Tonal alignment is contrastive in falling contours in Dinka

Bert Remijsen (The University of Edinburgh)

Abstract

This study investigates a contrast in tonal alignment, involving falling contours in Dinka. Such a contrast, first reported in Andersen (1987), calls into question the assumption that tonal alignment cannot distinguish contour tone patterns of the same shape within the syllable domain (Odden 1995:450, Silverman 1997:479-480, Yip 2002:29). A qualitative description of the Dinka phenomena is followed by a detailed production study, in which the realization of the contrast is investigated under manipulations of time pressure (cf. Caspers & van Heuven 1993). The primary correlate of the contrast is indeed tonal alignment: the early-aligned fall sets in during the onset or early in the vowel; the late-aligned peak sets in well into the vowel. The time pressure study also suggests that it is unlikely for more than two patterns of alignment to be accurately produced and perceived, given various phonetic limitations (Xu & Sun 2002; ’t Hart, Collier & Cohen 1990; House 2004). On this basis, it is argued that a representation with reference to mora edges (Prieto, D’Imperio & Gili Fivela 2005; Morén & Zsiga 2006) is less attractive: this representation is not sufficiently restrictive. Instead, the contrast is represented using a binary feature [± late-aligned]. This representation is adequate in an explanatory sense, in that the category boundary is in line with the quantal threshold hypothesized in the model of pitch perception of House (1990). The results also strengthen the case for three-level vowel length in Dinka (Andersen 1987, Remijsen & Gilley 2008).
1. Introduction

1.1 Motivation

Tonal alignment refers to the timing of fundamental frequency (f0) patterns relative to the sequence of speech segments. Many linguists assume that the alignment of tone contours within a syllable is not contrastive (Odden 1995:450,474; Silverman 1997:479-480; Yip 2002:29). This view is spelled out clearly by Odden (1995:474): “it might be that in some languages pitch changes are timed relatively early in the syllable, and in other languages they are timed relatively late. Such control would only be phonetic, never phonological.” This assumption implies that “[i]f the contours are composed of levels, the existence of two falls [in a tonal inventory] implies at least three levels” (Yip 2002:29). In other words, a contrast between two falling contours necessarily involves a difference in the height of the tone targets, rather than in their alignment. The same assumption is implicit in Silverman (1997), who argues that a falling contour resulting from tone sandhi would neutralize with an underlying falling contour within the same range. The common assumption is expressed as a hypothesis in (1).

(1) The alignment of falling f0 contours within a syllable is not distinctive.

The hypothesis in (1) is specific to the syllable domain. In other words, it does not rule out instances of contrastive alignment involving a high target followed by a low target, over a domain that extends beyond a single syllable. The latter configuration is well attested – e.g. Frota (2002) on two intonational tone patterns in European Portuguese. Also beyond the scope of (1) are phenomena whereby two contrasting tone patterns diverge in the alignment of a target, but the overall shape within the syllable is different, as in the contrast between the Fall and High tone categories of Thai (Morén & Zsiga 2006 and further references there). Restricted in this way, (1) is warranted in the face of the available evidence from the world’s languages. In his survey of 187 tone languages that have contour tones, Zhang (2001, 2002) makes no mention of phenomena that would challenge it. In fact, Zhang reports that the distribution of contour tones is often restricted to contexts that offer more sonorous duration within the syllable rhyme, such as environments involving a long vowel, a sonorant coda, stress, and utterance-final lengthening. If languages struggle to implement contour tones consistently across contexts, then it is unlikely that they could implement contrastive contour tones of the same shape within a syllable. Several phonetic limitations underlie this interaction between
contour tones and sonorous duration within the rhyme. On the production side, there is the time required to implement an f0 change. The report on this issue by Xu & Sun (2002) suggests that it takes 124 ms to realize a fall in f0 of 4 semitones (e.g. 150 to 120 Hz) – a time interval exceeding the duration of a vowel in many environments. On the perception side, one limitation is the glissando threshold, which reflects the hearer’s ability to perceive f0 changes as pitch contours rather than as pitch levels (Rossi 1971; Greenberg & Zee 1979; ’t Hart, Collier & Cohen 1990). This threshold is determined by the size and duration of the f0 change. A second perceptual limitation relates to the difference in timing relative to the segmentals for two otherwise identical sequences of tone targets to be reliably distinguished. Surveying the results of studies bearing on this question, House (2004) finds that the threshold is at or above 50 milliseconds (ms) in most studies. This threshold sensitivity to differences in the timing of a tone target can be wide enough to turn a falling f0 pattern over the syllable into a rise (D’Imperio & House 1997).

Against this background, this paper investigates the phenomenon of contrastive alignment in Dinka, a Nilo-Saharan language. In the context of a descriptive analysis of Dinka phonology, Andersen (1987) postulated a lexical and morphological contrast between two falling contours, which diverge primarily in terms of alignment within the syllable. On the basis of auditory impressions, Andersen (1987:20) describes one as “fall […] followed by level pitch”, and the other as “level pitch [...] followed by fall”. The two contours are identical in terms of Andersen’s numerical representation of the excursion size of the pitch fall. In an investigation of a different dialect, Remijsen & Ladd (2008:181) also report that there are two tone categories involving falling contours. Both of these studies present evidence from tone sandhi suggesting that contour tones in Dinka are composed of level targets. As this phenomenon contradicts the widely-held generalization in (1), it is worthwhile to examine it in detail. The phenomenon will be investigated through a production study in which the amount of sonorous duration is manipulated experimentally (cf. Caspers & van Heuven 1993). If Andersen’s hypothesis is correct and the contrast is indeed one of tonal alignment, then the contrasting patterns should diverge primarily in the timing of the targets with the segmental sequence, and this correlate should be stable across manipulations of time pressure.

A corroboration of Andersen’s hypothesis would then raise two important issues. First, as noted above, the phonetics of contour tones are constrained by limitations in speech production and
speech perception, which ultimately affect their distribution (Zhang 2001, 2002). If Dinka does show a contrast between two contour tones, it will be worthwhile to find out how this is possible. Dinka is particularly suitable to explore the impact of these phonetic limitations, because sonorous duration can be controlled effectively through contrastive vowel length.

The second issue on which the corroboration of Andersen’s hypothesis would have a bearing is phonological representation. In the autosegmental approach, where tonal phenomena are represented by associating tone targets with metrical units, there are two general strategies to represent contrastive alignment: one is to increase the level of detail of representation on the tonal tier. The other is to associate the tone targets with a smaller unit within the prosodic hierarchy. I will argue for a feature-based representation on the tonal tier.

The remainder of this introduction provides background on Dinka and its sound system. Section 2 presents a description of the alignment contrast between falling contours in Dinka, and provides preliminary support for the hypothesis that there are indeed two falling patterns in the surface phonology. This is followed by a production study (Section 3). In the Discussion I will evaluate hypothesis (1) in the face of the evidence from Dinka (Section 4.1), and then consider the above-mentioned issues, namely phonetic limitations (Section 4.2), and phonological representation (Section 4.3). The paper is concluded in Section 5.

1.2 Background on Dinka and its suprasegmental phonology

Dinka is a Western Nilotic language within the Nilo-Saharan language family. It has approximately 2 million speakers, primarily in South Sudan. The dialects of Dinka can be divided into four geographic clusters: Padang, Rek, Agar, and Bor (Roettger & Roettger 1989). This paper deals primarily with the Bor cluster, which includes – from South to North – the dialects known as Bor (in a narrow sense), Twic, Nyarweng, and Hol. The first three of these are under investigation in the production study in Section 3. With respect to tone, Twic and Nyarweng do not present any differences between themselves, but the two of them differ from Bor (narrow sense) in a way that is relevant to this study. I will therefore refer to Bor (narrow sense) as Bor South (BS), and to Twic and Nyarweng together as Bor North (BN).

A salient characteristic of Dinka phonology is its rich system of suprasegmental contrasts: tone, vowel length and voice quality are all three distinctive, independently of one another (Andersen
1987, Remijsen & Manyang 2009). The minimal set in (2) illustrates the vowel length contrast. Vowel length distinguishes both unrelated lexical items – e.g. the short root \{lel\} ‘isolate’ in (2a,b) vs. the long root \{leel\} ‘provoke’ in (2c) – and also related forms within paradigms – e.g. 2nd vs. 3rd singular of \{lel\} ‘isolate’ in (2a) vs. (2b), respectively.

(2) a. ràaan ā-lèl
   person DC\S-isolate\2S
   ‘You are isolating a person.’
   b. ràaan ā-lèel
   person DC\S-isolate\3S
   ‘He is isolating a person.’
   c. ràaan ā-lèel
   person DC\S-provoke\3S
   ‘He is provoking a person.’
   d. ṭǒoor āa-lèl
   men DC\P-isolate\2S
   ‘You are isolating men.’
   e. ṭǒoor āa-lèel
   men DC\P-isolate\3S
   ‘He is isolating men.’
   f. ṭǒoor āa-lèel
   men DC\P-provoke\3S
   ‘He is provoking men.’

As seen from (2a-c), the combination of lexical and morphological length combines into a three-level vowel length contrast in content morphemes. The hypothesis of three-level vowel length was first proposed for Dinka by Andersen (1987, 1993), and has been corroborated on the basis of phonetic evidence (Remijsen & Gilley 2008). The three levels of vowel length – short, long and overlong – can be represented through one, two, and three moras, respectively, associated with the syllable nucleus (Flack 2007, Remijsen & Gilley 2008:320). This is illustrated schematically in (3).

