the raw F1 and F2 values were *z*-transformed by speaker (Lobanov 1971).[16] Transforming the values for F1 and F2 by speaker is important to ensure that the variability around the means is not inflated by between-speaker variation in formant ranges. For each of the four vowels included in the design, i.e. /ɛ a ɔ u/, Fig. 3 displays three ellipses, one for each of the levels of Vowel length. These ellipses encircle one standard deviation of the values, after *z*-transformation by speaker. The mean values by vowel for each of the levels of vowel length are connected by lines. A first observation is that, for each vowel, there is considerable overlap between the ellipses that mark the levels of Vowel length, meaning that the difference in vowel quality as a function of Vowel length is limited. Second, for each of the four vowels, the ellipsis for the short level of Vowel length

[16] As noted above (§3.1.1), a few items with the vowel /i/ were included in the speaker-specific *z*-transformations, so as to ensure that the resulting *z*-score ranges represent the speakers' full formant ranges, especially in relation to F2.

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

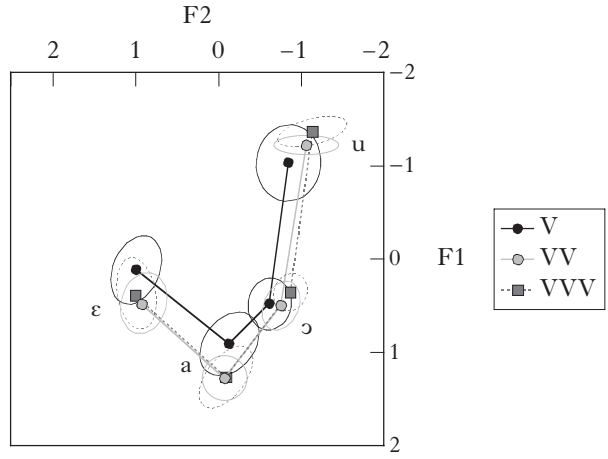


Figure 3
Means and standard deviations (as ellipses) for *z*-transformed F1 and F2, over speakers and items, by Vowel (/ɛ, a, ɔ, u/) and Vowel length (V, VV, VVV).

lies closest to the centre of the vowel space. Because of this, the polygon connecting the mean values of short vowels lies within the polygons connecting the means of long and the overlong vowels. This result suggests that the time pressure that is inherent in the production of short vowels indeed exerts a centralising effect on vowel quality.

To examine the influence of Vowel length on phonetic vowel quality quantitatively (i.e. through descriptive and inferential statistics), we calculated a measure of the degree of centralisation, derived from the raw F1 and F2 values. This is necessary because the centralisation observed in Fig. 3 affects different vowels differently in terms of their F1 and F2. For example, whereas in the case of /a/ the short vowels on average have a lower F1 value than the corresponding long and overlong vowels, in the case of /u/ the relation is the opposite, i.e. the short vowels have a higher F1 value than the long and overlong vowels.

The calculation of this measure of the degree of centralisation is based on Becker-Kristal (2010). We started by determining the centroid of each speaker's vowel space, on the basis of the speaker-specific mean values of the vowels /i a u/, which define the corners of the vowel triangle. Next, we calculated the Euclidean distance for each F1 × F2 measurement to the corresponding speaker-specific centroid. The descriptive statistics are displayed in Fig. 4. The mean values for this Euclidean-distance measure by levels of Vowel length are 390 Hz for short vowels, 440 Hz for long vowels and 465 Hz for overlong vowels. These descriptive statistics for centralisation are in line with the results in Fig. 3: short vowels are on average closer to the centroid of the vowel space than long and overlong

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

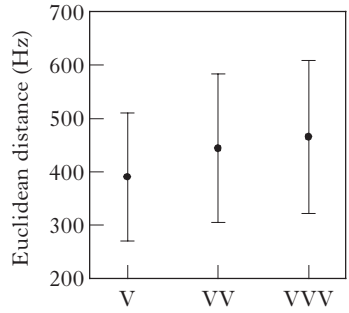


Figure 4

Means and standard deviations for the Euclidean distance from the speaker-specific centre of the vowel space, by Vowel length.

vowels, and overlong vowels tend to be more peripheral than short and long vowels. The variability around the mean values for short, long and overlong vowels is substantial, however, to the extent that the distributions largely overlap.

