THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Ternary vowel length in Shilluk

OPEN ACCESS

Ternary vowel length in Shilluk

Bert Remijsen, Otto Gwado Ayoker, and Signe Jørgensen

University of Edinburgh

Abstract

Three-level vowel length is typologically rare, and the supporting evidence is limited (Odden 2011, Prehn 2012). This paper presents the results of an investigation into the hypothesized case of this configuration in Shilluk, a Nilo-Saharan language spoken in South Sudan. The paper first describes the role of vowel length in Shilluk phonology and morphology. Then we report on an acoustic study with eight native speakers, in which eight monosyllabic minimal sets for vowel length (short, long, overlong) are measured for vowel duration, coda duration, vowel quality (F1, F2), and fundamental frequency (F0). The three levels of the hypothesized contrast differ significantly and substantially in terms of vowel duration, to the effect that 96 percent of the items can be classified successfully for vowel length in a Linear Discriminant Analysis on the basis of this measurement alone. Of the other measurements, only vowel quality registers a significant effect at alpha .01, and this effect is considerably smaller than that of vowel duration. The mean values for vowel duration – 68 ms for short vowels, 111 ms for long vowels and 150 ms for overlong ones – are similar to those reported for three-level vowel length in Dinka, a closely-related language.

1. Introduction

1.1 Data and theory in relation to three-level vowel length

Three-level or ternary vowel length refers to a three-way phonological distinction between short vs. long vs. 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 the typological range of vowel length is limited to binary contrast, i.e., short vs. long (Kohler 2001, Duanmu 2009). Commenting on the lack of evidence for this configuration, Odden (2011:465) writes that: "for the vast majority of [languages for which three-level vowel length has been postulated], so little is known that it is difficult to say that length ternarity is a real phenomenon, rather than an epiphenomenon arising from binary length and some orthogonal dimension."

Critical evaluations of hypothesized cases can be found in Odden (2011) and Prehn (2012). Both conclude that the analysis of three-level 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 three-level 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 three-level vowel length is supported by phonological and phonetic investigations on several dialects (Andersen 1987, 1990, 1992-1994; Flack 2007; Remijsen & Gilley 2008; Remijsen 2013, 2014).

Dinka content words alternate in vowel length either between short and long, or between long and overlong. This is illustrated in Table 1. 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 at the level of lexical roots, i.e., there are lexical roots with a short vowel such as {lel} and lexical roots 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 into a three-level contrast. This is illustrated in Table 1 by **ă-lèl**, **ă-lèel**, and **ă-lèeel**. The reader can ascertain this by playing the embedded sound examples.

Table 1. The interaction between lexical vs. 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.
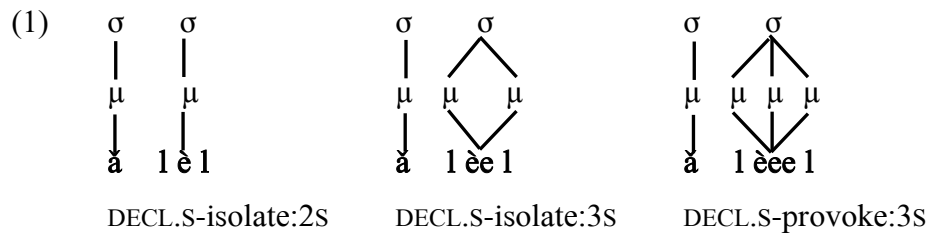
| | | Morphological vowel length (grade) | |
		Short	Long
Lexical vowel length	Short	ràaan **ǎ-lèl**	ràaan **ǎ-lèel**
		person DECL.S-isolate:2S	person DECL.S-isolate:3S
		'You are isolating somebody.'	'He is isolating somebody.'
	Long	ràaan **ǎ-léel**	ràaan **ǎ-lèeel**
		person DECL.S-provoke:2S	person DECL.S-provoke:3S
		'You are provoking somebody.'	'He is provoking somebody.'

A production study on three-level 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 milliseconds (ms), respectively, in utterance-medial context, and the distributions are 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 on three-level vowel length relates to a more general debate in phonology, as to the status of binarity in sound systems. In relation to tone, for example, inventories with three or more level tones have frequently been analysed through a nested combination of two binary features (Yip 1980, Bao 1999). However, such an architecture is ill-suited to represent not just rich inventories but also common contextual tone processes (Hyman 2011; Clements, Michaud & Patin 2011). With regards to metrical systems, constraints have been invoked with the specific aim to avoid postulating ternary feet (e.g. Elenbaas & Kager 1999, Hermans & Torres-Tamarit 2013, Torres-Matarit & Jurgec 2015). The appeal of binarity is also evident from the proposal within government phonology that all syllables are CV, i.e., with 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 presents itself. It is clear that, across the world's languages, vowel length tends to represent 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 morae (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 1 can be represented by associating one, two, or three weight units, as in (1).

(1)

σ σ	σ σ	σ σ
μ μ	μ μ μ	μ μ μ μ
à l è l	à l èe l	à l èee l
DECL.S-isolate:2S	DECL.S-isolate:3S	DECL.S-provoke:3S

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 distinguishing the structures in (1b) and (1c) in terms of the mora count of the stem syllable, they could be distinguished from one another, for example, by postulating mora-sharing with the coda consonant, an analysis similar to the one postulated in relation to coda quantity in Saami (Baal, Odden & Rice 2012).

In summary, the phenomenon of three-level vowel length raises two questions. First, can a language have a three-level 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 three-level 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 hypothesized cases. In relation to the first question, the phonetic investigations have to go beyond transcriptions, because transcriptions are subjective and therefore cannot serve as a common

ground when the facts themselves are in contention, as is often the case in relation to three-level vowel length. A well-established way to make the analysis more accountable is by carrying out acoustic measurements on a representative sample, which are then subject to statistical analysis. The data should be made publicly available, so that the analysis can be replicated and challenged. Another way to increase accountability is to include sound examples (Rice 2014). 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 hypothesized cases. In summary, the comprehensive investigation of three-level vowel length crosses the boundaries between phonetics, phonology, and morphology.

Given the importance of the Dinka case to this controversy, it is worthwhile to examine the evidence in closely-related languages. In fact, three-level vowel length has been postulated for three of these, all belonging to the West Nilotic subbranch within the Nilo-Saharan family: Shilluk (Tucker 1955; Remijsen, Ayoker & Mills 2011; Remijsen, Miller-Naudé & Gilley 2015), Thok Reel (Reid 2010), and Nuer (Monich 2016). 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 our paper is to answer the first question, i.e., whether Shilluk has a three-level contrast in vocalic quantity in the syllable nucleus. With respect to the second question – what is the most compelling phonological representation – here we will not evaluate the full range of options. Instead, our aim is to provide a descriptive analysis of the role of vowel length in Shilluk phonology and morphology, which can inform this debate. Sound examples are embedded, and the dataset as a whole is publicly available (Remijsen, Ayoker & Jørgensen 2018).

The paper is structured as follows. In Section 2, we first describe the syllable-internal quantity contrasts that serve lexical and morphological functions in Shilluk in terms of a hypothesized three-level vowel length contrast (Section 2.1). Then we introduce an alternative hypothesis of the same phenomena (Gilley 2003), which does not violate the hypothesis that vowel length is maximally binary (Section 2.2). Section 3 presents the results of an acoustic analysis of minimal sets, in which the hypothesis of three-level vowel length is put to the test. The conclusion and

discussion follow in Section 4. The remainder of this introduction presents some background information on the Shilluk sound system, except for the hypothesized 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 Shilluk 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 (2).[1] With few exceptions, these stem syllables tend to be closed. The stem can combine with one prefix syllable, typically /a-/ or /ʊ-/, and up to two suffixes. Polysyllabic word roots, such as úgǐik 'buffalo' in (2), 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) dɔ́ɔc   úgǐik   ú-cāaam
good   buffalo   FUT-eat:XV

líɲ   jǎaak̚-ɔ̄   á-mā̱ṭ-ì-á [á-mā̱ṭ-áa]
conflict   chief-s   PAST-drink-ITER-1S

The inventory includes ten vowel phonemes, which are organized in terms of an advanced tongue root (ATR) feature: there are the -ATR vowels /ɪ,ɛ,a,ɔ,ʊ/, and their +ATR counterparts /i,e,ʌ,o,u/, respectively (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, Ayoker & Mills 2011). ATR does not play a role in vowel harmony. Instead, there are several morphological processes in which -ATR changes to +ATR

---

[1] Our transcriptions follow the IPA standard, with the following exceptions. First, segmentable morphemes or combinations of morphemes are separated by hyphens, in line with the Leipzig Glossing Rules (Comrie, Haspelmath & Bickel 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). Note that various tone categories are represented using multiple tonal diacritics, either on a vowel character 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 on the phonological analysis. For example, in the case of the High Fall to Mid, as in kâaaŋ 'trumpet:PERT.S', the combination of diacritics denotes a syllable-level tone category realized 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 in the sense that a High Fall would be associated with the vowel and the Mid tone with the coda.