(3) CVC
    \[\mu\]
    CVVC
    \[\mu \mu\]
    \[\mu \mu \mu\]

All verb roots consist of a single closed syllable, a pattern which predominates throughout the lexicon (Andersen 1987, 1993). Short verb roots like \{lel\} ‘isolate’ have a phonologically short stem vowel (V) in some inflections, and a phonologically long stem vowel (VV) elsewhere – e.g. /lèl/ ‘isolate\2S’ vs. /lèel/ ‘isolate\3S’. Long roots such as \{leel\} ‘provoke’ have a phonologically long stem vowel in some inflections, and a phonologically overlong stem vowel (VVV) elsewhere – e.g. /lèel/ ‘provoke\2S’ vs. /lèeel/ ‘provoke\3S’. In function morphemes, vowel length is distinctive in a binary fashion, i.e., short vs. long. This contrast is illustrated in
by the declarative prefix, which is /ā-/ when the preceding topic noun is grammatically singular (1a-c), and /āā-/ when the preceding noun is grammatically plural (1d-f).

Like vowel length, tone is distinctive both between lexical items and within paradigms. Depending on the dialect, the inventory of underlying tone categories or tonemes consists of three or four patterns, associated at the level of the syllable. The Bor dialects all have four tonemes: Low, High, Mid, and Fall. Figure 1 illustrates the realizations of these patterns.²

![Figure 1](image.png)

**Figure 1.** Averaged interpolated f0 traces on a normalized time axis, showing the realization of the four tonemes – Low, High, Fall, and Mid – in /nòoon/ ‘grass\S’, /lòooom/ ‘rib\S’, /nàaap/ ‘sycamore\S’, and /làaac/ ‘urine\U’, respectively. Each trace is averaged across realizations by three speakers of Bor North. Panel 1A shows citation forms. Panel 1B shows the same target words in a frame sentence, between High tonemes. The vertical line in 1A and the one on the left in 1B both mark the start of the vowel of the target word. The second vertical line in 1B marks the end of the target word.

There are two phonological categories of voice quality: modal vs. breathy. Like tone and vowel length, voice quality distinguishes unrelated lexical items – e.g., /ròoor/ ‘forest’ vs. /ròoor/ ‘men’ – and also forms within morphological paradigms, as in verb forms marked for spatial deixis e.g. /lèeel/ ‘provoke\3S.FUGAL’ vs. /lèeel/ ‘provoke\3S.PETAL’. Beyond this phonological contrast between modal and breathy voice, I have noted two phonetic interactions that involve voice quality. First, overlong Low-toned vowels often sound creaky when they are preceded by another Low toneme and followed by an utterance boundary – i.e., when realized low in the speaker’s f0 range. Second, specifically in Bor South, the vowels /e,o/ are somewhat centralized when they are phonologically breathy.³
As noted above, in many languages contour tones are restricted to long vowels (Zhang 2001, 2002). In several Dinka dialects, however, contour tones are found on syllables with a short, long and overlong vowel alike. The Bor North dialects form part of this group, as do Agar (Andersen 1987) and Luanyjang (Remijsen & Ladd 2008). In Bor North, the Fall is found productively on syllables with a short vowel in one grammatical context only: the present-tense passive inflection of transitive verbs, illustrated in (4a).

(4)  
\[ a. \text{Dèn̥ ā-t̷n̥̣ ðàā n̥̣ l̥̣l} \quad \text{(Bor North)} \]  
\[ b. \text{Dèn̥ ā-t̷n̥̣ ðàā n̥̣ l̥̣l} \quad \text{(Bor South)} \]  
\[ \text{Deng DC\$see person provoke\$PASS} \]  
\[ \text{‘Deng sees the person who is being provoked.’} \]

In Bor South, in contrast, there is an interaction between the inventory of tone categories and syllable weight. Here the Fall toneme is found on syllables with a long or overlong vowel, but not on syllables with a short vowel. For example, as seen from (4b), short verb roots in Bor South have the Mid toneme on the present-tense passive, rather than the Fall.

2. Falling contours in Dinka

2.1 Description

This section provides a description of the contrast between the two falling contours, based primarily on data from Bor South. The tone categories or tonemes involved are the Low and the Fall. In (5), these tonemes appear in sentence-final position, preceded by a Low toneme.

(5)  
\[ a. \text{dèn̥ ā-t̷n̥̣ ðàā n̥̣ l̥̣l} \quad \text{b. dèn̥ ā-t̷n̥̣ ðàā n̥̣ l̥̣l} \]  
\[ \text{Deng DC\$see person isolate\$3S} \quad \text{Deng DC\$see person provoke\$PASS} \]  
\[ \text{‘Deng sees the person whom he is isolating.’} \quad \text{‘Deng sees the person who is being provoked.’} \]

Figure 2A shows descriptive statistics on the realization of this contrast in this particular context. The graph shows the f0 traces of the last two syllables in (5a,b), averaged over realizations by four speakers. The trace for the Fall toneme (interrupted line) reaches a maximum well into the vowel before falling steeply. The Low toneme (full line) slopes down at the bottom of the speakers’ range.
The realization of the Low toneme is markedly different when it is not preceded by another instance of the Low toneme. In (6), the two verb stems familiar from (5) form the head of a clause, and as such they take the Mid-toned declarative prefix /ä-/.

(6)  a. ràaan  ā-lēel  
    person  DC\S\-isolate\3S  
    ‘He is isolating a person.’

b. ràaan  ā-lēel  
    person  DC\S\-provoke\\PASS  
    ‘A person is being provoked.’

Averaged traces are presented in Figure 2B. A comparison of the interrupted traces in Figures 2A vs. 2B shows that the realization of the Fall toneme does not differ noticeably as a function of the preceding tonal context. But the realization of the Low toneme (full lines) is markedly different in the two contexts. Preceded by another Low toneme, the Low toneme is realized through a low and level f0 trace (Figure 2A). Hereafter this allophonic variant will be referred to as Low\textsuperscript{Level}. Preceded by the Mid toneme, the realization of the Low toneme involves an f0 peak early in the vowel (Figure 2B), followed by a falling contour. Hereafter this allophonic variant will be referred to as Low\textsuperscript{Fall}. The difference in realization between the Low\textsuperscript{Fall} and the Fall tonemes appears to be primarily a matter of the alignment of the f0 peak that defines the starting point of a fall in f0 within the rhyme. This peak is located near the beginning of the vowel in the case of the Low\textsuperscript{Fall}. In the case of the Fall, this peak is located well into the vowel. I do not perceive any difference in phonetic voice quality between the Low\textsuperscript{Fall} and the Fall.
These observations on the realization of the Low<sub>Fall</sub> and the Fall in Bor South are in line with Andersen’s (1987:20) study on the Agar dialect. He summarizes his syllable-level pitch impressions as “fall […] followed by level pitch” in the case of what I have referred to as the Low<sub>Fall</sub>, and “level pitch […] followed by fall” in the case of the Fall. Similarly, Andersen does not report a difference in perceived voice quality between the two falling contours.

The distributions of Low<sub>Level</sub> and Low<sub>Fall</sub> do not overlap. The Low<sub>Level</sub> realization is found following another instance of the Low toneme. The Low<sub>Fall</sub> is found elsewhere: following the Mid toneme (Figure 2B), the High toneme (Figure 1B), and also in the citation form (Figure 1A). The fact that the Low<sub>Fall</sub> is found in the citation form and following the Mid toneme is insightful, because, in these contexts, the high target at the beginning of the vowel cannot be attributed to a high end target of a preceding syllable. That is, the Mid toneme does not involve a high end target in the Bor dialects in any other context (cf. Figures 1A,B).

In most of the dialects of Dinka that have been investigated so far, the Fall is also found on short vowels. If a dialect allows this configuration, then one can expect to find minimal contrasts of Low<sub>Fall</sub> vs. Fall on syllables with a short vowel, as in the example from North Bor in (7). Comparable contrasts have also been reported for Agar Dinka (Andersen 1987:18) and for Luanyjang Dinka (Remijsen & Manyang 2009:120).

(7) a. ràaan ã-lèl
   person DC\S-isolate\2S
   ‘You are isolating a person.’
   b. ràaan ã-lèl
   person DC\S-isolate\PASS
   ‘A person is being isolated.’

In conclusion, the Low<sub>Fall</sub> and Fall tone patterns appear to differ primarily in the alignment within the syllable of a peak that defines the beginning of a falling contour (cf. Figure 2B). This observation is in line with the description in Andersen (1987). However, such a state of affairs is at odds with the hypothesis (1) that the timing of f0 changes within the syllable is not contrastive. This hypothesis predicts that, if the Low toneme is realized as a falling contour in certain contexts, it should neutralize with the underlying Fall (cf. Silverman 1997). The following two subsections present evidence as to whether or not the Low<sub>Fall</sub> and Fall neutralize, from a perception experiment (Section 2.2), and from a tone sandhi process (Section 2.3).
2.2 Perception experiment

To the best of my knowledge, native speakers of Dinka have no difficulty in disambiguating forms in which the lexical and/or grammatical meaning rests solely on the contrast between the Low^Fall and the Fall. The experiment reported here constitutes a small-scale attempt to verify this. The subjects are presented with a stimulus whose interpretation hinges crucially on whether the verb has the Low^Fall or the Fall. After hearing the stimulus utterance, the subjects select between the English translations appropriate for the two interpretations (forced-choice design). For example, one stimulus is a recording of the sentence in (6a), and the subjects respond by selecting either the translation in (6a) or the one in (6b). There are 10 stimuli for tone, made up of five minimal sets for the contrast between Low^Fall and Fall, including the sets in (6) and (7). Three minimal sets have a short vowel in the stem syllable, and two have a long vowel. The stimuli come from three different speakers. There is also a control set involving a contrast in vowel height^4, as an independent criterion on the basis of which to exclude responses from participants who would have difficulties with the task.