We then used this measure of centralisation (i.e. the Euclidean distance measure, in Hz) as the dependent variable in a linear mixed-effects model, with the fixed factors Vowel length (short, long, overlong) and Vowel quality (/i ɛ a ɔ/), and random factors Speaker and Item. There is a significant main effect on the degree of centralisation for Vowel length ($\chi^2(2) = 10.02$, $p = 0.0067$), but not for Vowel ($\chi^2(3) = 8.32$, $p = 0.0399$). The interaction between Vowel length and Vowel does not have a significant effect on the degree of centralisation ($\chi^2(6) = 12.26$, $p = 0.0564$). Vowel length displays a significant main effect, and because this factor has more than two levels, we followed up with pairwise comparisons (*t*-tests with Bonferroni correction), to find which combinations of factor levels are significant. The short level of length is significantly different from overlong ($p = 0.0057$), while the other two pairwise combinations are not significant (short *vs*. long: $p = 0.0776$; long *vs*. overlong: $p = 1$).

3.2.4 *Fundamental frequency*. The motivation for including F0 among the measurements stems from our concern to determine whether a phonological dimension other than vowel length could be plausibly invoked to represent any of the three levels of the hypothesised 3VL contrast. The 3VL hypothesis does not predict a difference in F0. In relation to 2VL + stress, Gilley (2004) states that tone may be a correlate of stress. Moreover, Gilley invokes stress in relation to morphophonological contrasts which we hypothesise to be marked solely by tone.

The phonological specification for tone is the same for all target words: a level High tone (cf. Table II). Because of the level nature of this specification, the measurement is taken at the temporal mid point of the vowel. The

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

measurements reveal very little divergence in the mean F0 values at the midpoint of short, long and overlong vowels – 193, 190 and 191 Hz respectively. As a function of Vowel, the mean values diverge somewhat more, in step with vowel height: they are 201 Hz for the high vowel /u/, 190 and 189 Hz for lower-mid /ɛ/ and /ɔ/ respectively, and 186 Hz for the low vowel /a/. This is in line with research which shows that, all else being equal, high vowels inherently have a higher F0 than low vowels (House & Fairbanks 1953, Whalen & Levitt 1995).

We examined the significance of Vowel length and Vowel as fixed factors in a linear mixed-effects model with Speaker and Item as random factors. The dependent is the F0 at the midpoint of the vowel. In the likelihood ratio tests, Vowel length does not exert a significant influence on F0 ($\chi^2(1) = 1.45$, $p = 0.2288$). Neither does Vowel, although this result approaches significance ($\chi^2(1) = 6.03$, $p = 0.0141$). Finally, the interaction between Vowel length and Vowel does not have a significant effect on F0 ($\chi^2(1) = 0.49$, $p = 0.4818$).

3.2.5 *Linear Discriminant Analysis.* Linear Discriminant Analysis allows us to establish for each item whether its factor level in terms of the hypothesised 3VL contrast can be predicted on the basis of a given measurement or set of measurements. It is important to establish this, because it answers the question as to whether a particular measurement distinguishes between the category levels; in this respect this technique is complementary to the inferential tests reported in §3.2.1 to §3.2.4. We carried out five LDAs. In each of these, the 192 items in the dataset were classified for Vowel length (short, long, overlong). In the first, the classification was based on all four of the measurements for which inferential statistics were reported in §3.2.1 to §3.2.4: vowel duration, coda duration, vowel quality and F0 at the temporal mid point. In the remaining four LDAs, classification was carried out on the basis of each of these four measurements individually. The overall classification results of these analyses are reported in Fig. 5. With three levels to be discriminated between, the chance-level base stands at 33%.