(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-mâ̰a̰t**   **á-mʌ̂ʌʌ̰t**   b. **á-máaa̰t**   **á-mʌ́ʌʌ̰t**

  PAST-drink:OV PAST-drink:ATP   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/. In addition there are 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 – /w/ or /j/ – can follow another consonant, as in **kwáac** 'fishscales' and **rjēw** 'second'. There are two additional phonotactic restrictions. First, dental and alveolar obstruents do not co-occur within the phonological word (cf. Mackenzie 2016).[2] Second, voicing in plosives is not contrastive in the stem-final consonant. In this position, plosives vary freely in voicing.[3]

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.[4] 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 & Ayoker 2014). Eight of these are illustrated in (4); see also footnote 1. On affixes, the inventory is limited to Low, Mid and High.

---

[2] Remijsen et al. (2011:113) present a few forms that violate this consonantal place harmony. However, following publication, we have realized 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 **n̪ṵ̂uu̪ɪ̄** 'dried cowdung', **d̪ṵ̂uu̪t-ɔ̀** 'overflow from pressure buildup', **t̪ṵ̀uu̪t̪** 'ideophone for aggressive vocalisation of a cat or a snake, or an aggressive outbust of a person'.

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

[4] The contrast between Low Rise and High Rise was not postulated in Remijsen et al. (2011); we have discovered it only recently. A minimal set example is **pʌ̌ʌʌt̪** 'twist.rope.using.hands:ATP:IMP' "Twist!" vs. **pʌ̌ʌʌt̪** 'fall:IMP' "Fall!". In base forms of nouns, Low Rise is found among others in **ŋ̌ooom** 'awl' and **ť̪aaak** 'hat', and High Rise among others in **ɲǎaaw** 'house cat' and **mwɔ̌ɔɔl** 'morning'.

(4)   á-kɔ̀l (Low)   á-kɔ̄l (Mid)   á-kɔ́l (High)
       PAST-disturb:2S   PAST-disturb:XV   PAST-take.out:XV:2S

       á-kɔ̂l (Low Fall)   á-kɔ̂l (High Fall)   á-kɔ̂l (High Fall to Mid)
       PAST-take.out:2S   PAST-disturb:OV   PAST-take.out:XV

       á-kɔ́lˋ (Late Fall)   á-kɔ̌l (High Rise)
       PAST-take.out:FUG   PAST-disturb:XV:2S

2. The phonology of syllable-internal quantity in Shilluk

Section 2.1 presents a descriptive analysis of Shilluk's syllable-internal quantity contrasts at the lexical and morphological levels. Here we use the three-level vowel length analysis to represent the patterns of quantity. After that, in Section 2.2, we will 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. In Table 2 these patterns are illustrated on the basis of transitive verb inflections. Here the lexical length of the root vowel is illustrated by the stem form 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, there are transitive verbs that have a short root vowel and that do not display morphological lengthening of the stem vowel. The verb {ŋɔl} 'cut', for example, has a short vowel throughout its morphological paradigm. We will refer to these as Fixed Short roots. Second, there are verbs that have a short root vowel, but which appear with increased vowel length in many inflections, including the past tense 2nd singular. This pattern, which we refer to as Short with Grade, is illustrated in Table 2 by {càm} 'eat'. Third, there are verbs that have a long root vowel, and which also undergo morphological lengthening. These are the 'Long with Grade' verbs; {lὲὲŋ} 'throw' is one of them. Note that morphological lengthening in Shilluk is 'overlengthening': if a verb displays morphological lengthening, then it lengthens to a third level of vowel length, 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).

Table 2. The three patterns of inflectional vowel length alternation, in transitive verbs.

	PAST OBJECT VOICE	PAST 2ND SINGULAR	
Fixed Short	á-ŋɔ́l	á-ŋɔ̀l	'cut'
Short with Grade	á-cɑ̂m	á-càaam	'eat'
Long with Grade	á-lɛ́ɛŋ	á-lɛ̀ɛɛŋ	'throw'

The same quantity alternations appear elsewhere in the verb paradigm. Consider for example the antipassive. This derivation is marked through a package of stem-internal markers: quantity alternation, a raising in vowel quality, a change of coda /l,r/ to /t/ (cf. Remijsen, Miller-Naudé & Gilley 2016). The antipassives of the three verbs in Table 2 are á-ŋʊ̀t, á-cʌʌʌm, 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 Table 3 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 Table 3, these base forms either have a short vowel or a long one. Noun inflections such as the pertensive[5] and the construct state display morphological vowel lengthening. On a par with transitive verb paradigms, we find Fixed Short nouns that have a short stem vowel and do not display a length alternation; Short with Grade nouns, that alternate between short and overlong vowel length; and Long with Grade nouns, that alternate between long and overlong.

Table 3. The three patterns of inflectional vowel length alternation, in suffixless singular nouns.

	BASE	PERTENSIVE SG.	PERTENSIVE PL.	
Fixed Short	ḍɔ́k	ḍɔ́k̄	ḍɔ́k	'mouth'
Short with Grade	káŋ	kɑ̂aaŋ	káaaŋ	'trumpet'
Long with Grade	gʌ̀ʌt	gʌ̀ʌʌt	gʌ̌ʌʌt	'river bank'

---

[5] The pertensive – term from Dixon (2010:268) – is used when a noun is modified by a possessor. In the case of **káŋ** 'trumpet' for example, there is **kɑ̂aaŋ kʊ̀l** 'Kul's trumpet', and **káaaŋ ɟìr** 'people's trumpet'. If the head noun is singular, the pertensive also marks the number of the possessor modifier. As seen from Table 2, the pertensive with singular possessor never ends in a high tone target; in contrast, the pertensive with plural possessor invariably does.

There are also singular nouns that have a suffixed base form. Almost invariably, the suffix at issue is singulative **-ɔ**. As seen from Table 4, we find the same three patterns of alternation here, albeit with a twist. Note that, in Short with Grade and in Long with Grade paradigms, the overlong stem vowel is to be found in the base form, and the underlying length of the lexical root in the inflected form. That is, in suffixed singular nouns the length alternation applies in the opposite direction, as compared with the suffixless nouns.

Table 4. The three patterns of inflectional vowel length alternation, in suffixed singular nouns.

	BASE	PERTENSIVE SG	PERTENSIVE PL	
Fixed Short	pîc-ɔ̀	pîc-ì	pîc-í	'tethering pole'
Short with Grade	pâaal-ɔ̀	pâl-ì	pâl-í	'knife'
Long with Grade	bɔ̀ɔɔt̪-ɔ̀	bɔ̀ɔt̪-ì	bɔ̀ɔt̪-í	'craftsman'

The patterns of length alternation interact in part with vowel quality. The Short with Grade verbs have a low vowel /ʌ,a/, or they have a complex onset that includes a semivowel, as in {gwêt} 'write' and {tjêk} 'surround'. The Fixed Short have monophthong vowels other than /ʌ,a/.[6] The Long with Grade vowels, finally, are found with any vowel quality.

In the patterns of length alternation surveyed so far, lexical items appear either with a short or with 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 for Dinka a binary vowel length distinction – 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 presents content-word morphemes that are only found with an overlong vowel, i.e., roots that are Fixed Overlong. Table 5 presents three 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

---

[6] 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 with **páac** 'amniotic fluid'.

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.

Table 5. Intransitive verbs that have an overlong root vowel throughout their paradigm; the paradigm is illustrated by several inflections.

PAST	NEVP	IMPF	FUT	INV	INV 1SG	
á-jûuuṭ	jûuuṭ-ɔ̂	ɔ́-jûuuṭ-ɔ̂	ɔ́-júuuṭ	jűuuⱦ	júuuṭ-á	'finish'
á-dʌ̂ʌʌr	dʌ̂ʌʌr-ɔ̂	ɔ́-dʌ̂ʌʌr-ɔ̂	ɔ́-dʌ́ʌʌr	dʌ́ʌʌ̄r	dʌ́ʌʌr-á	'become tired'
á-bûuuṭ	bûuuṭ-ɔ̂	ɔ́-bûuuṭ-ɔ̂	ɔ́-búuuṭ	bűuuⱦ	búuuṭ-á	'become bored'
á-tûuul	tûuul-ɔ̂	ɔ́-tûuul-ɔ̂	ɔ́-túuul	tűuul̄	túuul-á	'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 among others in the formation of the antipassive derivation of certain Long with Grade verbs that have a Low Fall as their lexical tone.[7] This is illustrated in Table 6: the subject voice past tense form displays the lexical specifications for tone and vowel length of the root (cf. Remijsen & Ayoker 2018). As Long with Grade verbs, the verbs at issue alternate inflectionally between long and overlong vowel lengths; this is illustrated by the 2nd singular past tense forms in Table 6. In the antipassive derivation, however, the vowel is short if the vowel height 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 {cûʊl} 'pay for' appear with short, long and overlong stem vowels within their paradigm. As for the change of /ʊ/ to /u/ in the latter, verb, this is predictable. The vowel qualities /ɪ,ʊ,e,o/ are avoided in morphologically derived short vowels. That is, when the vowel is either /ɪ/ or /e/ underlyingly, it shifts most often to /i/ and sometimes to /ɛ/. If it is either /ʊ/ or /o/, it shifts most often to /u/ and sometimes to /ɔ/.