The design was implemented in Praat (Boersma & Weenink 2010) using the Multiple Forced Choice tool for designing and running listening experiments. To gauge the level of consistency in the responses, each stimulus recurs 5 times in the experiment. The stimuli were presented in five randomized blocks consisting of 12 stimuli each, with each stimulus occurring only once in each block. The stimuli were presented over loudspeakers, and the responses were recorded by clicking the selected English translation. Because some of the subjects were not accustomed to using a computer, the subjects pointed out their selection, and I recorded their choice by clicking the sentence they had pointed at. Six native speakers of Dinka participated as subjects. Three of them are speakers of Bor North (2 Twic, 1 Nyarweng); the other three speak dialects from the Rek cluster (2 Twic^5, 1 Malual). All of the dialect varieties represented among the participants allow for the Fall toneme on short vowels.

None of the participants had difficulties with the task: each subject invariably chose the correct translation for the control-set items. These are not considered further. The responses on tone stimuli number 300 in total (6 subjects * 10 stimuli * 5 repetitions). The correct classification score on these stimuli is 92% (276 correct out of 300). Separated by the dialect region of the
participants, the correct classification result is higher for the Rek subjects (95.3% correct) than for the Bor North subjects (88.6% correct). Across all subjects, correct classification was better on stimuli with a short vowel (93.9% correct) than on the stimuli with a long vowel (89.2%). It is clear that the participants were able to infer the lexical and morphological meaning in minimal pairs for tone (Low<sub>Fall</sub> vs. Fall) well above chance level. In addition, it is noteworthy that the Rek speakers were also successful at the task, even though the stimuli were drawn from a Bor dialect. This confirms that the contrast between falling contours is representative beyond the Bor dialects (cf. Andersen 1987, 1993; Remijsen & Manyang 2009).

2.3 Tone sandhi

The hypothesis in (1) predicts that when the Dinka Low toneme is realized as a falling contour, it should no longer be phonologically distinct from the Fall toneme (cf. Silverman 1997). Concretely, this hypothesis entails that underlying stem forms like /lèel/ ‘isolate\3S’ and /lèel/ ‘prove\<PASS>’ are indistinguishable as falls in the surface phonology, among others in forms involving the Mid-toned declarative prefix. That is, underlying /lèel/ and /lèel/ would both yield surface-phonological /ä-lèel/. The Bor South dialect presents a tone sandhi process that makes it possible to evaluate this prediction. The argument depends on the assumption that phonological processes that apply within words take effect before phonological processes between words (Kiparsky 1985). In the Bor South dialect, the Fall toneme is replaced by the High toneme whenever it is not immediately followed by an utterance boundary. Similar sandhi processes are reported for other dialects (Andersen 1987, Remijsen & Ladd 2008). This process is formulated in autosegmental terms in (8), where the Fall is represented as High-Low. The second component of a High-Low sequence associated with a single syllable is deleted if the tone-bearing unit is followed by another word within the same utterance.

\[
(8) \quad \sigma \quad \# \quad \sigma \quad \text{Fall Truncation}
\]

Consider the application of this rule on the verb forms familiar from (6). In (9), these verb forms are followed by an adverb, so that the contextual condition for the rule in (8) is met. If the prefix /ä-/ has the effect of turning a Low toneme into a Fall toneme in the context of the
word-internal phonology, then the verb stem in (9a) has the Fall toneme at the post-lexical stage of derivation, the point where Fall Truncation applies.

(9) a. ṭ̣ạṇ-ạḷeḷ ẹ̀-tēṇe
    person DC\S-isolate\3S PREP:here
    ‘He is isolating a person here.’

b. ṭ̣ạṇ-ạḷeḷ ẹ̀-tēṇe
    person DC\S-provoke\PASS PREP:here
    ‘A person is being provoked here.’

The Low\Fall pattern is conditioned not only by a word-internal non-Low target, as in (6a), but also by a non-Low target separated by a word boundary, as in Figure 1B. Given so, any conditioning process producing the Low\Fall would also have to apply post-lexically. By implication, the Low\Fall may also be conditioned after Fall Truncation (8) has applied. I assume that any post-lexical conditioning of the Low\Fall would be vacuous when the conditioning context is already met word-internally.

Figure 3 presents averaged traces based on data from four speakers. In the case of the underlying Fall (9b), the graph shows f0 rising throughout the rhyme, indicating that the underlying Fall (HL) has turned into High, in line with Fall Truncation. This contrasts with the falling f0 contour in utterance-final position (Figure 2B, interrupted line).

**Figure 3.** Averaged f0 traces on a normalized time axis, illustrating the realization of syllables underlyingly specified for Low and Fall tonemes, averaged across realizations by four speakers of Bor South. The traces represent the last word in the sentences in (9a) and (9b), respectively.

In contrast, the stem with underlying specification for the Low toneme (Figure 3, full line) shows no evidence of Fall Truncation. Its averaged trace is very similar to the one in Figure 2B: there is a falling contour over the stem syllable of the verb, and the beginning point of this contour is close to the beginning of the vowel. On the assumption that the phonological
processes are ordered as outlined here, Fall Truncation does not affect the Low\textsuperscript{Fall}, suggesting that Low\textsuperscript{Fall} is distinct from the Fall in phonological terms.

2.4 Summary

The findings from the perception experiment and from the tone sandhi process are in line with the observations reported in Section 2.1 and in Andersen (1987), i.e., that Dinka has two falling contour tones in its surface phonology. On this basis, I present the hypothesis in (10):

\begin{equation}
(10) \quad \text{Tonal alignment distinguishes the Low}\textsuperscript{Fall} (early-aligned in syllable) from the Fall (late-aligned in syllable) in the surface phonology of Dinka.}
\end{equation}

The hypothesis in (10) is not compatible with the hypothesis in (1). In the following section I will report on a production study in which the phonetic realizations of the Low\textsuperscript{Fall} and the Fall are studied in detail. What is at issue is whether the contrast under investigation is primarily one of tonal alignment. The hypothesis in (1) would be corroborated if the two falling contours diverge primarily in terms of another correlate, such as f0 height (Yip 2002:29). If, in contrast, the hypothesis in (10) is corroborated, then it will be worthwhile to find out how a contrast in tonal alignment can be implemented phonetically, and to consider how such a contrast can best be represented.

3. Production study

The goal of this production study is to gain insight into the phonetic realization of the Low\textsuperscript{Fall} and Fall in Dinka. The design hinges on the manipulation of time pressure, which enables the researcher, “to separate the properties of pitch movements into categories of greater or lesser importance” (Caspers & van Heuven 1993:162). Factors that are often manipulated in this context include rate of speech (Caspers & van Heuven 1993; Xu 1998, 2001), vowel length and the composition of the syllable (Caspers & van Heuven 1993; Xu 1998; Schepman, Lickley & Ladd 2006), and word length (Myers 2003, Yu 2008). In the context of the current study, any parameters that saliently and consistently distinguish Low\textsuperscript{Fall} and Fall from one another across manipulations of time pressure have a bearing on the phonological nature of the two contours, and by extension, on the evaluation of the competing hypotheses (1,10).
3.1 Methods
3.1.1 Materials

The phonological patterns of tone and vowel length in the target words are illustrated in (11). All of the target words involve the Mid-toned declarative prefix /ā(a)/-, which conditions the Low$^{\text{Fall}}$ realization of the Low toneme, thereby setting up a minimal contrast between Low$^{\text{Fall}}$ and Fall. Time pressure in the two syllables nearest to the falling f0 change was manipulated independently, by controlling vowel length in the prefix syllable and in the stem syllable. It is relatively straightforward to manipulate time pressure in this way in Dinka, because vowel length is determined by lexical and morphological meaning (cf. Section 1.2).

(11) ā-lèl ā-lèel ā-lèeel
DC\S-isolate\2S DC\S-isolate\3S DC\S-provoke\3S
āa-lèl āa-lèel āa-lèeel
DC\P-isolate\2S DC\P-isolate\3S DC\P-provoke\3S
ā-lèl* ā-lèel
DC\S-isolate\PASS DC\S-provoke\PASS
āa-lèl* āa-lèel
DC\P-isolate\PASS DC\P-provoke\PASS

The asterisked forms in (11), i.e., the combination of the Fall with a short stem vowel, are only found in Bor North, and as a result they were only recorded from speakers of that dialect. In Bor South, the same inflection is marked instead by the Mid toneme, as in (4b). These Mid-toned forms were recorded from the speakers of Bor South, but they are not included in the analysis of the falling contours reported in Section 3.2. Not included at all in the design is the combination of the Fall toneme with an overlong stem vowel. This combination is part of the Dinka phonological system, but it was left out for practical reasons. This is not a problem, because time pressure is smallest in forms with an overlong vowel.

In (11), the design of the materials is illustrated on the segmental sequence /lel-leel-leeel/. The same phonological patterns were also collected for three other sets: /maŋ-maŋ-maan/, /ŋop-ŋoop-ŋooott/, /wel-weel-weeel/. In each case, the onset is a sonorant, and the segments are matched for phonetic type within each set. The target verb forms are embedded in final position in a two-word sentence, preceded by a noun that is monosyllabic and Low-toned, as in (2,6,7).
3.1.2 Speakers
Thirteen speakers of Dinka (11 men, 2 women) were recorded. There are seven speakers of Bor North (BN – 6 male, 1 female) and six of Bor South (BS – 5 male, 1 female). All speakers were born in Dinka villages, but spent one to two decades in neighboring countries as refugees. Through a questionnaire, I made sure that all of the participating speakers had spent their time away living with people of the same dialect. The speakers also speak English, plus one or more other languages, such as Arabic, Swahili, and Acholi.