The LDA based on all four of the measures yields a correct classification result of 99%. Of the LDAs with a single measure, only the one with vowel duration results in a comparably high result, yielding correct classification in 96% of the cases. For the other three measures, the correct classification results are substantially lower, at 50% for coda duration, 40% for vowel quality and 35% for F0.

The similarity between the correct classification scores for the LDA with all four measures and the LDA with vowel duration alone suggests that in the former case vowel duration is the primary predictor, while the contribution of the other three factors is limited. This interpretation is supported by the correlation coefficients of the first two discriminant functions in the LDA with all four of the measurements; see Table III. Note that, in the most important discriminant function, Function 1, vowel duration has the most striking correlation coefficient (−3.59). In Function 2, it is the

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

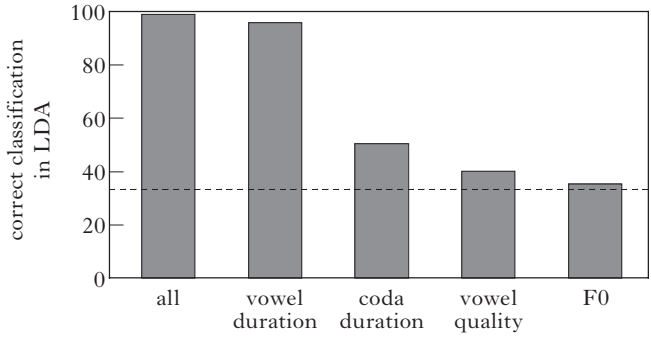


Figure 5

Correct classification for Vowel length (short, long, overlong) for LDAs: for all four of the measurements together, and on the basis of each of the measurements by itself. The horizontal line represents the chance-level baseline (33%).

correlation coefficient of coda duration that stands out most (−1.04). However, Function 2 accounts for very little of the discriminable variation: just 1.5% of the trace, as compared to Function 1 (98.5%) (the trace is the proportion of variance between factor levels, explained on the basis of successive discriminant functions).

	Function 1	Function 2
vowel duration	−3.59	−0.05
coda duration	−0.26	−1.04
vowel quality	−0.37	0.27
F0	−0.12	−0.24

Table III

Correlation coefficients of the four measures in relation to the first and second discriminant functions in an LDA with all four measurements.

Table IV reports the correct classification results separately for each of the three levels of Vowel length in each of the five LDAs. In line with the overall results shown in Fig. 5, there is little difference between the results for the LDA with all measurements and the LDA with vowel duration only. Among the results of individual levels of Vowel length for other measurements, the correct classification result for long vowels on the basis of coda duration stands out, at 70%. This reflects the fact that, on average, coda consonants have a shorter duration when they follow a long vowel than when they follow either a short or an overlong vowel. However, this tendency is not as consistent as the effect of vowel duration, which allows for a correct classification of long vowels in 94% of items. In

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

summary, the only measure that reliably distinguishes between levels of Vowel length is vowel duration.

	short	long	overlong
all	100	98	98
vowel duration	98	94	95
coda duration	27	70	54
vowel quality	55	16	50
F0	50	52	5

Table IV

Correct classification by level of Vowel length, in an LDA with all four measurements together, and in LDAs with the four measurements individually.

4 Discussion and conclusion

The descriptive analysis of the role of syllable-internal quantity in Shilluk phonology and morphophonology in §2.1 shows that the hypothesised ternary vowel-length contrast has a high functional load, both at the lexical level and in the morphology. All three levels of vowel length, i.e. short, long and overlong, appear in lexical roots in Shilluk. The latter finding is noteworthy, because in closely related Dinka, overlong vowels can invariably be conceived of as morphologically complex (Dimmendaal 1995), so that they do not need to be postulated as part of the lexical representation of roots, where a binary contrast is sufficient. Against this background, Shilluk presents a system where ternary vowel length has a wider scope, in that some lexical items have an overlong vowel throughout their paradigm (cf. (8)).