---

[7] The key motivation to treat 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-modality, voice and subject marking, parallel to the corresponding inflectional paradigm of the transitive base.

Table 6. Morphological lengthening and shortening in Long with Grade Low Fall verbs. The verb root is CV̂VC.

PAST SUBJECT VOICE	PAST 2ᴺᴰ SINGULAR	ANTIPASSIVE PAST	
á-bûuk	á-bûuuk	á-bûk	'cover w. powder'
á-côʊl	á-côʊʊl	á-cût	'pay for'
á-gûur	á-gûuur	á-gûr	'grind'
á-mî̪n	á-mî̪ɪ̪n	á-mîn	'pierce'
á-mêɛɲ	á-mêɛɛɲ	á-mèeeɲ	'illuminate'
á-mâa̪ʈ	á-mâaa̪ʈ	á-mʌʌʌ̪ʈ	'drink'

Vowel shortening is regularly found in other morphological operations on transitive verbs as well, such as the derivations for benefactive and iterative for transitive verbs (Remijsen, Miller-Naudé & Gilley 2016). For example, Long with Grade {mâa̪ʈ} 'drink' has a short stem vowel in both – the past tense object voice forms of the iterative derivation is á-má̪ʈ-ì and the past tense object voice forms of the benefactive derivation is á-mʌ̪́ʈ-ì.

Outside the system of transitive verbs, morphological shortening of the vowel is also found in nouns, in particular in number marking. We will need to provide some background information on this part of the grammar before we can discuss morphological shortening. Number marking is neither productive nor regular in Shilluk. In relation to productivity, roughly 26 percent of singular nouns in our lexicographic data do not have a corresponding plural.[8] As for regularity, number can be marked through suffixation and a variety of stem-internal changes (tone, vowel length, stem-final consonant, and vowel quality / breaking), and which combination of these exponents is involved is not predictable (Xu 2017). Table 7 presents a set of singular nouns that are morphologically unmarked,[9] all of which a Low Fall and a long vowel, i.e., the same

---

[8] Our lexicographic data includes 917 singular nouns and 678 plural nouns.

[9] It is not invariably the case that the singular is morphologically unmarked whereas the plural is morphologically 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 Table 7, 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 marking prefix, as in lùm á-lî̪ɲʈ-à 'I saw the grasses', the past tense prefix has increased duration, i.e., [á:lî::ḍà] rather than [álî::ḍà]. Presumably, floating quantity in these forms is a diachronic reflex of a lost plural-marking suffix. An unmarked

specification for tone and vowel length as the transitive verbs in Table 6. The irregularity of number marking is illustrated by the fact that the nouns on the right 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 as an inflectional phenomenon (cf. Haspelmath 1996:47).

Table 7. Examples of morphological alternation in vowel length and tone in number marking in nouns with a CV̂VC root in the singular.

Shortening in plural formation		Lengthening in plural formation	
mîîŋ – mìŋ	'deaf.person:S/P'	dâak – dǎaak	'herd:S/P'
pîîl – pèl	'grindstone:S/P'	kêɛl – kěɛɛl	'kind.of.cheetah:S/P'
lûum – lùm	'grass:S/P'	mʌ̂ʌṭ – mʌ́ʌʌṭ	'friend:S/P'
tûuŋ – tùŋ	'horn:S/:P'	jîîṭ – jìiiṭ	'well:S/P'
kîic – kìc	'orphan:S/P'	jêel – jêeel	'bracelet:S/P'
lôʊl – lùl	'steep.bank:S/P'	pôʊl – pʊ́ʊʊl	'widest.part.of.lower.leg:S/P'

The importance of the data in Table 7 to the present discussion on quantity is that they show that the stem vowel can either lengthen or shorten in the morphological marking of number. Shortening is illustrated by the nouns on the left: these illustrate a decrease in vowel length in the marking of plural number. (As explained in relation to vowel shortening in transitive verbs, the vowel qualities /ɪ,ʊ,e,o/ are avoided in derived short vowels, through raising or through lowering.) 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 Tables 3 and 4, i.e., Fixed Short, Short with Grade, and Long with Grade. By implication, all of the singular nouns in Table 7 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 'grass' is lúuum̀. Just as in transitive verbs, then, morphological shortening is derivational and not inflectional. Also in nouns and verbs alike, morphological shortening of the stem vowel is only found in stems that have a high

---

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-lîîɲṭ-à 'I saw the forest'. The unmarked singular tìm does not have floating quantity: there is no increased duration of the following past-tense marking prefix, which is realised [álî::ḍà].

(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 hypothesized three-level vowel length contrast has both lexical and morphological functions. As for the lexical role of vowel length, short and long vowels are commonly found in lexical roots. Roots with an overlong vowel are also attested, but they are few in number (see Table 5). Presumably, such roots represent the diachronic outcome of valency-increasing processes whereby the base has been lost subsequently. As for the morphological role of quantity level, we find the following three alternations: CVC~CVVC, CVC~CVVVC, CVVC~CVVVC. The latter two patterns of alternation have the highest functional load, and they 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 several inflections and derivations (CVC~CVVVC, CVVC~CVVVC). In addition, long lexical roots with a Low Fall tone additionally display shortening of the stem vowel (CVC~CVVC). Across word classes, vocalic shortening is invariably derivational in nature. In conclusion, the hypothesized three-level vowel length contrast serves to distinguish unrelated lexical forms, and it also marks a wide range of morphological processes.

2.2 Competing hypotheses

In Section 2.1, we described Shilluk syllable-internal quantity in terms of the three-level 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 does not violate 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 discerned. Three decades ahead of Andersen's (1987) paper postulating this configuration for Dinka, Tucker (1955) makes the following observations:

> 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." [Tucker 1955:427]

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 character for a half-long vowel to transcribe long vowel length. Further on in the same paper, he emphasizes the importance of this three-way distinction, i.e., "very short, medium, very long" in morphophonological alternations (Tucker 1955:432). More recently, three-level vowel length was hypothesized for Shilluk phonology in Remijsen, Ayoker & Mills (2011), and its role in Shilluk morphology is described in Remijsen, Miller-Naudé & Gilley (2015).

The 3VL hypothesis for Shilluk represents lexical contrast and morphological contrast in terms of the same phonological dimension of contrast, i.e., vowel length (cf. Section 2.1). This is a crucial difference with a competing analysis of syllable-internal quantity in Shilluk, presented in Miller & Gilley (2001, 2007) and in Gilley (2003).[10] According to this hypothesis, syllable-internal quantity in Shilluk is represented through a combination of a) binary vowel length contrast, marking lexical contrast, and b) a stress contrast, marking morphological contrast. We refer to this hypothesis hereafter 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 the lack of stress is found in morphologically marked forms (Gilley 2003:118). For example, the past tense object voice form, which we hypothesize to display the quantity of the root (cf. Table 2), has a stressed stem syllable according to the 2VL + Stress hypothesis (Miller & Gilley 2001). The realization of stress is described as follows:

---

[10] In Gilley (2003), the analysis is extended to Dinka.

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. In [the following illustration] the different lengths and stress patterns are shown schematically.

short stressed ˈ|___|
short unstressed ˈ|_____|
long stressed ˈ|_____|
long unstressed ˈ|_____| [Gilley 2003:104]

As seen from Gilley's description, she finds that vowels in Shilluk "sound as though they exhibit three different lengths". But this apparent three-level contrast is represented as – to speak with Odden (2011: 465) – "an epiphenomenon arising from binary length and some orthogonal dimension", in this case stress. The primary phonetic realization of the hypothesized stress feature is anomalous: across the world's languages, greater duration of segments in stressed syllables compared to corresponding segments in unstressed syllables is the most reliable acoustic correlate of stress (e.g. Ortega-Llebaria & Prieto 2011; van Heuven 2014). The primary correlate of the stress feature postulated in Gilley (2003) is the opposite: stressed syllables display reduced vowel duration.

Apart from vowel duration, Gilley mentions several other phonetic characteristics of the hypothesized 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 2003:104). She also states that stress affects tone, suggesting that this may be due to time pressure. Elsewhere, she describes unstressed syllables as more peripheral (Gilley 2003:102). In Section 3 we will report on a detailed investigation of the phonetic realization of the quantity contrast.