3.1.3 Procedure
The dataset is built up around eight lexical verb roots, listed in Table I. That is, each of the four segmentally controlled sets involves two roots – e.g. {leel} ‘isolate’ and {leel} ‘provoke’ in the case of the /leel-leel-leel/ set. Inflections of the same verb root were elicited together. The eight root-based blocks were ordered so as to avoid lexical confusion between segmentally similar roots. Table I presents the list of stem forms that were elicited in forward order. This order was reversed for half of the speakers, to counterbalance for order-of-presentation effects. Within each block, the order of elicitation was fixed: 1st/3rd singular came first, then 2nd singular, then passive. For each stem inflection, forms preceded by a singular topic noun were elicited before forms with a plural topic noun. Two or three realizations were recorded of each form. The data were elicited using English by the author, initially with the help of a Dinka assistant, John Penn de Ngong, to help ensure that the intended form was recorded. The recordings were made in Juba, South Sudan, using a directional headset-mounted microphone and a solid-state recorder.

Table I. The inflections that were elicited for each verb root, in forward order.

<table>
<thead>
<tr>
<th>Verb root</th>
<th>Inflections (each elicited with prefixes /a/- ‘DC\S’ and /aa/- ‘DC\P’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>{leel} ‘provoke’</td>
<td>3S: -léeel</td>
</tr>
<tr>
<td>{wel} ‘turn over’</td>
<td>3S: -wéeel</td>
</tr>
<tr>
<td>{maan} ‘hate’</td>
<td>1S: -mään</td>
</tr>
<tr>
<td>{öop} ‘take a gulp’</td>
<td>3S: -ööop</td>
</tr>
<tr>
<td>{maŋ} ‘slap’</td>
<td>3S: -mâaŋ</td>
</tr>
<tr>
<td>{ōoot} ‘spit’</td>
<td>3S: -ōoot</td>
</tr>
</tbody>
</table>
3.1.4 Data processing and analysis

The recordings were processed and analyzed using Praat (Boersma & Weenink 2010). The target word within each sentence was segmented based on information in the waveform and in the spectrogram. In three of the four sets, the onset is a nasal or /l/. These consonants lend themselves well to the study of tonal phenomena, because they carry fundamental frequency without perturbing it upward or downward. However, it is not uncommon for the f0 trace to show a spike at the boundary between such a consonant and a vowel. These spikes were corrected using the trimming algorithm described in Xu (1999). The threshold value above which spikes are trimmed was set at 1 Hz. The effect of this correction can be seen in Figure 4. In both of the examples, the spikes in the raw trace (grey) would have shifted the timing measurement of the f0 peak towards the segment boundary. The corrected trace is overlaid in black. Raw and trimmed traces were compared visually for each target word token, to examine the output of the trimming algorithm and to correct any tracking errors. Figure 4 also illustrates that, for the Low\textsuperscript{Fall} and the Fall alike, there is a salient f0 peak in the context of a preceding Mid toneme. As a result, the height and timing properties of the peak are easy to measure.

Figure 4. Raw (grey) and trimmed (black) f0 traces for the target words in two utterances by a female speaker of Bor North (Twic). Left panel: /aa-\textipa{l}el/; right panel: /\textipa{a}-\textipa{l}el/.

The combination of 13 speakers each producing 10 forms for each of 4 segmental sets yields an expected total of 520 cells. However, 15 of these are empty. These gaps are either due to items
being accidentally skipped during elicitation, or because I failed to notice during elicitation that a speaker uttered a sentence other than the intended one.

The following measures were extracted for each token: (a) the duration of the stem vowel; (b) the duration of the prefix vowel; (c) the difference in time between the f0 peak within the stem syllable and the beginning of the stem vowel, the end of the stem vowel, and the beginning of the prefix vowel; (d) the f0 value at this maximum; and (e) the excursion size of the f0 fall. The values for these measurements were averaged across any repetitions of the same sentence by the same speaker, before running the statistics.

All of these measurements will be investigated through descriptive statistics, and those that may constitute correlates of the contrast between Low\textsuperscript{Fall} and Fall will be examined further through linear mixed effects modeling and likelihood ratio tests, as implemented in R (R Development Core Team 2010). The design consists of the random factors Speaker (13) and Set (4), and the fixed factors Tone (Low\textsuperscript{Fall} / Fall), Stem length (V / VV / VVV), Prefix length (V / VV), Dialect (BS / BN), and Coda type (Sonorant / Plosive). Stem forms that have the Mid toneme (i.e., passive forms of short verbs in Bor South) are not included in these analyses. Test values and probabilities are reported for all main effects, and also for any interactions that significantly improve the model in a likelihood ratio test (Bates 2005).

3.2 Results
3.2.1 Vowel duration

The measurements for vowel duration – in particular regarding the short and long levels of vowel length in both prefix and stem syllables – are ancillary to the main focus of the production study, in that they demonstrate the extent to which sonorous duration has been varied experimentally. These measurements are also of interest in themselves, because three-level vowel length is highly uncommon among the world’s languages (Odden 2011).

The descriptive statistics for vowel length in the stem syllable are presented in Table II. Long vowels are 59 percent (40 ms) longer than short vowels, and overlong vowels are about 77 percent (80 ms) longer than long vowels. A separate set of statistics is provided for the forms that have the Low\textsuperscript{Fall} on the stem syllable, because these represent a dataset for three-level
vowel length in Dinka in which tone is controlled. In the prefix, where the vowel length distinction is binary rather than ternary, the average durations were 76.2 ms for the short vowel /a/- and 137.9 ms for long /aa/-, i.e., a difference of about 80 percent (60 ms).

**Table II.** Means and standard deviations (between brackets) for the duration of the stem vowel, in ms. These statistics are provided both for the whole dataset, and also for a subset consisting of those forms that have the Low^{Fall} on the stem syllable.

<table>
<thead>
<tr>
<th>Stem length</th>
<th>Low^{Fall} and Fall</th>
<th>Low^{Fall} subset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short (V)</td>
<td>70.7 (12.1)</td>
<td>68.9 (11.7)</td>
</tr>
<tr>
<td>Long (VV)</td>
<td>112.8 (16.2)</td>
<td>109.6 (15.7)</td>
</tr>
<tr>
<td>Overlong (VVV)</td>
<td>194.4 (30.1)</td>
<td>194.4 (30.1)</td>
</tr>
</tbody>
</table>

Figure 5A shows the descriptive statistics for the whole dataset as a graph. Comparable results for the Luanyjang dialect, from Remijsen & Gilley (2008), are presented alongside (Figure 5B). In both datasets, the target words are in utterance-final position. The pattern is similar in the two dialects. The separation between levels of vowel length is clearer in the Bor dialects, where there is no overlap between the distributions at the level of one standard deviation. Factors that are likely to give rise to variation within levels of vowel length include between-speaker variation in rate of speech and intrinsic vowel duration.

**Figure 5.** Means and standard deviations for vowel duration (ms) in the stem syllable, as a function of Stem length (V, VV, VVV). Panel A: this study (13 speakers of Bor Dinka). Panel B: comparable data from Remijsen & Gilley (2008) on Luanyjang Dinka.
Table III reports the results of the linear mixed effects model with vowel duration in the stem syllable as the dependent. These results confirm that there is a substantial difference in the duration of the stem vowel as a function of Stem length. The comparisons between levels of Stem length are all significant, characterized by double-digit $t$ values. There is no significant effect for Dialect. This indicates that the vowel length distinction is realized similarly in Bor North and Bor South. Prefix length and Coda type also do not register a significant effect. There is, however, a significant effect of Tone. Vowel durations are 5-6 ms shorter when the stem is specified for the Low$^\text{Fall}$ than when it carries the Fall. The mean values for Low$^\text{Fall}$ and Fall are 68.9 and 74.1 ms, respectively, when then the stem vowel is short, and 109.6 vs. 115.8 when the vowel is long. In summary, the three-level vowel length distinction is saliently realized in Dinka Bor. Specifically for the purposes of the current study, it is clear that time pressure was manipulated effectively by controlling Stem length: differences between adjacent levels of vowel length in the prefix and in the stem range from 60 to 80 percent relative to the lower of the two levels. Vowel duration is also significantly affected by Tone, although this difference is proportionally small: the vowel is between 5 and 7.5 percent longer if the syllable has the Fall toneme than if it has the Low$^\text{Fall}$.

**Table III.** Summary of the results of a linear mixed effects model with the duration of the vowel of the stem syllable as the dependent.

<table>
<thead>
<tr>
<th>Factor and levels</th>
<th>$t$ value</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stem length</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– V vs. VV</td>
<td>24.7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>– V vs. VVV</td>
<td>63.0</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>– VV vs. VVV</td>
<td>41.9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Prefix length</td>
<td>– V- vs. VV-</td>
<td>-1.5</td>
</tr>
<tr>
<td>Coda type</td>
<td>– sonorant vs. plosive</td>
<td>-1.06</td>
</tr>
<tr>
<td>Tone</td>
<td>– Low$^\text{Fall}$ vs. Fall</td>
<td>3.31</td>
</tr>
<tr>
<td>Dialect</td>
<td>– BS vs. BN</td>
<td>0.1</td>
</tr>
</tbody>
</table>
3.2.2 Alignment of the f0 peak

Figure 6 displays averaged f0 traces over the stem syllable, as a function of Stem length, Tone and Coda type. In this section we will consider the alignment of the f0 peak. After that we will consider the height of this f0 peak (Section 3.2.3) and then the excursion size of the fall that follows the peak (Section 3.2.4).