The morphological function of vowel length in Shilluk is extensive. All three of the logically conceivable alternations between levels of vowel length are attested: short ~ long, short ~ overlong and long ~ overlong. The short ~ overlong and long ~ overlong alternations have a high functional load in the inflectional paradigms of verbs and nouns, as a result of overlengthening processes. The short ~ long alternation is also found in both verbs and nouns, but its functional scope is limited to derivational processes. Here the direction of the quantity change is primarily shortening, i.e. from long to short.

In §2.2 we evaluated a competing analysis, in which Shilluk is postulated to have a binary vowel-length contrast plus stress (Miller & Gilley 2001, Gilley 2004). We rejected this hypothesis on the basis of language-internal considerations and theoretical concerns. The language-internal considerations are the following. First, 2VL + stress predicts that lexically short and lexically long verbs remain distinct when they undergo morphological lengthening, which is not the case. Second, not all morphological processes

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

for which stress has been invoked consistently involve increased quantity. In relation to fixed short roots, there is no change in quantity at all. In addition, stress has been invoked to represent the morphophonological marking of antipassive formation, which is quantity-increasing for most transitive verbs but quantity-decreasing for others. Finally, we noted that stress is invoked to represent morphological processes that are in fact marked by tone. From a theoretical point of view, the hypothesised stress contrast is anomalous in several respects: it is neither obligatory nor culminative, and its durational marking is the opposite of what is attested cross-linguistically, i.e. stressed syllables are characterised by a reduction in vowel duration.

The acoustic analysis of the phonetic realisation of the hypothesised ternary vowel-length contrast in §3 shows that short, long and overlong vowels diverge saliently in terms of vowel duration. The mean durations for short, long and overlong vowels are well spaced, at 68, 111 and 150 ms respectively. Compared with three other acoustic measures, this difference in duration conditions the biggest effect size for the factor Vowel length ($\chi^2(1) = 41.7$, $p < 0.0001$) of all of the likelihood ratio tests on linear mixed-effects analyses carried out. In addition, short vowels have a more central vowel quality than long vowels, and these in turn have a more central vowel quality than overlong vowels. The mean values, in terms of distance from the speaker-specific centroid of the vowel space, are 390 Hz for short vowels, 440 Hz for long vowels and 465 Hz for overlong vowels. Vowel quality also conditions a significant effect in linear mixed-effects analysis ($\chi^2(2) = 10.02$, $p = 0.0067$). The two other parameters which we investigated were coda duration and vowel F0. These do not display a significant effect for Vowel length, although coda duration comes close to significance ($\chi^2(1) = 3.95$, $p = 0.047$).

Linear Discriminant Analyses, both with the four measurements individually and also with the four entered together, lead to the conclusion that vowel duration is the only measure that reliably distinguishes between levels of Vowel length. This measure yields a correct classification result of 96%, similar to the correct classification result obtained with all four measures entered together, which is 99%. Of the other three measures, coda duration registers the highest correct classification result, at 50%.

In summary, the quantitative analyses indicate that vowel duration is the primary correlate of the contrast under investigation. On the basis of the various descriptive and inferential statistics, centralisation and coda duration can be interpreted as secondary correlates, at best. This state of affairs is in line with the ternary vowel-length hypothesis, but not with interpretations that account for the quantity contrast under investigation with a combination of binary vowel length and some other dimension of phonological contrast.

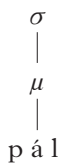
At the beginning of this paper, we set out two questions in relation to Shilluk. First, does Shilluk have a three-way quantity contrast marked by vowel duration in the syllable nucleus in its surface phonology? On the basis of the production study, this question can be answered in the

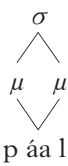
https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

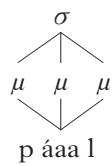
affirmative. This conclusion is important, because the vast majority of contrasts in vocalic quantity in the world's languages are binary in nature, and the evidence for three-way contrasts in vowel length is limited (Odden 2011, Prehn 2012). In this context, it is interesting that there are nevertheless some languages that present robust evidence for ternary vowel length. This indicates that the bias towards binarity in quantity contrasts is not absolute. Blevins (2004) attributes the development of ternary vowel length in Dinka to its morphological load; this explanation can be extended to Shilluk.