We can compare the competing hypotheses – three-level vowel length (3VL) vs. two-level vowel length plus stress (2VL + Stress) – on the basis of evidence from antipassives, using sound data embedded for the sake of accountability. Antipassives are derived from transitive verbs through a range of suprasegmental changes, and quantity is one of these. In Section 2.1, we represented the changes in terms of 3VL. According to 2VL + Stress, antipassive verbs are invariably unstressed,

whereas the corresponding transitive verb forms are stressed (Miller & Gilley 2001:43). Table 8 presents a lexical minimal set for vowel length, consisting of the transitive verbs {nàŋ} 'lick' vs. {nɛ́aŋ} 'seal, fill'.[11] Both of these verbs display increased quantity in the antipassive. If the change in quantity is realized by a stress feature in line with the 2VL + Stress hypothesis, then the antipassive forms of {nàŋ} 'lick' and {nɛ́aŋ} 'seal, fill' 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 hypothesized stress feature, this would bring a short unstressed vowel to the duration of a long stressed vowel. In turn, a long unstressed vowel should display a vowel that is saliently longer in duration. According to the 3VL, morphological lengthening is overlengthening (cf. Section 2.1). This means that verbs with a lexically short vowel and those with a lexically long vowel both display an overlong vowel in the antipassive, thereby neutralizing the lexical contrast.

Table 8. Transcriptions showing morphological quantity alternations in Short with Grade {nàŋ} 'lick' and Long with Grade {nɛ́aŋ} 'seal'. Separate transcriptions are shown for the competing hypotheses, i.e., three-level vowel length (3VL) and vowel length plus stress (2VL + Stress).

3VL	á-nɛ́ŋ	á-nʌʌʌŋ	á-nɛ́aŋ	á-nʌʌʌŋ
2VL + Stress	á-ˈnāŋ	á-nĀŋ	á-ˈnāaŋ	á-nĀʌŋ
	PAST-lick:OV	PAST-lick:ATP	PAST-seal:OV	PAST-seal:ATP
	'It was licked.'	'S/he was licking.'	'It was sealed.'	'S/he was sealing.'

Playing the embedded sounds, the reader will find 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 Table 9.[12]

---

[11] Shilluk displays topic-drop, to the effect that preverbal internal arguments can be omitted if they have a 3rd person singular referent and they can be inferred from the preceding context. As a consequence, the utterances in Tables 8 and 9 are grammatically well-formed clauses.

[12] The verbs {càm} 'eat' and {cɛ́am} 'feed' are undoubtedly related; this derivation is sporadic / rare.

Table 9. Transcriptions showing morphological quantity alternations in Short with Grade {càm} 'eat' and Long with Grade {cá̂am} 'feed'. Separate transcriptions are shown for the competing hypotheses, i.e., three-level vowel length (3VL) and vowel length plus stress (2VL + Stress).

3VL	á-cá̂m	á-cʌ̄ʌʌm	á-cá̂am	á-cʌ̄ʌʌm
2VL + Stress	á-ˈcām	á-cʌ̄m	á-ˈcāam	á-ˈcʌ̄ʌm
	PAST-eat:OV	PAST-eat:ATP	PAST-feed:OV	PAST-feed:ATP
	'It was eaten.'	'S/he was eating.'	'S/he was fed.'	'S/he was feeding.'

The 2VL + Stress hypothesis runs into further difficulties. If the morphological marking of antipassive through quantity is represented as a loss of stress, we should find greater vowel duration in the antipassive for all transitive verbs. The data contradict this in two ways. First, there are the Fixed Short verbs, which do not display any quantity alternation: e.g. á-ŋɔ́l 'PAST-cut' > á-ŋ̀ùt 'PAST-cut:ATP'. Second, as noted above (Table 6), there is a set of verbs – i.e., members of the Long with Grade / Low Fall class that have a closed vowel, i.e. /ɪ,i,u,ʊ/ – where the stem vowel displays reduced quantity in the antipassive. Table 10 presents transcriptions of an example in terms of the competing hypotheses. As heard in the sound examples, the antipassive of {gûur} 'grind' has a shorter vowel than the corresponding transitive.

Table 10. Transcriptions showing the morphological quantity alternations in Long with Grade {gûur} 'grind'. Separate transcriptions are shown for 3VL and 2VL + Stress.

3VL	á-gû́ur	á-gûr
2VL + Stress	á-ˈgūr	á-gùr
	PAST-grind:OV	PAST-grind:ATP
	'It was ground.'	'S/he was grinding.'

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

18

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, 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. Illustration (5) displays two examples from Gilley (2003:107), which are to show the role of stress in a syntactic voice alternation. The preverbal topic slot accommodates the thematic subject in (5a), but the beneficiary in (5b). Stress is hypothesized to be the sole dimension of morphophonological contrast here. Given that the 2VL + Stress postulates that stress goes with decreased vowel duration, Gilley's transcription predicts that the stem vowel of the verb has an audibly shorter duration in (5b) than in (5a). Our own transcriptions are included as well. As seen from these, we represent this alternation through tone rather than quantity.[13]

(5) a. **ānéj á-cwōttì   kī ḍáṉ-ɔ̀**   [2VL + Stress]
         ànéj á-cwòtì   kí ḍâaaṉ-ɔ̀   [3VL]
         Aney PAST-call:AMB PRP person-S
         'Aney called a person.'

     b. **ānéj á-ˈcwòttì   kī ḍáṉɔ̀**   [2VL + Stress]
         ànéj á-cwôtì   kí ḍâaaṉɔ̀   [3VL]
         Aney PAST-call:AMB:OV PRP person-S
         'Somebody called a person for Aney.'

The sound examples 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 (5b), whereas it has a level low pitch in (5a). In contrast, the verbs do not sound 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

---

[13] Specifically, when the ambitransitive form of the verb has a Low-toned stem syllable in the morphologically unmarked form, i.e., AV(O), 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.

the concept, stress is a) culminative and b) obligatory (Hyman 2006). As a culminative feature, stress singles out a syllable within a string of syllables. As an obligatory feature, there should be no phonological words that are not marked for stress. The hypothesized stress feature of Miller & Gilley (2001, 2007) and Gilley (2003) displays neither of these features. Intransitive verbs, for example, are hypothesized not to carry stress (Miller & Gilley 2001); and monosyllabic words can carry stress: e.g. ˈlwák 'cattle byre' in Gilley (2003:108). The hypothesized stress feature is equally anomalous in terms of its phonetic realization: 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 re. the morphological alternations in which it is hypothesized to be involved, such as antipassive and syntactic voice alternations. Second, it is conceptually anomalous.

3. An acoustic study of three-level vowel length in Shilluk

We now put the 3VL hypothesis to the test by means of a production study. The hypothesized vowel length contrast is manipulated while other factors are 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 Section 3.1, and the results are reported in Section 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, then this would offer support for an alternative interpretation, whereby a phonological dimension orthogonal to vowel length is involved (cf. Odden 2011). For example, if the hypothesized levels of short vs. long vowels can be differentiated reliably by F1 or F2, then the bottom half of the contrast can be represented in terms of phonemic vowel quality, and a binary vowel length contrast is sufficient.[14] The orthogonal dimension postulated in Gilley (2003) is stress, and she mentions centralization of

---

[14] This is in fact the analysis proposed in relation to the controversy of three-level vowel length in Low German by Kohler & Tödter (1984) and Kohler (2001).

vowels, i.e., vowel quality, and tone as phonetic characteristics (Gilley 2003:102).[15] By measuring F1, F2, and F0, we can evaluate that alternative hypothesis. We also measure coda duration, on the one hand because gemination of stem-final consonants has been postulated in earlier work on Shilluk (Gilley 1992, Noske 1995), and also because coda duration may interact with a vowel length contrast, in a complementary fashion, with consonant duration being greater following vowels that have a lower degree of quantity, either in a phonetic sense or in a phonological sense (Elert 1964, Schaeffler, Wretling & Strangert 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 11. 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 morphological contrast. All of the targets are nouns in a morphosyntactic sense. Many are agent-oriented nominalisations derived from transitive verbs. The third level of vowel length is invariably represented through the pertensive inflection, an instance of head-marking in possessive noun phrases (cf. footnote 5).

Table 11. The target words used in the production study.

Vowel length	Short	Long	Overlong
/ɛ/	léŋ 'beat.drum:NOM'	léɛŋ 'throw:NOM'	léɛɛŋ 'throw:NOM:PERT:P'
	kél 'peel.w.teeth:NOM'	kéɛl 'separate:NOM'	kéɛɛl 'separate:NOM:PERT:P'
/a/	pál 'dodge:NOM'	páal 'surgery.knife'	páaal 'surgery.knife:PERT:P'
	lám 'pray:NOM'	láam 'prayer'	láaam 'pray:NOM:PERT:P'
/ɔ/	kɔ́l 'take.out:NOM'	kɔ́ɔl 'herd:NOM'	kɔ́ɔɔl 'herd:NOM:PERT:P'
	t̪ɔ́l 'eat.prematurely:NOM'	t̪ɔ́ɔl 'rope'	t̪ɔ́ɔɔl 'rope:PERT:P'
/u/	dúp 'mess.up:NOM'	dúup 'mouse'	dúuup 'mouse:PERT:P'
	gút 'stab:NOM'	gúut 'navel'	gúuut 'navel:PERT:P'

[15] Gilley (2003) also mentions that there may be pause following a word that ends in a stressed syllable. We have not found this to be the case, and in our dataset, there are no pauses to be measured following the target words.