Figure 6. Averaged f0 traces on a normalized time axis, by Stem length and Tone. Separate panels for sets ending in a sonorant coda (panel A) vs. sets ending in a plosive (panel B).

The graphs in Figure 6 suggest that the peak is aligned close to the beginning of the stem vowel in the case of the Low\textsuperscript{Fall}, and well into the stem vowel in the case of the Fall. To begin with, we need to determine the segmental landmark relative to which the alignment of these peaks can be expressed best. This question can be answered by examining alignment across manipulations of vowel length: the most appropriate landmark is the one relative to which peak alignment varies least. As seen from Table IV, if we use beginning of the stem vowel as a point of reference, then the peak of the Low\textsuperscript{Fall} is aligned on average 9.6 ms later on a long vowel than on a short vowel, and 13 ms later on an overlong vowel than on a short vowel. In contrast, if peak alignment for the Low\textsuperscript{Fall} is expressed relative to the end of the vowel, the variation in peak alignment are much greater: 31.3 ms difference between short and long vowels, and as a much as 111.4 ms difference between short and overlong vowels. Table IV also presents comparable descriptive statistics for the Fall. Peak alignment for the Fall differs by 13.1 ms between short and long vowels when it is expressed relative to the beginning of the vowel. Expressed relative to the end of the vowel, the difference is 28.6 ms, i.e., more than the double. In summary, for the Low\textsuperscript{Fall} and the Fall alike, peak alignment can be expressed more
appropriately relative to the beginning of the stem vowel than to its end, given that alignment varies less across contexts relative to the former landmark than relative to the latter.

**Table IV.** Means for f0 peak alignment (ms) calculated relative to the beginning and the end of the stem vowel, as a function of Tone and Stem length

<table>
<thead>
<tr>
<th>Tone</th>
<th>Stem length</th>
<th>Peak alignment, expressed relative to:</th>
<th>Start of stem vowel</th>
<th>End of stem vowel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low\textsuperscript{Fall}</td>
<td>CVC</td>
<td>-5.6</td>
<td>-74.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CVVC</td>
<td>4.0</td>
<td>-105.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CVVVC</td>
<td>8.6</td>
<td>-185.8</td>
<td></td>
</tr>
<tr>
<td>Fall</td>
<td>CVC</td>
<td>32.6</td>
<td>-41.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CVVC</td>
<td>45.7</td>
<td>-70.1</td>
<td></td>
</tr>
</tbody>
</table>

Particularly in relation to the Low\textsuperscript{Fall}, it is also worthwhile to consider the phonetic alignment of the peak relative to the prefix vowel. Given that the Mid associated with the prefix is one of the contexts in which the Low\textsuperscript{Fall} is found, it could be that peak is aligned in a constant timing relation with this prefix vowel. As seen from Table V, this is not the case: the Low\textsuperscript{Fall} reaches its peak about 59 ms later following a long prefix vowel than after a short prefix vowel. The same difference as a function of prefix length is found for the Fall. In both cases, this distance between the peak and the beginning of the prefix vowel is instead greatly affected by vowel length: the duration of long vowels is 62 ms greater than that of short vowels.

**Table V.** Means for f0 peak alignment (ms) calculated relative to the beginning of the prefix vowel, as a function of Tone and Prefix length.

<table>
<thead>
<tr>
<th>Tone</th>
<th>Prefix length</th>
<th>Peak alignment relative to start of prefix vowel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low\textsuperscript{Fall}</td>
<td>V-</td>
<td>163.5</td>
</tr>
<tr>
<td></td>
<td>VV-</td>
<td>222.3</td>
</tr>
<tr>
<td>Fall</td>
<td>V-</td>
<td>199.3</td>
</tr>
<tr>
<td></td>
<td>VV-</td>
<td>258.0</td>
</tr>
</tbody>
</table>

Given these comparisons, further descriptive and inferential statistics on peak alignment will be expressed relative to the beginning of the vowel of the stem syllable. Across the dataset, the
mean values for peak alignment for the $\text{Low}^{\text{Fall}}$ and the Fall are 2.3 ms and 41.0 ms into the vowel, respectively, in each case with a standard deviations of 15 ms. Means by levels of Dialect, Stem length and Prefix length are presented in Table VI.

**Table VI.** Means for the alignment of the peak that defines the beginning of the falling contour, by Dialect, Tone, Stem length, and Prefix length.

<table>
<thead>
<tr>
<th>Dialect</th>
<th>Tone</th>
<th>Stem length</th>
<th>Pref. length V-</th>
<th>Pref. length VV-</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bor North</td>
<td>$\text{Low}^{\text{Fall}}$</td>
<td>V</td>
<td>-5.6</td>
<td>-12.1</td>
<td>-8.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>VV</td>
<td>-2.3</td>
<td>-1.8</td>
<td>-2.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>VVV</td>
<td>7.9</td>
<td>0.0</td>
<td>3.6</td>
</tr>
<tr>
<td>Fall</td>
<td>V</td>
<td>32.9</td>
<td>32.3</td>
<td>32.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VV</td>
<td>45.0</td>
<td>41.9</td>
<td>43.5</td>
<td></td>
</tr>
<tr>
<td>Bor South</td>
<td>$\text{Low}^{\text{Fall}}$</td>
<td>V</td>
<td>1.2</td>
<td>-4.7</td>
<td>-1.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>VV</td>
<td>12.1</td>
<td>8.8</td>
<td>10.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>VVV</td>
<td>16.1</td>
<td>13.3</td>
<td>14.7</td>
</tr>
<tr>
<td>Fall</td>
<td>VV</td>
<td>48.6</td>
<td>48.9</td>
<td>48.7</td>
<td></td>
</tr>
</tbody>
</table>

Figure 6 and Table VI both show a difference in peak alignment between $\text{Low}^{\text{Fall}}$ and Fall. This difference is displayed more clearly in Figure 7. This graph shows peak alignment as a function of Tone in the subsets where its influence can be compared in the same level of Stem length. The means for peak alignment differ by around 40 ms between $\text{Low}^{\text{Fall}}$ and Fall. The size of this difference is similar in Bor South and in Bor North. Moreover, in the Bor North dialect, where the Fall is also found on short vowels, the difference in peak alignment between the $\text{Low}^{\text{Fall}}$ and the Fall is similar in size between short and long stem vowels.
Figure 7. Means and standard deviations for peak alignment by Dialect, Tone and Stem length (V and VV only).

It is worthwhile to examine the peak alignment of the Low\textsuperscript{Fall} in greater detail – this tone pattern was collected in six different length conditions, through the orthogonal crossing of Stem length and Prefix length – cf. (11). The result is a highly detailed perspective on the influence of time pressure on the alignment of tonal targets.

Figure 8. Means and standard deviations for peak alignment in the realization of Low\textsuperscript{Fall}, as a function of various combinations of Stem length, Prefix length, and Dialect.

As seen from Figure 8A, the peak is aligned progressively later as Stem length increases from short (V) to long (VV), and further to overlong (VVV). In contrast, an increase in Prefix length has the opposite effect: from a short (V-) to a long (VV-) prefix vowel, peak alignment shifts...
earlier. In both cases, there is a shift in alignment towards the region where there is more segmental duration.

The third factor under consideration in Figure 8 is Dialect. As seen from Figure 8C, its effect is similar to that of Prefix length (Figure 8B): the Low\textsuperscript{Fall} reaches its high initial target earlier in Bor North than in Bor South. There is no obvious explanation for this difference, in particular given that a comparable difference can be observed in relation to the Fall (Figure 7, Table VI).

A final factor to consider is Coda type. Three of the four sets have a sonorant coda (N); the fourth set has a plosive coda (P). Figure 9 shows the descriptive statistics. Peak alignment is earlier when the coda is a plosive than when it is a sonorant, across levels of Tone and Stem length. This difference is small in the case of the Low\textsuperscript{Fall}, which is aligned 3.8 ms earlier when the coda consonant is a voiceless plosive rather than a sonorant, but more substantial in the case of the Fall, where peak alignment is on average 7.5 ms earlier when the coda is a voiceless plosive. The difference is particularly great when the stem vowel is short and the syllable carries a Fall. In this structure, the one where time pressure is greatest, the peak is aligned 14.7 ms earlier when the coda is a voiceless plosive than when the coda is a sonorant.

![Figure 9](image)

**Figure 9.** Means and standard deviations for peak alignment by Stem length (separate panels for short and long vowels), Tone and Coda Type.

The direction of the difference in the peak alignment as a function Coda type can be explained in the same way as the differences observed for Stem length and Prefix length. When the coda is a plosive, the sonorous duration on which the tone pattern can be realized is smaller than when the coda is a sonorant. This reduction in sonorous duration shifts the peak earlier, away
from the area where the segmental space is reduced. And given that a voiceless plosive gives rise to increased time pressure at the end of the target syllable, it is natural for the later-aligned Fall to be affected more than the earlier-aligned LowFall.

In summary, Tone conditions a substantial difference in peak alignment, of about 40 ms on average, and the distributions for LowFall and Fall do not overlap at the level of one standard deviation. All of the other factors (Stem length, Prefix length, Dialect, Coda type) give rise to smaller differences, of 5 to 10 ms per factor, and here there is substantial overlap between the distributions of factor levels.

Table VII reports the results of the mixed-effects model with dependent peak alignment (expressed relative to the beginning of the stem vowel). Tone, Stem length, Dialect, and Prefix length all yield a significant effect, in that order of importance. Coda type does not condition a significant main effect, but its interaction with Tone is significant. This indicates that the significant influence of Coda type on peak alignment is limited to the Fall (cf. Figure 9). Tone has the biggest effect on peak alignment, with a t value of 31.3. The significant effects conditioned by other factors are much smaller in size, with t values under 10.