Our findings strengthen the case for ternary vowel length in West Nilotic. As noted in the introduction, ternary vowel length has been postulated in relation to several West Nilotic languages (Andersen 1987 on Dinka, Reid 2010 on Thok Reel, Tucker 1955 on Shilluk and Monich 2017 on Nuer). However, the phonetic evidence has so far been limited to Dinka (Remijsen & Gilley 2008, Remijsen 2013, 2014). The research reported here strengthens this body of evidence.

Given the conclusion that Shilluk presents a three-way contrast in vocalic quantity in the syllable nucleus, the second question presents itself: what is the best phonological representation of this contrast? In relation to Dinka, Hayes (1989) argues that ternary vowel length can be represented straightforwardly by associating one, two and three weight units or moras with syllables containing a short, long and overlong vowel respectively. This representation can be extended to Shilluk. This is shown in (16), using one of the minimal sets of the production study.

(16) a. 

b. 

c. 

The phonological representation in (16) represents phonetic quantity, specifically vowel duration, directly in terms of phonological quantity in the syllable nucleus. We consider this direct relation between the primary correlate of a contrast and its phonological representation to be an advantage of this analysis. There are a variety of alternative ways in which the contrast can be represented without postulating a three-way phonological quantity contrast in the nucleus. For example, we could postulate mora sharing between nucleus and coda, along the lines of the analysis proposed in Baal *et al.* (2012) to represent the three-way contrast in consonantal quantity in the syllable coda in Saami. The disadvantage of postulating such an analysis for Shilluk is that a contrast that resides within the syllable nucleus is represented with reference to a different constituent, the coda. As seen from the LDA results, for each of the three pairwise comparisons between levels of the hypothesised ternary vowel-length contrast, i.e. short *vs.* long, short *vs.* overlong and long *vs.* overlong, vowel

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

duration offers a clearer separation than coda duration. In addition, we do not find support for such an analysis in the morphophonology.

We hope that our work will stimulate further research on the realisation, representation and explanation of ternary vowel length in Shilluk and other languages. If it is correct to conclude that 3VL exists, a key question is why it is so uncommon (cf. Blevins 2004, Remijsen & Gilley 2008). One area of research that holds promise for shedding light on this question is speech perception. As it stands, there is no research available on how a ternary vowel-length contrast is perceived. It is conceivable that the just noticeable difference for the perception of durational contrasts is a constraining factor (cf. Remijsen & Gilley 2008 and references therein). Equally, it would be worth finding out whether native speakers of languages with 3VL (e.g. Shilluk and Dinka) perceive durational differences in speech in a more fine-grained manner, so that their just noticeable difference for duration would be smaller than that of native speakers of languages with less complex systems of vocalic quantity.

REFERENCES

Andersen, Torben (1987). The phonemic system of Agar Dinka. *Journal of African Languages and Linguistics* **9**. 1–27.

Andersen, Torben (1990). Vowel length in Western Nilotic languages. *Acta Linguistica Hafniensia* **22**. 5–26.

Andersen, Torben (1992–94). 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.

Baal, Berit Anne Bals, David Odden & Curt Rice (2012). An analysis of North Saami gradation. *Phonology* **29**. 165–212.

Baayen, R. H., D. J. Davidson & D. M. Bates (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* **59**. 390–412.

Bao, Zhiming (1999). *The structure of tone*. Oxford: Oxford University Press.

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tilly (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language* **68**. 255–278.

Bates, Douglas (2005). Fitting linear mixed models in R: using the lme4 package. *R News* **5:1**. 27–30.