The target words were elicited in utterance-medial position, using the frames illustrated in (6a,b,c) for forms with short, long, and overlong vowels, respectively. 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 (6a,b), the target nouns are in the base form and followed by a prepositional phrase. In the overlong condition (6c), the following constituent is a possessor modifier which expresses the agent of the referent event of the source verb.[16]

(6)   a. **dâa**   **lám**   **kì**   **kêɲ**

              EXIST pray:NOM PRP place:CS:DEM

              'There is praying here.'

     b. **dâa**   **láam** **kì**   **kêɲ**

              EXIST prayer PRP place:CS:DEM

              'There is a prayer here.'

     c. **dâa**   **láaam**   **jìi**

              EXIST pray:NOM:PERT:P people

              'There is praying by people.'

We also included in the design three items with an overlong vowel /i/ in the stem syllable: **kíiic** 'bee:PERT:P', **àbíiik** 'sorghum.flour:PERT:P', and **míiilɔ̀** 'Milo (name)'. These items are not taken into account in relation to the quantitative analyses. Instead, the formant measurements on the syllables with vowel /i/ are used to display the formant measurements on the target vowels in the context of the vowel space as a whole. To do this, we need 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, and that is why these items are collected in addition.

---

[16] The carrier sentence for the overlong test word in (6c) is one syllable shorter than the carrier sentences in (6a) and (6b). 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 salient only between one-word and two-word sentences (Windmann, Šimko & Wagner 2015). To the best of our knowledge, there is no evidence suggesting that the difference between three vs. four syllables has a significant influence on segment durations.

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.[17] Six of them grew up in the southern part of the Shilluk region, known as Lwak, and the other two grew up in the northern part, known as 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:P') – 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 the tone on the target to be High. 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, both in relation to the noun block and in relation to the verb block, yielding four different orders overall 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 the verb block were preceded by a practice item.

---

[17] The second author is one of the speakers.

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 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, which all speakers had some proficiency in. If need be, we described the target meaning in Shilluk using synonyms. There were only two following contexts, cf. (6) 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 kêɲ** 'here' context. Elicitation was carried out by the two 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 e.g. background noise affected one of the realisations.

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 realizations (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 the second formant (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 (F2) enabled us to determine the location of the boundary between a vowel and the closure phase of a following plosive. In the segmentation of 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: the duration of the nucleus; the duration of the coda; vowel quality (F1, F2), and fundamental frequency (F0). In relation to F1 and F2, we created the formant candidates using the Burg algorithm, 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 and 800 if the vowel was /u/. In four cases, this did not yield realistic values either, and these tokens were discarded. In relation to 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 at this point in the vowel.

Ahead of the descriptive and inferential statistics, we aggregated the measurements over repetitions, so as not to treat repetitions as independent observations. Doing so, we went from 361 tokens to 192 types. This is indeed the expected total, as there were eight minimal sets with three members each, collected from eight speakers, i.e., 8*8*3 = 192. So there are no missing values.

3.1.5 Statistical analyses

If the 3VL hypothesis is true, then short, long and overlong vowels differ primarily and substantially in terms of vowel duration. The other measures (coda duration, F1 and 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: linear mixed-effects (LME) analysis combined with likelihood ratio testing (Bates 2005; Baayen et al. 2008; Winter 2013). The crucial advantage of using LME 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, and also any interaction between the main effects. Then we take out interactions and factors from the model, one at a time. Each time, 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 all linear mixed effects models, both subjects and items were modeled 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 (alpha) at 0.01, i.e., lower than the usual level of 0.05.[18] Both in terms of the modeling of the random factors, and also in setting the significance threshold, we are mindful to 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, in relation to phonemic contrast, statistical significance is a low bar. That is, inferential tests such as the above-mentioned 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 phonological contrast, we can expect that measure to reliably predict category membership. LDA allows us to find out, for each item, whether its factor level in terms of the hypothesized 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 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 Figure 3; and MASS (Venables & Ripley 2002) for LDA.

---

[18] For a critical evaluation of the meaning of this threshold value, see Cohen (1994).

3.2 Results

3.2.1 Vowel duration

Figure 1.A 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 milliseconds, 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, neither between short vs. long, nor between long vs. overlong. This suggests that the three levels of Vowel length are produced with little overlap in terms of vowel duration.
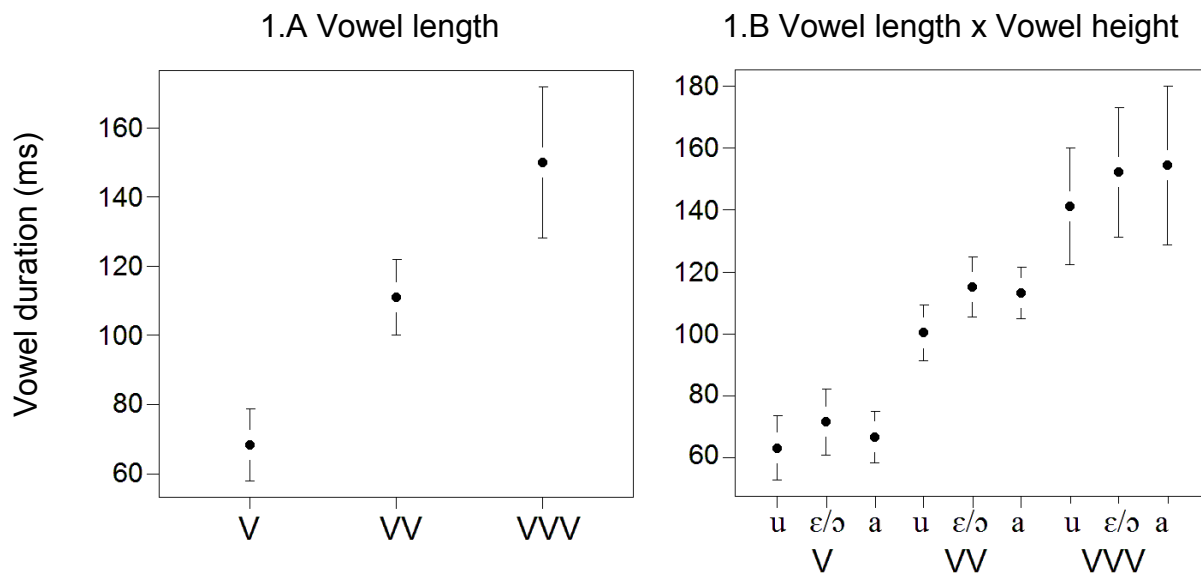
Figure 1. Means and standard deviations for vowel duration (ms) over items and speakers, by Vowel length – short (V) vs. long (VV) vs. overlong (VVV) – in Figure 1.A, and by Vowel length and Vowel height – /u/ vs. /ɛ,ɔ/ vs. /a/ – in Figure 1.B.

The variability around the mean 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, Figure 1.B shows that the high vowel /u/ has the shortest duration, across the 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 [$\chi^2(1) = 3.86$, $p = 0.0494$].[19]

3.2.2 Coda duration

The descriptive statistics for coda duration are presented in Figure 2. The mean values as a function of Vowel length for short, long and overlong vowels are 96, 84 and 104 ms, respectively. These are displayed in Figure 2.A. Figure 2.B shows the results broken down by type of coda consonant, i.e., /l/, N[asal] or P[losive]. Figure 2.B shows that there is not much difference in the mean duration between the different types of consonants, in 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 realized weakly. Both overall (Figure 2.A) and by coda consonant (Figure 2.B), 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.

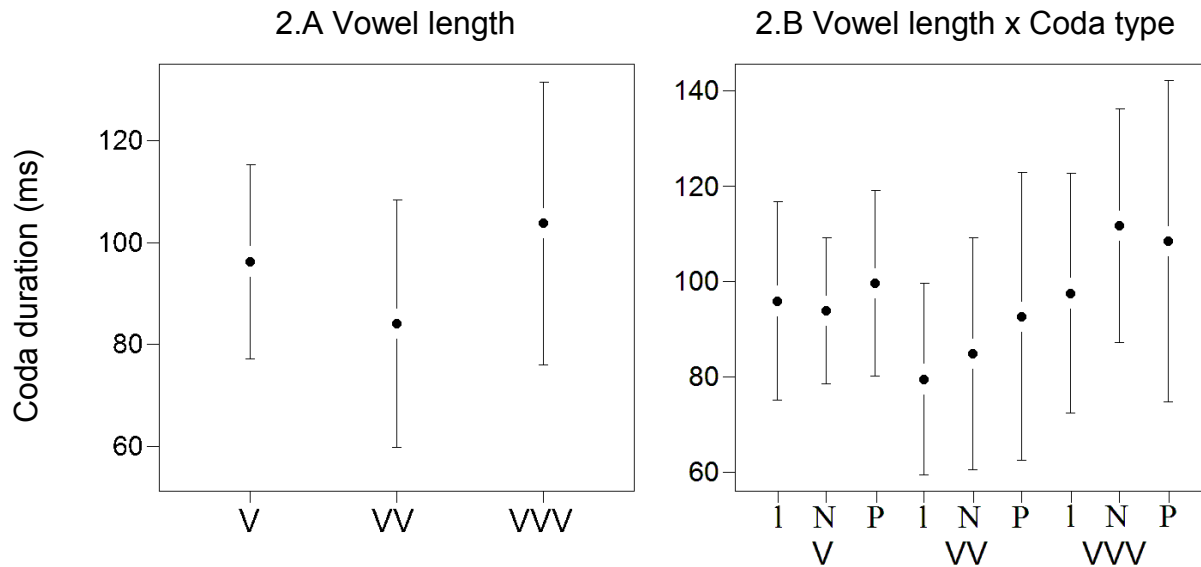[19] Alpha is set at 0.01 (cf. Section 3.1.5).