Table VII. The results of a linear mixed effects model with dependent peak alignment.

<table>
<thead>
<tr>
<th>Factor and levels</th>
<th>t value</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stem length</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– V vs. VV</td>
<td>7.5</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>– V vs. VVV</td>
<td>9.5</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>– VV vs. VVV</td>
<td>3.2</td>
<td>0.002</td>
</tr>
<tr>
<td>Prefix length</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– V- vs. VV-</td>
<td>-3.1</td>
<td>0.002</td>
</tr>
<tr>
<td>Coda type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– sonorant vs. plosive</td>
<td>-1.3</td>
<td>0.181</td>
</tr>
<tr>
<td>Tone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– LowFall vs. Fall</td>
<td>31.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Dialect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– BS vs. BN</td>
<td>-3.7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Coda type * Tone</td>
<td>-2.9</td>
<td>0.003</td>
</tr>
</tbody>
</table>
3.2.3 Peak height

The influence of declination – i.e., the overall downward slope of f0 throughout the utterance – was controlled in the dataset by keeping constant both the number of syllables in the sentences and also the tonal context. That is, the target – a disyllabic verb form – is in sentence-final position, preceded by a Low-toned monosyllabic noun. As a result, any difference in the height of the f0 peak at the beginning of the falling contour can be attributed to the factors under investigation. The traces in Figure 6 do not reveal a salient difference between Low\(^{\text{Fall}}\) and Fall in the f0 value of the peak. This is confirmed by the mean values for peak height, which are only 2.1 Hz apart: 150.7 Hz for the Low\(^{\text{Fall}}\) vs. 148.6 Hz for the Fall.

The results of the inferential test with peak height as the dependent are presented in Table VIII. Tone (Low\(^{\text{Fall}}\) vs. Fall) does not yield a significant effect. The only factor that does reach significance is Prefix length. The peak is slightly higher when the prefix vowel is short than when the prefix vowel is long. The mean values are 151.4 Hz and 148.7 Hz, respectively, i.e., a difference of 2.7 Hz. This difference can be attributed to declination: when the vowel of the prefix syllable is long, the stem syllable appears slightly further along the declination slope, thereby yielding a lower peak value. The difference in peak height between Low\(^{\text{Fall}}\) and Fall, while not significant, can be explained in the same way: f0 has declined slightly further in the case of the later-aligned Fall (mean peak height: 148.6 Hz) than in the earlier-aligned Low\(^{\text{Fall}}\) (mean peak height 150.7 Hz).

**Table VIII.** Summary of the results of a linear mixed effects model with dependent peak height.

<table>
<thead>
<tr>
<th>Factor and levels</th>
<th>t value</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stem length</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– V vs. VV</td>
<td>-0.2</td>
<td>0.820</td>
</tr>
<tr>
<td>– V vs. VVV</td>
<td>-0.1</td>
<td>0.999</td>
</tr>
<tr>
<td>– VV vs. VVV</td>
<td>0.2</td>
<td>0.853</td>
</tr>
<tr>
<td>Prefix length</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– V- vs. VV-</td>
<td>-3.6</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Coda type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– sonorant vs. plosive</td>
<td>1.62</td>
<td>0.106</td>
</tr>
<tr>
<td>Tone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– Low(^{\text{Fall}}) vs. Fall</td>
<td>-0.5</td>
<td>0.635</td>
</tr>
<tr>
<td>Dialect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– BS vs. BN</td>
<td>-0.2</td>
<td>0.801</td>
</tr>
</tbody>
</table>
3.2.4 Excursion size of the f0 fall

In the previous section, it became clear that the height of the f0 peak that defines the beginning of the falling contour varies little as a function of the factors under investigation. Given so, variability in the excursion size of the f0 fall is determined primarily by its end value. A comparison between the panels in Figure 6 above suggests that the size of the f0 fall is reduced when the coda is a voiceless plosive and therefore does not carry f0. In addition, Figure 6A suggests that, specifically in relation to stem forms with a sonorant coda, a greater proportion of the f0 fall takes place in the coda consonant a) when Stem length is reduced – from VVV over VV to V – and b) when the Tone involves later alignment – i.e., from Low,\textsuperscript{Fall} to Fall.

These differences can be observed more clearly in Figure 10, which shows descriptive statistics for the excursion size of the fall. Figure 10A shows the results for syllables with a sonorant coda. Here there is little difference in the size of the f0 fall: the mean excursion sizes for the f0 fall as a function of Stem length and Tone are all close to 40 Hz. Figure 10B displays the size of the fall, again in syllables with a sonorant coda, but now only up to the end of the vowel.

These statistics show that the size of the f0 fall over the vowel becomes smaller as Stem length decreases and as Tone involves later alignment. This is relevant, because the vowel is the most salient part of the rhyme in perceptual terms.

**Figure 10.** Means and standard deviations for the size of the f0 fall, as a function of Tone and Stem length, starting from the peak. Separate panels by Coda type (Sonorant / Plosive) and domain (to end of nuc[leus] / C[oda]).

Figure 10C shows similar decreases in the size of the fall as a function of Stem length and Tone when the coda is an unvoiced plosive. When time pressure is maximal, i.e., when the
stem vowel is short, the tonal specification is the late-aligned Fall, and the coda is a voiceless plosive, the size of the f0 fall is reduced to 12.6 Hz. This amounts to 8.3 percent of the relevant peak value that defines the beginning of the fall (151.7 Hz).

The results of the inferential test with the size of the f0 fall as the dependent are reported in Table IX. The most sizeable main effect is conditioned by Coda type. As seen from the negative \( t \) value, the size of the f0 fall is smaller when the coda is a voiceless plosive than when it is a sonorant. There is also a main effect of Stem length, between short and both long and overlong vowels, i.e., where time pressure is more acute, but not between long and overlong vowels. Prefix length also conditions a significant difference in fall size, although the effect size is small: the probability value is close to the significance threshold (.05).\(^8\)

**Table IX.** Summary of the results of a linear mixed effects model with excursion size of f0 fall over the voiced part of the rhyme as the dependent.

<table>
<thead>
<tr>
<th>Factor and levels</th>
<th>( t ) value</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stem length – V vs. VV</td>
<td>-2.6</td>
<td>0.011</td>
</tr>
<tr>
<td>– V vs. VVV</td>
<td>-3.4</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>– VV vs. VVV</td>
<td>1.3</td>
<td>0.19</td>
</tr>
<tr>
<td>Prefix length – V- vs. VV-</td>
<td>-2.1</td>
<td>0.035</td>
</tr>
<tr>
<td>Tone – Low(^{Fall}) vs. Fall</td>
<td>-0.14</td>
<td>0.888</td>
</tr>
<tr>
<td>Dialect – BS vs. BN</td>
<td>0.02</td>
<td>0.984</td>
</tr>
<tr>
<td>Coda type – sonorant vs. plosive</td>
<td>-6.3</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Tone * Coda type</td>
<td>-2.5</td>
<td>0.013</td>
</tr>
<tr>
<td>Stem length (V vs. VV) * Coda type</td>
<td>3.5</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Stem length (V vs. VVV) * Coda type</td>
<td>5.1</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Stem length (VV vs. VVV) * Coda type</td>
<td>-2.2</td>
<td>0.032</td>
</tr>
</tbody>
</table>

There are also two significant interactions, both involving Coda type. One involves Tone. This reflects the fact that Low\(^{Fall}\) vs. Fall differ in the excursion size of the f0 fall more when the coda is a plosive (Figure 11C) than when it the coda a sonorant (Figure 11A). The second
significant interaction involves Stem length. The size of the f0 fall varies as a function of Stem length particularly when the coda is a plosive (Figure 11C).

In summary, the main effects and interactions show that the excursion size of the fall is affected by factors that have a bearing on the amount of sonorous duration in the stem syllable – in particular Stem length, Coda type – and by Tone, which affects the proportion of this sonorous domain that over which the fall is realized.

4. Discussion

4.1 Evaluation of the competing hypotheses

The null hypothesis (1) states that the alignment of falling contour tones is not phonologically distinctive. The alternative hypothesis (10) challenges this, specifically in relation to the surface phonology of Dinka.

Likelihood ratio tests were carried out to test these hypotheses directly, on the basis of the dependent measures examined in Section 3. In each test, the fully-developed mixed-effects model is compared to a baseline model in which the factor Tone (LowFall vs. Fall) and any interactions involving Tone are not included. The outcomes are reported in Table X. The contrast between LowFall and Fall significantly contributes to explaining the variability in tonal alignment, the size of the f0 fall, and duration, in that order of importance.

Table X. Outcomes of likelihood ratio tests of the models reported in Tables VII, IX, III, and VIII, respectively, each relative to a baseline in which Tone (LowFall vs. Fall) is not included.

<table>
<thead>
<tr>
<th>Dependent</th>
<th>Chi Square</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alignment of peak relative to vowel onset</td>
<td>558.1</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Excursion size of the fall</td>
<td>52.6</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Vowel duration</td>
<td>10.8</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Peak height</td>
<td>0.2</td>
<td>0.641</td>
</tr>
</tbody>
</table>

Peak alignment is to be interpreted as the primary correlate of the contrast, on several grounds. As seen from the Chi Square values in Table X, this is the dependent that yields the biggest
effect size as a function of the addition of Tone to the baseline model. In other words, the contrast between $\text{Low}^{\text{Fall}}$ and Fall affects peak alignment more than any other measure.