Bates, Douglas, Martin Mächler, Benjamin M. Bolker & Steven C. Walker (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**. 1–48.

Becker-Kristal, Roy (2010). *Acoustic typology of vowel inventories and Dispersion Theory: insights from a large cross-linguistic corpus*. PhD dissertation, University of California, Los Angeles.

Blevins, Juliette (2004). *Evolutionary Phonology: the emergence of sound patterns*. Cambridge: Cambridge University Press.

Boersma, Paul & David Weenink (2015). *Praat: doing phonetics by computer* (version 5.4.22). http://www.praat.org.

Childers, Donald G. (1978). *Modern spectrum analysis*. New York: IEEE Press.

Clements, G. N., Alexis Michaud & Cédric Patin (2010). Do we need tone features? In Goldsmith *et al*. (2010). 3–24.

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

Cohen, Jacob (1994). The earth is round ($p < .05$). *American Psychologist* **49**. 997–1003.
Comrie, Bernard, Martin Haspelmath & Balthasar Bickel (2015). The Leipzig glossing rules. https://www.eva.mpg.de/lingua/resources/glossing-rules.php.
Dimmendaal, Gerrit J. (1995). Vowels as complex segments in Nilotic. In Nicolaï & Rottland (1995). 147–168.
Dixon, R. M. W. (2010). *Basic linguistic theory*. Vol. 2: *Grammatical topics*. Oxford: Oxford University Press.
Duanmu, San (2009). *Syllable structure: the limits of variation*. Oxford: Oxford University Press.
Elenbaas, Nine & René Kager (1999). Ternary rhythm and the lapse constraint. *Phonology* **16**. 273–329.
Elert, Claes-Christian (1964). *Phonologic studies of quantity in Swedish*. Uppsala: Almqvist & Wiksell.
Flack, Kathryn (2007). Templatic morphology and indexed markedness constraints. *LI* **38**. 749–758.
Gilley, Leoma G. (1992). *An autosegmental approach to Shilluk phonology*. Arlington: Summer Institute of Linguistics.
Gilley, Leoma G. (2004). The feature of stress in Nilotic. In Rose-Juliet Anyanwu (ed.) *Stress and tone: the African experience*. Cologne: Rüdiger Köppe. 99–120.
Goldsmith, John A., Elizabeth Hume & W. Leo Wetzels (eds.) (2010). *Tones and features: phonetic and phonological perspectives*. Berlin & Boston: De Gruyter Mouton.
Haspelmath, Martin (1996). Word-class-changing inflection and morphological theory. *Yearbook of Morphology 1995*. 43–66.
Hayes, Bruce (1989). Compensatory lengthening in moraic phonology. *LI* **20**. 253–306.
Hermans, Ben & Francesc Torres-Tamarit (2014). Ternary constituents are a consequence of mora sluicing. In John Kingston, Claire Moore-Cantwell, Joe Pater & Robert Staubs (eds.) *Proceedings of the 2013 Meeting on Phonology*. http://dx.doi.org/10.3765/amp.v1i1.26.
Heuven, Vincent J. van (2014). Acoustic correlates and perceptual cues of word and sentence stress: mainly English and Dutch. In Carlos Gussenhoven, Yiya Chen & Dan Dediu (eds.) *Proceedings of the 4th International Symposium on Tonal Aspects of Languages*. 211–217. Available (November 2018) at http://www.isca-speech.org/archive/tal_2014.
House, Arthur S. (1961). On vowel duration in English. *JASA* **33**. 1174–1178.
House, Arthur S. & Grant Fairbanks (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *JASA* **25**. 105–113.
Hyman, Larry M. (2006). Word-prosodic typology. *Phonology* **23**. 225–257.
Hyman, Larry M. (2010). Do tones have features? In Goldsmith *et al*. (2010). 50–80.
Kohler, Klaus J. (2001). Überlänge im Niederdeutschen? In Robert Peters, Horst P. Pütz & Ulrich Weber (eds.) *Vulpis Adolatio: Festschrift für Hubertus Menke zum 60. Geburtstag*. Heidelberg: Winter. 385–402.
Kohler, Klaus J. & Regina Tödter (1984). Experimentalphonetische Untersuchungen zur 'Überlänge' im Nieder- und Hochdeutschen Schleswig-Holsteins. *Arbeitsberichte Institut für Phonetik der Universität Kiel* **21**. 61–131.
Lehtonen, Jaakko (1970). *Aspects of quantity in Standard Finnish*. Jyväskylä: Jyväskylä University Press.
Letterman, Rebecca S. (1994). A phonetic study of Sinhala syllable rhymes. *Working Papers of the Cornell Phonetics Laboratory* **9**. 155–181.
Lobanov, Boris M. (1971). Classification of Russian vowels spoken by different speakers. *JASA* **49**. 606–608.
Lowenstamm, Jean (1996). CV as the only syllable type. In Jacques Durand & Bernard Laks (eds.) *Current trends in phonology: models and methods*. Salford: ESRI. 419–441.