Figure 2. Means and standard deviations for coda duration (ms) over speakers and items, by Vowel length (V, VV, VVV) in Figure 2A, and by Vowel length and Coda type (/l/ vs. nasal [N] vs. plosive [P]) in Figure 2.B.

Considering the results in a cross-linguistic perspective, there is evidence that coda duration is inversely correlated with vowel duration (Maddieson 1985, Lehtonen 1970). 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 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 centralizing pressure on the realization of vowels, which is why longer vowel categories are likely to be more spaced out in the F1xF2 space than shorter vowels (Lehtonen 1970, Remijsen & Gilley 2008). Hence, the 3VL hypothesis predicts that short vowels may be centralized 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 centralization, then this lends support to the interpretation that another phonological dimension is involved.

Figure 3 displays the descriptive statistics for F1 and F2. Because speakers vary considerably in terms of the formant values, the raw F1 and F2 values were z-transformed by speaker (Lobanov 1971).[20] Z-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 the formant range. For each of the four vowels included in the design, that is /ε,a,ɔ,u/, Figure 3 displays three ellipses, one for each of the levels of Vowel length. These ellipses encircle 1 standard deviation of the values, after z-transformation by speaker. The mean values by vowel for each of the levels of vowel length have been 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. This means that the difference in vowel quality as a function of Vowel length is limited.
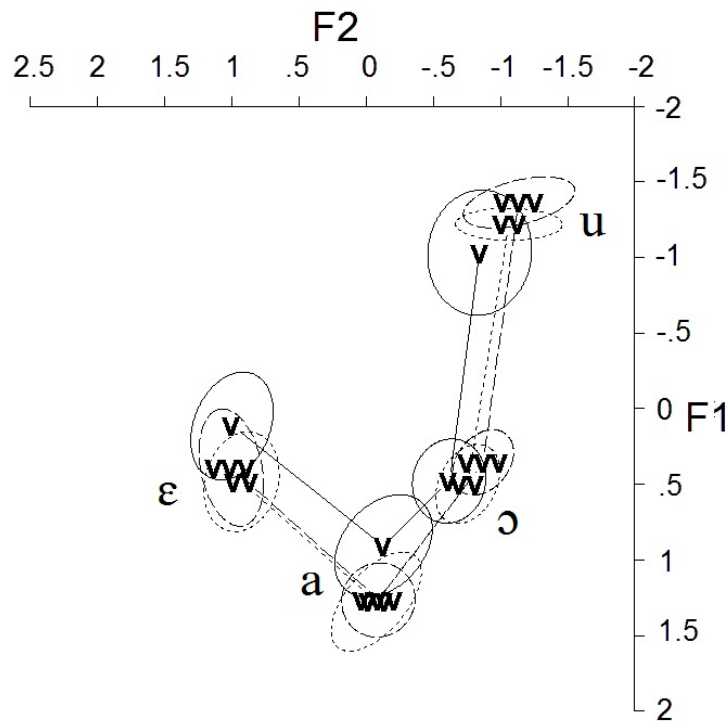

Figure 3. Means and standard deviations (as ellipses) for z-transformed F1 and F2, over speakers and items, by Vowel (/ε,a,ɔ,u/) and Vowel length (V, VV, VVV). The ellipses and connecting lines marking different levels of Vowel length are also distinguished by line style: full (V), dotted (VV), interrupted (VVV).

---

[20] As noted above (Section 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.

Second, for each of the four vowels, the ellipsis for the short level of length lies closest to the center of the vowel space. Because of this, the polygon connecting the mean values of short vowels (full line) lies within the polygons connecting the means of long and the overlong vowels. This result suggests that the time pressure that is inherent to the production of short vowels indeed exerts a centralizing 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 centralization, derived from the raw F1 and F2 values. This is necessary because the centralization observed in Figure 3 affects different vowels differently in terms of their F1 and F2. For example, whereas in the case of vowel /a/ the short vowels on average have a lower F1 value than the corresponding long and overlong vowels, in the case of the vowel /u/ the relation is the opposite, i.e., for this vowel the short vowels have a higher F1 value than the long and overlong vowels (see Figure 3).

The calculation of this measure of the degree of centralization is based on Becker-Kristal (2010). Specifically, we started with 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, for each F1xF2 measurement, its Euclidean distance to the corresponding speaker-specific centroid. The descriptive statistics are displayed in Figure 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 centralization are in line with the results in Figure 3: short vowels are on average closer to the centroid of the vowel space, as compared to long and overlong vowels, and overlong vowels tend to be more peripheral, as compared to 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.
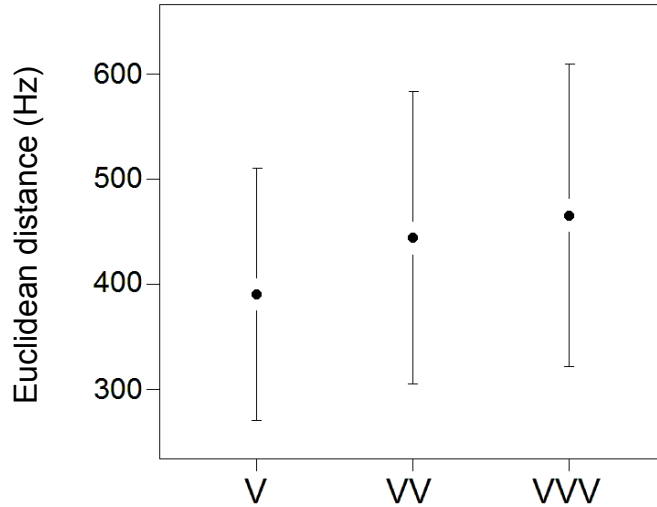
Figure 4. Means and standard deviations for the Euclidean distance from the speaker-specific center of the vowel space, by Vowel length (V, VV, VVV).

We then went on to use this measure of centralization (i.e., the Euclidean distance measure, in Hz) as the dependent variable in a linear mixed-effects model, with 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 centralization 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 also does not have a significant effect on the degree of centralization [$\chi^2(6) = 12.26$, $p = 0.0564$]. Vowel length registered 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 ($p = .0057$), while the other two pairwise combinations are not significant (short vs. long: $p = .0776$; long vs. overlong: $p = 1$).

3.2.4 Fundamental frequency

Our motivation in including F0 among the measurements stems from the concern to determine whether a phonological dimension other than vowel length could be plausibly invoked to represent any of the three levels of the hypothesized 3VL contrast. The 3VL hypothesis does not predict a difference in F0. In relation to 2VL + Stress, Gilley (2003) states that tone may be a

correlate of stress. Moreover, Gilley (2003) invokes stress in relation to morphophonological contrasts we hypothesize to be marked solely by tone.