And whereas Tone influences peak alignment overall, i.e., as a main effect, the effect of Tone on the size of the f0 fall is conditional on the composition of the target syllable. The addition of an interaction between Coda type and Tone significantly improves on a model in which no interaction is modeled, and, with this interaction included in the model, Tone is no longer significant as a main effect (Table IX). The third measure significantly affected by the addition of Tone to the baseline model is vowel duration. It is not uncommon in tone languages for tone patterns to have a secondary effect on duration (e.g. Ho 1976 on Mandarin). Finally, the addition of Tone to the model for peak height is not significant. This is in line with the descriptive statistics, which show a difference of only 2.1 Hz. This outcome dispels an interpretation whereby the contrast of $\text{Low}^{\text{Fall}}$ vs. Fall is analysed in terms of tone height.

This outcome corroborates the alternative hypothesis, thereby confirming Andersen’s (1987:20) descriptive analysis of contrastive alignment in falling contours in Dinka. This raises the question of phonological representation: how can contrastive alignment be dealt with best in the phonology? This issue will be addressed in Section 4.3. First, however, the findings are considered in the light of phonetic limitations relating to tonal alignment. These limitations can usefully inform the question of phonological representation.

### 4.2 Phonetic limitations on alignment contrasts in contour tones

Across the world’s languages, the distribution of contour tones is often conditioned by factors determining the extent of the sonorous rhyme duration (Zhang 2001, 2002) – factors such as vowel length, the presence and nature of a coda, stress, phrase-final lengthening, etc. Ultimately, the relevance of these factors follows from phonetic limitations on the realization of f0 contours and on the perception of the resulting pitch impressions. In this section, the results of the production study in Section 3 are considered in relation to these limitations.

The duration of the interval over which a fall in f0 is realized varies greatly within the dataset. Relevant statistics are presented in Table XI. The fourth column shows the mean durations from the f0 peak to the end of the sonorant domain, as a function of three factors that affect time pressure. The means range from 48.4 ms for the Fall on a syllable with a short vowel and
a voiceless coda (CVP), up to a domain that is more than five times greater, for the Low$^{\text{Fall}}$ on a syllable with voiced coda and overlong vowel (CVVVN).

**Table XI.** Mean values as a function of Tone, Coda type, and Stem length, for the duration of the f0 fall (ms), size of the f0 fall (Hz and ST). The last column shows the difference between the mean size of the fall and the relevant glissando threshold, calculated using the formula in (12).

<table>
<thead>
<tr>
<th>Factors affecting time pressure</th>
<th>Duration of f0 fall</th>
<th>Size of f0 fall</th>
<th>Size of f0 fall – glissando threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone P V</td>
<td>48.4</td>
<td>12.6</td>
<td>1.5</td>
</tr>
<tr>
<td>Tone P VV</td>
<td>76</td>
<td>18.6</td>
<td>2.3</td>
</tr>
<tr>
<td>Tone N V</td>
<td>141.7</td>
<td>37.7</td>
<td>5</td>
</tr>
<tr>
<td>Tone N VV</td>
<td>157.7</td>
<td>36.6</td>
<td>4.8</td>
</tr>
<tr>
<td>Low$^{\text{Fall}}$ P V</td>
<td>75</td>
<td>21.7</td>
<td>2.7</td>
</tr>
<tr>
<td>Low$^{\text{Fall}}$ P VV</td>
<td>103.1</td>
<td>29.4</td>
<td>3.8</td>
</tr>
<tr>
<td>Low$^{\text{Fall}}$ P VVV</td>
<td>178.2</td>
<td>35.1</td>
<td>4.6</td>
</tr>
<tr>
<td>Low$^{\text{Fall}}$ N V</td>
<td>166.2</td>
<td>41.7</td>
<td>5.8</td>
</tr>
<tr>
<td>Low$^{\text{Fall}}$ N VV</td>
<td>200.9</td>
<td>42.3</td>
<td>5.7</td>
</tr>
<tr>
<td>Low$^{\text{Fall}}$ N VVV</td>
<td>274.1</td>
<td>43.1</td>
<td>5.8</td>
</tr>
</tbody>
</table>

The duration of the domain over which the f0 fall is realized influences its excursion size. Relative to mean values of around 150 Hz for the f0 peak that defines the beginning of the f0 fall, the mean size of the fall ranges from 12.6 Hz (Fall on CVP) to an excursion size that is more than three times greater: 43.1 Hz (Low$^{\text{Fall}}$ on CVVVN). This dependency of the size of f0 fall on the duration of the f0 fall suggests that the speed of f0 change is limited (cf. Xu & Sun 2002). Xu & Sun point out that f0 changes in speech are shaped like S-curves, i.e., there is a fast-moving central section which is bounded on either side by inert sections near the turning points. This can be seen from Figure 6A: when the Fall is realized on a syllable with a short vowel, the inert section around the peak takes up most of the vowel’s duration. When the vowel is long (VV), the speakers reach the fast-moving section well within the vowel. As for the bottom end of the falling f0 change, Figures 6A and 10A show that the inert final section absorbs the effects of time pressure, thereby buffering the high-speed central section from
truncation. This characteristic, along with final lengthening, can help to explain why the utterance-final position is conducive to the realization of contour tones (cf. Zhang 2001, 2002). The fast-moving central section is only truncated if the loss of sonorous duration is more substantial (Figures 6B, 10C).

The reduced excursion size of the f0 fall under time pressure has implications for pitch perception, because f0 changes are only perceived as pitch contours if they exceed a threshold excursion size (Rossi 1971; Greenberg & Zee 1979; ’t Hart, Collier & Cohen 1990). F0 changes below this ‘glissando threshold’ are perceived as pitch levels. On the basis of their own experiments and the results of several earlier studies, ’t Hart et al. (1990:32) model the glissando threshold as in (12). As seen from this formula, the glissando threshold increases exponentially as the duration of the domain over which the f0 change is realized (D) is reduced. For example, while a change from 150 to 135 Hz (1.83 ST) may be perceived as a pitch contour over 100 ms, the same change may not perceived as an f0 change over 50 ms.⁹

(12) \[ \text{Glissando threshold (ST per second)} = \frac{0.16}{D^2} \]

This formula enables us to evaluate whether the f0 falls in the production study are more likely to be perceived as levels or as contours. The fifth and sixth columns in Table XI provide the size of the f0 fall, and the seventh column has the size of the fall expressed relative to the glissando threshold, calculated on the basis of (12). In most conditions, the f0 falls realizing Low\textsuperscript{Fall} and Fall are two or more semitones above the threshold. But when the Fall is realized on CVP, the f0 fall is 1.8 ST below the glissando threshold. In this context, the Fall would be perceived as a level pitch. In addition, the values for a Fall on CVVP and for the Low\textsuperscript{Fall} on CVP are close to the threshold. A comparison of the values indicates that a plosive coda is more influential in bringing the size of the f0 fall down towards the glissando threshold than a short vowel. This is due to the fact that a sonorant coda represents a greater contribution to sonorous space – D in (12) – than an increase in vowel length. This helps to explain why a plosive conditions a leftward shift on peak alignment (cf. Figure 9).

The evaluation of the size of f0 falls in relation to the glissando threshold reveals that the range of pitch percepts of the Fall toneme probably comprises a level pitch pattern. This level allophonic variant of the Fall is found in the Bor North dialect of Dinka, when the Fall is realized on a syllable with a short vowel and a voiceless coda. Ohala (1989) has argued that
diachronic change in sound patterns is “drawn from a pool of synchronic variation”. One of the main mechanisms in this context is hypocorrection, whereby the listener fails to correct for a contextual perturbation. In this way, the level tone percept of the Fall could be reinterpreted as reflecting a level tone pattern in a phonological sense. This scenario of hypocorrection is precisely what appears to have happened in Bor South: here the Mid toneme is found synchronically on a short vowel in contexts where other Dinka dialects (Bor North, Luanyjang, Agar) have the Fall toneme. Figure 12 contrasts the Fall of Bor North with the Mid of Bor South, as a function of Coda type. The data are the passive forms of verbs with a short vowel (cf. Table I). When the Fall of Bor North is maximally truncated (full black line), the acoustic difference with the Mid of Bor South (full grey line) is small. Perceptually, the difference would be smaller still, because the level-tone pitch percept of a frequency change is skewed towards the end value (Nábělek, Nábělek & Hirsh 1970; Rossi 1971). In conclusion, the truncated realization of the Fall in Bor North is similar to its cognate in Bor South.

Figure 11. Averaged f0 traces on a normalized time axis of the Fall of Bor North and the Mid of Bor South, on syllables with a short stem vowel. Separate traces by Coda type and Dialect.

A third phonetic consideration is the sensitivity to the difference in timing between f0 patterns. Most perception tests relating to this issue report that it takes a difference of at least 50 ms in timing relative to the segmentals for two otherwise identical sequences of tone targets to be reliably distinguished (House 2004 and further references there). However, Bruce (1977) reports a timing sensitivity below 30 ms, albeit on the basis of stimuli that may not be ecologically valid.10 In the Dinka data under investigation here, the differences in peak alignment between LowFall and Fall all lie between 30 and 50 ms (Table XII). The difference is smallest when time pressure is maximal: in the Bor North dialect in CVP forms. However, as
noted above, in this context the contrast may be one of a pitch contour vs. a level pattern, in perceptual terms.

**Table XII.** Differences between mean peak alignment as a function of Tone (Low\(^\text{Fall}\) vs. Fall), in various conditions.