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

McCloy, Daniel R. (2016). Package 'phonR': tools for phoneticians and phonologists. R package (version 1.0-7). https://cran.r-project.org/web/packages/phonR/phonR.pdf.
Mackenzie, Sara (2016). Consonant harmony in Nilotic: contrastive specifications and Stratal OT. *Glossa* **1**(1):12. 1–38. http://doi.org/10.5334/gjgl.8.
Maddieson, Ian (1985). Phonetic cues to syllabification. In Victoria A. Fromkin (ed.) *Phonetic linguistics: essays in honor of Peter Ladefoged*. Orlando: Academic Press. 203–221.
Miller, Cynthia L. & Leoma G. Gilley (2001). Evidence for ergativity in Shilluk. *Journal of African Languages and Linguistics* **22**. 33–68.
Miller, Cynthia L. & Leoma G. Gilley (2007). Evidentiality and mirativity in Shilluk. In Doris L. Payne & Mechthild Reh (eds.) *Advances in Nilo-Saharan linguistics*. Cologne: Rüdiger Köppe. 191–206.
Monich, Irina Vitta (2017). Vowel length in Nuer. In Karen Jesney, Charlie O'Hara, Caitlin Smith & Rachel Walker (eds.) *Proceedings of the 2016 Annual Meeting on Phonology*. http://dx.doi.org/10.3765/amp.v4i0.4006.
Nicolaï, Robert & Franz Rottland (eds.) (1995). *Actes du Cinquième Colloque de Linguistique Nilo-Saharienne*. Cologne: Rüdiger Köppe.
Nooteboom, S. G. (1972). *Production and perception of vowel duration: a study of the durational properties of vowels in Dutch*. PhD dissertation, University of Utrecht.
Noske, Manuela (1995). Consonant gemination in Shilluk. In Nicolaï & Rottland (1995). 217–243.
Odden, David (2011). The representation of vowel length. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.) *The Blackwell companion to phonology*. Malden, Mass.: Wiley-Blackwell. 465–490.
Ortega-Llebaria, Marta & Pilar Prieto (2011). Acoustic correlates of stress in Central Catalan and Castilian Spanish. *Language and Speech* **54**. 73–97.
Prehn, Maike (2012). *Vowel quantity and the fortis–lenis distinction in North Low Saxon*. PhD dissertation, University of Amsterdam.
R Core Team (2016). R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. http://www.r-project.org.
Reid, Tatiana (2010). *Aspects of phonetics, phonology and morphophonology of Thok Reel*. MSc dissertation, University of Edinburgh.
Remijsen, Bert (2013). Tonal alignment is contrastive in falling contours in Dinka. *Lg* **89**. 297–327.
Remijsen, Bert (2014). Evidence for three-level vowel length in Ageer Dinka. In Johanneke Caspers, Yiya Chen, Willemijn Heeren, Jos Pacilly, Niels O. Schiller & Ellen van Zanten (eds.) *Above and beyond the segments: experimental linguistics and phonetics*. Amsterdam & Philadelphia: Benjamins. 246–260.
Remijsen, Bert & Otto Gwado Ayoker (2014). Contrastive tonal alignment in falling contours in Shilluk. *Phonology* **31**. 435–462.
Remijsen, Bert & Otto Gwado Ayoker (2018). Forms and functions of the base paradigm of Shilluk transitive verbs. In *A grammar of Shilluk*. Honolulu: University of Hawai'i Press. Available (November 2018) at http://hdl.handle.net/10125/24777.
Remijsen, Bert, Otto Gwado Ayoker & Signe Jørgensen (2018). Shilluk_2016_controlled_ThreeLevelVowelLength. (Dataset.) Edinburgh Datashare. http://dx.doi.org/10.7488/ds/2424.
Remijsen, Bert, Otto Gwado Ayoker & Timothy Mills (2011). Shilluk. *Journal of the International Phonetic Association* **41**. 111–125.
Remijsen, Bert & Leoma G. Gilley (2008). Why are three-level vowel length systems rare? Insights from Dinka (Luanyjang dialect). *JPh* **36**. 318–344.
Remijsen, Bert, Cynthia L. Miller-Naudé & Leoma G. Gilley (2015). Stem-internal and affixal morphology in Shilluk. In Matthew Baerman (ed.) *The Oxford handbook of inflection*. Oxford: Oxford University Press. 577–596.