The phonological specification for tone is the same for all target words: a level High tone (cf. Table 11). Because of the level nature of this specification, the measurement is taken at the temporal mid point of the vowel. The measurements reveal that the mean F0 values at the midpoint of short, long and overlong vowels diverge very little – they are 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 earlier research, which shows that, all else being equal, high vowels inherently have 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 random factors Speaker and Item. 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 analyses

Linear Discriminant Analysis (LDA) allows us to find out, for each item, whether its factor level in terms of the hypothesized 3VL contrast can be predicted on the basis of a given measurement or set of measurements. This is worthwhile to find this out, because it answers the question as to whether a particular measurement distinguishes between the category levels, and in this respect this technique is complementary to the inferential tests reported in Sections 3.2.1 to 3.2.4. We carried out five linear discriminant analyses. In each of these, the 192 items in the dataset were classified for Vowel length (short, long, overlong). In a first one, the classification was based on all four of the measurements for which inferential statistics are reported in Section 3.2.1 to 3.2.4: nucleus duration, coda duration, centralization, and F0 at temporal mid point. In the remaining four LDAs, classification was carried out on the basis of each of these four measurements by itself. The overall classification results of these analyses are reported in Figure 5. With three levels to be discriminated between, the chance-level base stands at 33 percent.
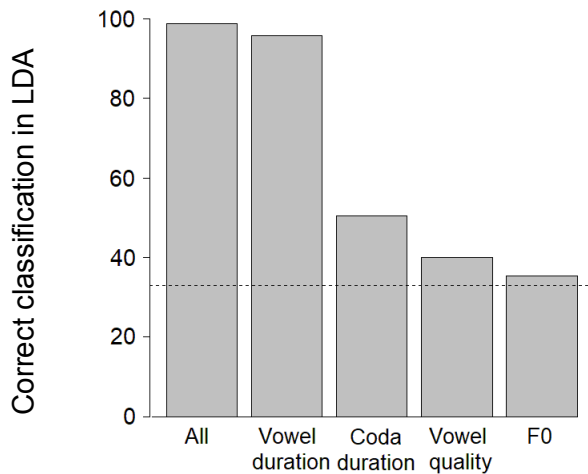
Figure 5. Correct classification for Vowel length (short, long, overlong) for LDAs: with all four of the measurements together, and on the basis of each of the measurements by itself. The chance-level baseline (33%) is represented as an interrupted line.

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

The similarity between the correct classification scores for the LDA with all four measures on the one hand and the one with vowel duration alone on the other hand suggests that, in the former, 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 function in the LDA with all four of the measurements in Table 12. Note that, in the most important discriminant function, Function 1, vowel duration has the correlation coefficient that stands out most (-3.59). In Function 2, the correlation coefficient of coda duration stands out most (-1.04). However, Function 2 accounts for very little of the discriminable variation: just 1.5 percent of the trace, as compared to Function 1, which accounts for 98.5 percent of trace.[21]

---

[21] The trace is the proportion of variance between factor levels, explained on the basis of successive discriminant functions.

Table 12. Correlation coefficients of the four measures in relation to the first and second discriminant functions in a LDA with all four measurements.

	Function 1	Function 2
Vowel duration	-3.59	-.05
Coda duration	-0.26	-1.04
Centralization	-0.37	0.27
F0	-0.12	-0.24

Table 13 reports the correct classification results for each of the three levels of Vowel length (short, long, overlong) separately in each of the five LDAs. In line with the overall results reported graphically in Figure 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 percent. 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 percent of items with a long vowel. In summary, the only measure that reliably distinguishes between levels of Vowel length is vowel duration.

Table 13. Correct classification by level of Vowel length, in an LDA with all four measurements together, and in LDAs with the four measurements by themselves.

Vowel length	Short	Long	Overlong
All	100	98	98
Vowel duration	98	94	95
Coda duration	27	70	54
Vowel quality (centralization)	55	16	50
F0	50	52	5

4. Discussion and conclusion

The descriptive analysis of the role of syllable-internal quantity in Shilluk phonology and morphophonology in Section 2.1 shows that the hypothesized three-level 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 three-level vowel length has a wider scope, in that some lexical items have an overlong vowel throughout their paradigm (cf. Table 5).

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 alternations short~overlong and long~overlong have a high functional load in the inflectional paradigms of verbs and nouns, through processes of overlengthening. 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 Section 2.2 we evaluated a competing analysis, whereby Shilluk is postulated to have a binary vowel length contrast plus stress (Miller & Gilley 2001, Gilley 2003). We rejected this hypothesis (2VL + Stress) 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 in fact is not the case. Second, the morphological processes for which stress has been invoked in fact do not all involve increased quantity consistently. 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 in fact quantity-increasing for most transitive verbs but quantity-decreasing for others. Finally, we noted that stress was invoked to represent morphological processes that are in fact marked by tone. As for the theoretical concerns, the hypothesized stress contrast is anomalous in several ways: it is neither obligatory nor culminative, and its durational marking is the opposite of what is attested cross-linguistically, i.e., stressed syllables are characterized by a reduction in vowel duration.

The acoustic analysis of the phonetic realisation of the hypothesized three-level vowel length contrast in Section 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 out, at

68, 111, and 150 milliseconds, 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 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 percent, similar to the correct classification result obtained with all four measures entered together, which is 99 percent. Of the other three measures, coda duration registers the highest correct classification result, at 50 percent.

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, centralization and coda duration can be interpreted as secondary correlates at best. This state of affairs is in line with the three-level vowel length hypothesis, but not with alternative interpretations that account for the quantity contrast under investigation through a combination of binary vowel length and another dimension of phonological contrast.

At the beginning of this report, 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 affirmative. This conclusion is important, because the vast majority of contrasts in vocalic quantity in the world's languages are binary in nature, i.e., short vs. long, and the evidence for three-way contrast 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 three-level vowel length. This indicates that the bias towards binarity in quantity contrasts is not absolute.

Blevins (2004) attributes the development of three-level vowel length in Dinka to its morphological load; this explanation can be extended to Shilluk.

Our findings strengthen the case for three-level vowel length in West Nilotic. As noted in the introduction, three-level 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, Monich 2016 on Nuer). However, so far the base of phonetic evidence has been limited (Remijsen & Gilley 2008 on Dinka). The research reported here strengthens this base of evidence.

Given the conclusion that Shilluk presents a three-way contrast in vocalic quantity in the syllable nucleus, the question presents itself, as to what the best phonological representation of this contrast is. In relation to Dinka, Hayes (1989) argued that three-level vowel length can be represented straightforwardly by associating one, two, and three weight units or morae (μ) with syllables with a short, long, and overlong vowel, respectively. This representation can be extended to Shilluk. This is shown in (7), using one of the minimal sets of the production study.

(7)

σ	σ	σ
p á l	p áa l	p áaa l
dodge:NOM	surgery.knife	surgery.knife:PERT:P
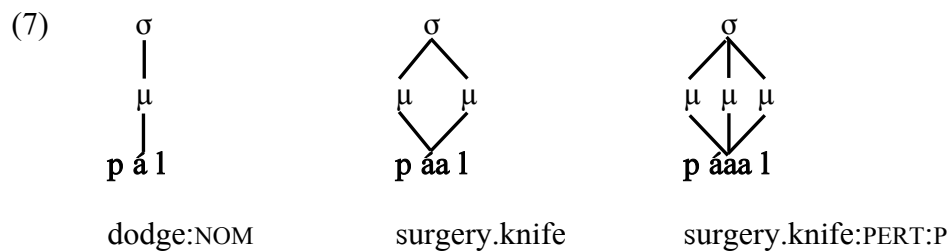
The phonological representation illustrated in (7) represents phonetic quantity, specifically vowel duration, directly in terms of phonological quantity in the syllable nucleus. We interpret this direct relation between the primary correlate of a contrast and its phonological representation as an advantage 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, one could postulate mora-sharing between nucleus and coda, along the lines of the analysis proposed in Baal, Odden & Rice (2012) to represent a 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, there is no pairwise combination of levels of the contrast that is distinguished by coda duration as well as it is by vowel 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 realization, representation and explanation for three-level vowel length in Shilluk and in other languages. If it is correct to conclude that 3VL is real, a key question is why it is so uncommon (cf. Blevins 2004, Remijsen & Gilley 2008). One area of research that holds the promise to shed light on this question is speech perception. As it stands, there is no research available on how a three-level vowel length contrast is perceived. It is conceivable that the just-noticeable difference (JND) for the perception of durational contrasts represents a constraining factor (cf. Remijsen & Gilley 2008 and further references there). Equally, it would be worthwhile to find out whether native speakers of languages that present 3VL (e.g. Shilluk, Dinka) perceive durational differences in speech in a more fine-grained manner, so that their just-noticeable difference (JND) for duration would be smaller than that of native speakers of languages that present less complex systems of vocalic quantity.

Abbreviations

AMB	Ambitransitive	ITER	Iterative derivation
ATP	Antipassive	OV	Object Voice
CS	Construct state	NEVP	Non-evidential derivation
DECL	Declarative	P	Plural
DEM	Demonstrative	PET	Centripetal derivation
EXIST	Existential marker	NOM	Nominalisation
FUG	Centrifugal derivation	PERT	Pertensive
FUT	Future	PRP	Preposition
IMP	Imperative	S	Singular
IMPF	Imperfective	XV	Applicative voice
INV	Inversion		

Acknowledgements

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 (1992-1994). Morphological stratification in Dinka: on the alternations of voice quality, vowel length and tone in the morphology of transitive verbal roots in a monosyllabic language. *Studies in African Linguistics* 23(1), 1-63.

Baal, Berit A.B., David Odden, Curt Rice (2012). An analysis of Saami gradation. *Phonology* 29, 165-212.

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

Baayen, R. H., D. J. Davidson, and Douglas 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.