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press

Remijsen, Bert, Cynthia L. Miller-Naudé & Leoma G. Gilley (2016). The morphology of Shilluk transitive verbs. *Journal of African Languages and Linguistics* **37**. 201–245.
Rice, Keren (2014). Sounds in grammar writing. In Toshihide Nakayama & Keren Rice (eds.) *The art and practice of grammar writing*. Honolulu: University of Hawai'i Press. 69–89.
Schaeffler, Felix, Pär Wretling & Eva Strangert (2002). On the development of a quantity typology for Swedish dialects. In Bernard Bel & Isabelle Marlien (eds.) *Speech prosody 2002*. Aix-en-Provence. 627–630. Available (November 2018) at http://www.isca-speech.org/archive/sp2002.
Torres-Tamarit, Francesc & Peter Jurgec (2015). Lapsed derivations: ternary stress in Harmonic Serialism. *LI* **46**. 376–387.
Tucker, A. N. (1955). The verb in Shilluk. *Mitteilungen des Instituts für Orientforschung* **3**. 421–462.
Venables, W. N. & B. D. Ripley (2002). *Modern applied statistics with S*. 4th edn. New York: Springer.
Warnes, Gregory R., Ben Bolker, Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber, Andy Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen Moeller, Marc Schwartz & Bill Venables (2016). Package 'gplots': various R programming tools for plotting data. R package (version 3.0.1). https://cran.r-project.org/web/packages/gplots/gplots.pdf.
Whalen, D. H. & Andrea G. Levitt (1995). The universality of intrinsic F_0 of vowels. *JPh* **23**. 349–366.
Windmann, Andreas, Juraj Šimko & Petra Wagner (2015). Polysyllabic shortening and word-final lengthening in English. *Proceedings of Interspeech 2015. 36–40*. https://www.isca-speech.org/archive/interspeech_2015/i15_0036.html.
Winter, Bodo (2013). Linear models and linear mixed effects models in R with linguistic applications. Available (November 2018) at http://arxiv.org/pdf/1308.5499.pdf.
Xu, Tongling (2017). *Number marking in Shilluk*. MA dissertation, University of Edinburgh.
Yip, Moira (1980). *The tonal phonology of Chinese*. PhD dissertation, MIT.
Yu, Alan C. L. (2008). The phonetics of quantity alternation in Washo. *JPh* **36**. 508–520.

https://doi.org/10.1017/S0952675719000058 Published online by Cambridge University Press