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

Bates, Douglas, Martin Maechler, Ben Bolker, Steve 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. UCLA PhD dissertation. [http://phonetics.linguistics.ucla.edu/research/RBecker_diss.pdf]

Blevins, Juliette (2004). *Evolutionary Phonology – The Emergence of Sound Patterns*. Cambridge University Press

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

Clements, Nick, Alexis Michaud & Cédric Patin (2011). Do we need tone features? In John A. Goldsmith, Elizabeth Hume & Leo Wetzels (eds.) *Tones and Features.* Berlin: De Gruyter Mouton, 3-24.

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 Robert Nicolaï & Franz Rottland (eds.) *Cinquième Colloque de Linguistique Nilo-Saharienne / Fifth Nilo-Saharan Linguistics Colloquium. Nice, 24-29 août 1992.* Cologne: Rüdiger Köppe Verlag, 147-168.

Dixon, R.M.W. (2010). *Basic linguistic theory – volume 2 grammatical topics*. Oxford University Press.

Duanmu, San (2009). *Syllable Structure: The Limits of Variation*. Oxford University Press.

Elenbaas, Nine & René Kager (1999). Ternary rhythm and the lapse constraint. *Phonology* 16, 273-329.

Elert, C.-C. (1964). *Phonological Studies of Quantity in Swedish*. Uppsala: Almqvist & Wiksell.

Flack, Kathryn (2007). Templatic morphology and indexed markedness constraints. *Linguistic Inquiry* 38, 749-758.

Gilley, Leoma (1992). *An Autosegmental Approach to Shilluk Phonology*. Dallas (TX): Summer Institute of Linguistics and The University of Texas at Arlington Publications in Linguistics 103.

Gilley, Leoma (2003). The feature of stress in Nilotic. In Rose-Juliet Anyanwu (ed.). *Stress and Tone – the African Experience*. Frankfurter Afrikanistische Blätter 15, 99-120. Cologne: Rüdiger Köppe Verlag.

Haspelmath, Martin (1996). Word-class changing inflection and morphological theory. In Geert Booij and Jaap van Marle (eds.) *Yearbook of morphology 1995*. Dordrecht: Kluwer, 43-66.

Hayes, Bruce (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20, 253-306.

Hermans, Ben & Francesc Torres-Tamarit (2013). Ternary constituents are a consequence of mora sluicing. In John Kingston, Claire Moore-Cantwell, Joe Pater & Robert Staubs (eds.), *Proceedings of Phonology 2013* (University of Massachusetts, Amherst). Linguistic Society of America.

House, Arthur S. (1961). On vowel duration in English. *Journal of the Acoustical Society of America* 33, 1174-1178.

House, Arther S. & Grant Fairbanks (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America* 25, 105-113.

Hyman, Larry M. (2006).Word-prosodic typology. *Phonology* 23, 225-257.

Hyman, Larry M. (2011). Do tones have features? In John A. Goldsmith, Elizabeth Hume & Leo Wetzels (eds.) *Tones and Features.* Berlin: De Gruyter Mouton, 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: C. Winter, 385-402.

Kohler, Klaus J. & Regina Tödter (1984). Experimentalphonetische Untersuchungen zur 'Überlänge'im Nieder- un Hochdeutschen Schleswig-Holsteins. *Institut ür Phonetik Universität Kiel, Arbeitsberichte* 21, 61-132.

Lehtonen, Jaakko (1970). *Aspects of Quantity in Standard Finnish.* Jyväskylä University Press, Jyväskylä, Finland.

Letterman, Rebecca (1994). A phonetic study of Sinhala syllable rhymes. *Working Papers of the Cornell Phonetics Laboratory* 9, 155-182.

Lobanov, B.M. (1971). Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America* 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.* University of Salford, Manchester, 419-441.

Mackenzie, Sara (2016). Consonant harmony in Nilotic: contrastive specifications and Stratal OT. *Glossa: a journal of general linguistics* 1(1), 12. [DOI: http://doi.org/10.5334/gjgl.8]

Maddieson, Ian (1985). Phonetic Cues to Syllabification. In Victoria Fromkin (ed.) *Phonetic Linguistics: Essays in Honor of Peter Ladefoged.* Orlando: Academic Press, 203-221.

Miller, Cynthia L. and Leoma G. Gilley (2001). Evidence for ergativity in Shilluk. *Journal of African Languages and Linguistics* 22: 33–68.

Miller, Cynthia L. and Leoma G. Gilley (2007). Evidentiality and mirativity in Shilluk. In Mechthild Reh & Doris L. Payne (eds.). *Advances in Nilo-Saharan linguistics: Proceedings of the 8th International Nilo-Saharan Linguistics Colloquium,University of Hamburg, August 22–25, 2001.* 191–206. Cologne: Rüdiger Köppe Verlag.

Monich, Irina (2016). Vowel length in Nuer. Proceedings of the 2016 Annual Meeting on Phonology [http://journals.linguisticsociety.org/proceedings/index.php/amphonology/index].

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]

Nooteboom, Sieb G. (1972). *Production and perception of vowel duration – A study of durational properties in Dutch.* PhD dissertation, University of Utrecht

Noske, Manuela (1995). Consonant Gemination in Shilluk. In Robert Nicolaï & Franz Rottland (eds.) *Cinquième Colloque de Linguistique Nilo-Saharienne / Fifth Nilo-Saharan Linguistics Colloquium. Nice, 24-29 août 1992*. Cologne: Rüdiger Köppe Verlag, 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*. Wiley-Blackwell, 465-490.

Ortega-Llebaria, Marta & Pilar Prieto (2011). Acoustic correlates of stress in Central Catalan and Castilian Spanish. *Language & Speech* 54, 73-97.

Prehn, Maike (2012). *Vowel quantity and the fortis-lenis distinction in North Low Saxon*. Amsterdam: University of Amsterdam PhD dissertation (LOT dissertation series volume 291).

R Core Team (2016). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. [URL https://www.R-project.org/].

Reid, Tatiana (2010). Aspects of phonetics, phonology and morphophonology of Thok Reel. University of Edinburgh MSc dissertation.

Remijsen, Bert & Leoma Gilley (2008). Why are three-level vowel length systems rare? Insights from Dinka (Luanyjang dialect). *Journal of Phonetics* 36, 318-344.

Remijsen, Bert (2013). Tonal alignment is contrastive in falling contours in Dinka. *Language* 89, 297-327.

Remijsen, Bert, Otto G. Ayoker & Timothy Mills. Shilluk. *Journal of the International Phonetic Association* 41, 111-125

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

Remijsen, Bert & Otto G. Ayoker (2014). Contrastive tonal alignment in falling contours in Shilluk. *Phonology* 31, 435-462.

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 University Press, 577-596.

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.

Remijsen, Bert, & Otto G. Ayoker (2018). Forms and functions of the base paradigm of Shilluk transitive verbs. *Language Documentation & Conservation*, Special publication no. 14: A grammar of Shilluk, Chapter 1. [http://hdl.handle.net/10125/24777]

Remijsen, Bert, & Otto G. Ayoker & Signe Jørgensen (2018). Shilluk_2016_controlled_ThreeLevelVowelLength. Edinburgh Datashare. [http://dx.doi.org/10.7488/ds/2424]

Rice, Keren (2014). Sounds in grammar writing. In Toshihide Nakayama and Keren Rice (eds.) *The art and practice of grammar writing. Language Documentation & Conservation*, Special Publication 8, 69-89.

Schaeffler, Felix, P. Wretling, P., & E. Strangert (2002). On the Development of a Quantity Typology for Swedish Dialects. *Proceedings of Speech Prosody 2002*, 627-630.

Torres-Tamarit, Francesc & Peter Jurgec (2015). Lapsed derivations: Ternary stress in harmonic serialism. *Linguistic Inquiry* 46, 376–387.

Tucker, A.N. (1955). The verb in Shilluk. *Mitteilungen des Instituts für Orientforschung* 3, 421-462.

van Heuven, Vincent J. (2014). Acoustic correlates and perceptual cues of word and sentence stress: Mainly English and Dutch. In Carlos Gussenhoven, Yiya Chen, & Dan Dediu (eds.) *The 4th International Symposium on Tonal Aspects of Languages, Nijmegen, The Netherlands, May 13-16, 2014*. International Speech Communication Association ISCA. 211-217.

Venables, W. N. & Ripley, B. D. (2002) *Modern applied statistics with S*. Fourth Edition. Springer, New York.

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 f0 of vowels. *Journal of Phonetics* 23, 349-366.

Windmann, Andreas, Juraj Šimko & Petra Wagner (2015). Polysyllabic shortening and word-final lengthening in English. *Proceedings of Interspeech 2015*, 36-40.

Winter, Bodo (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499. [http://arxiv.org/pdf/1308.5499.pdf]

Yip, Moira (1980). *The tonal phonology of Chinese*. Massachusetts Institute of Technology PhD dissertation. [https://dspace.mit.edu/handle/1721.1/15971]

Yu, Alan C.L. (2008). The phonetics of quantity alternation in Washo. *Journal of Phonetics* 36, 508-520.

Xu, Tongling (2017). Number marking in Shilluk. University of Edinburgh Hons dissertation.