<table>
<thead>
<tr>
<th>Dialect</th>
<th>Context</th>
<th>Alignment difference Low(^\text{Fall}) vs. Fall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bor North</td>
<td>CVVC</td>
<td>45.6 ms</td>
</tr>
<tr>
<td></td>
<td>CVN</td>
<td>41.4 ms</td>
</tr>
<tr>
<td></td>
<td>CVP</td>
<td>30.9 ms</td>
</tr>
<tr>
<td>Bor South</td>
<td>CVVC</td>
<td>38.2 ms</td>
</tr>
</tbody>
</table>

The small-scale perception experiment reported in Section 2.2 suggests that speakers can use these patterns of alignment to distinguish lexical and grammatical meanings. It is not implausible that the use of contrastive alignment in Dinka makes its speakers more sensitive to differences in tonal alignment, as compared to speakers of languages where this parameter is not exploited to the same extent in the phonology. Such a state of affairs would be in line with the fact that linguistic experience of word-level tonal specification in first language influences pitch perception (e.g. Gandour 1983, Pfordresher & Brown 2009).

In summary, I have considered the results of contrastive alignment in Dinka against the background of experimental findings on the phonetics of tone. It is clear that the contrast in tonal alignment between falling contours in Dinka presents a phonetic challenge. First, the observed differences in alignment (30 to 50 ms) are low relative to the range reported in perception experiments. Second, the evaluation of the f0 falls relative to the glissando threshold reveals the difficulty in realizing the late-aligned fall as a contour in perceptual terms. The perceptual challenge is not only to distinguish the two falling contours from one another, but also to distinguish the late-aligned fall from any level tone categories. The cross-dialect comparison confirms that the late-aligned fall can be reinterpreted as a level tone pattern.

These findings are relevant to the upcoming discussion of phonological representation: they suggest that it would be highly unlikely for a language to distinguish more than two falling contour tone categories primarily through the alignment of the same tone targets.
4.3 The phonological representation of contrastive alignment

How should the contrast between $\text{Low}^{\text{Fall}}$ and Fall be dealt with in terms of phonological representation? A first approach to consider is not to phonologically represent the variation in the realization of the Low toneme at all, i.e., to leave it to phonetic implementation. This is the approach taken in Remijsen & Ladd (2008) in relation to Luanyjang, another dialect of Dinka. Such an analysis is plausible if the initial high target of the $\text{Low}^{\text{Fall}}$ is predictable from the preceding tonal context. The Luanyjang tone category that corresponds to the Mid toneme of Bor Dinka is realized as a rising contour in utterance-medial contexts. It can be interpreted as a Rise (underlying Low-High), with a High end target that is realized in the following syllable, triggering the $\text{Low}^{\text{Fall}}$ variant of a following Low.

In relation to the Bor data under investigation here, such an account is problematic for several reasons. First, there is no independent evidence that the Mid involves a high target in its implementation. That is, when the Mid toneme is not followed by the $\text{Low}^{\text{Fall}}$, it is realized as a level tone pattern, as in Figures 2 and 11. Second, if the initial high target of the $\text{Low}^{\text{Fall}}$ would be due to a pattern associated with the preceding syllable, then we would expect that the alignment of this high target would be determined more by vowel length in this preceding syllable. The descriptive statistics in Tables IV, V and the inferential statistics in Table VII indicate that the opposite is the case: the high target of the $\text{Low}^{\text{Fall}}$ is aligned most consistently with the beginning of the vowel of the stem syllable. Finally, phonetic implementation cannot account for all of the contexts in which $\text{Low}^{\text{Fall}}$ is found. Its distribution includes the isolation form (cf. Figure 2A), where there is no earlier syllable with which the high target of the $\text{Low}^{\text{Fall}}$ could be associated. The divergence in the initial f0 value between words specified for Mid vs. Low toneme suggests that there is no intonational edge tone involved here.

In conclusion, an account based on phonetic implementation is not satisfactory. Given so, the alignment contrast between $\text{Low}^{\text{Fall}}$ and Fall needs to be represented somehow in the surface phonology. The autosegmental approach to tone and intonation offers two general strategies to achieve this. One is to associate tone units to more specific elements on the metrical tier. The other is to make to develop the specification of the tonal units. These two analytic options are foreshadowed by Pierrehumbert & Beckman (1988:159), who write: “If further research uncovers cases where the alignment is contrastive within a language, these might be handled by
the use of an alignment feature on the prosodic nodes or on the substantive elements”. I will consider these approaches in turn in the following subsections, in that order.

**4.3.1 Representing contrasts on the prosodic nodes**

The metrical unit below the syllable is the syllable-internal weight unit or mora (Hyman 1984, Hayes 1989). In the study of tone, the mora is commonly invoked as the tone-bearing unit (TBU) when the distribution of tone patterns is richer in heavy syllables than in light ones (Odden 1995, Yip 2002, Gussenhoven 2004). For example, Myers (2003) reports that, in Kinyarwanda, Low contrasts with High on syllables with a short vowel, but on syllables with a long vowel Low contrasts with two tone patterns. One is falling, the other is high or rising. Myers represents these facts by postulating that short and long vowels project one and two moras, respectively, and that the mora is the TBU. It follows that a High target can associate with the first or with the second mora of a long vowel, yielding HL and LH, respectively, and that only one point of association is available on a syllable with a short vowel.

Can the Dinka contrast between Low$^\text{Fall}$ and Fall be represented through association with moras? The three-level vowel length indicates that syllable-internal weight structure matters in Dinka, and this contrast can be represented by postulating representations involving one, two and three moras for short, long and overlong vowels respectively. However, in many dialects the distribution of tone patterns does not interact with syllable weight: level and contour tones alike are found on syllables with a short, long, or overlong vowel, as in (13). This implies that the syllable as a whole is the TBU (Odden 1995, Yip 2002, Gussenhoven 2004).
In his argument for an alternative criterion to determine metrical structure within the syllable, Duanmu (1994) proposes that the association of two tone targets with the same tone-bearing unit is ruled out, so that contour tones necessarily motivate two moraic TBUs. This analysis is problematic for Dinka, given that there are minimal contrasts such as /lēl/ vs. /lēel/ vs. /lēeel/: if we invoke two moras in the case of /lēl/ so as to represent the contour tone, then by extension we end up with four moras for /lēeel/. This is typologically undesirable (e.g. Odden 2011). Moreover, even if we had two TBUs in stem syllables with a short vowel, this still does not enable us to differentiate between two patterns that both involve a High target followed by a Low one. In summary, if contrastive alignment is to be represented by making metrical specification more specific, merely postulating the mora as a TBU does not present a solution.

Two studies – Prieto, D’Imperio & Gili Fivela (2005) and Morén & Zsiga (2006) – have independently invoked a more detailed representation, whereby tone targets are associated with the left and right ‘edges’ of moras. This proposal hinges on an interpretation of the mora as a concrete phonetic constituent. For example, Morén & Zsiga (2006:126) postulate for Thai that, in a syllable whose rhyme includes a long vowel, a sonorant coda, or both, the right edge of the first mora lies halfway the phonetic duration of the rhyme. The contrast between the Fall and High tonemes of Thai is represented through a H target associated with the first vs. the second mora, respectively, in each case in relation to the mora’s right edge. Prieto et al. make the case for alignment with mora edges more generally. For a tone target within a syllable with a long (bimoraic) vowel, their model provides the possible docking sites in (14): a tone can be linked with either edge of the syllable or of the moras within it (Prieto et al. 2005:373ff).
The innovation of Prieto et al. (2005) and Morén & Zsiga (2006) expands the set of possible patterns of association, so that one mora offers two points of association rather than just one. Applied to Dinka, this enables us to represent the two falling tone patterns as in (15). The high initial targets of Low\textsuperscript{Fall} and the Fall are represented through association with the mora’s left and right edges, respectively.

\[(15)\]
\begin{align*}
\text{a. Low}\textsuperscript{Fall}: \quad \left[ \mu \right. \nu ]_{\mu} - 1 \left[ \nu \delta \right. \nu ]_{\mu} & \quad \text{b. Fall}: \quad \left[ \mu \right. \nu ]_{\mu} - 1 \left[ \nu \delta \right. \nu ]_{\mu}
\end{align*}

This approach hinges on an interpretation of the mora as a concrete timing unit. As such it can be tested on the basis of the results of the time pressure study. In the case of the Low\textsuperscript{Fall}, the representation in (15a) is in line with the phonetic data: the f0 peak at the beginning of the falling contour is aligned closest to the beginning of the stem vowel. In the case of the Fall, however, it runs into problems. If the vowel is short (monomoraic), we would expect the f0 peak to be reached at the end of the vowel, i.e., at 100% of the mora’s domain. Instead, the peak is reached about 44% into this domain: at 32.6 ms into a vowel of 74.1 ms. Contrary to the representation in (15b), the peak of the Fall is aligned closer to the left edge of the stem vowel than to its right edge. In summary, the phonetic evidence does not support an interpretation whereby the mora is used as a physical timing unit with whose edges the tone targets are lined up in a predictable way.

Detailed acoustic studies on tonal alignment in other languages are equally challenging for an interpretation whereby tones associate with the edges of moras. First, the alignment of tone targets often varies greatly across contexts. For example, the LH pattern of Kinyarwanda reaches its peak 25% earlier in the syllable in phrase-final position than phrase-medially. Based on these and other instances of variability in Kinyarwanda, Myers (2003:94) concludes that “the mora is not [to be] treated as a unit of time, or as a specific stretch of the soundwave”.

38
Second, the tone targets of a contour may be aligned well outside the relevant domain. In a time-pressure study on Mandarin Chinese, Xu (2001) reports that the High and Rise tonemes both vary in peak alignment across the right edge of the syllable with which they are underlyingly linked. This is unexpected, at least if the turning points are interpreted as an accurate reflection of the phonological tone targets.11

A more fundamental problem for moraic alignment is that it is not sufficiently restrictive. As seen in (14), Prieto et al. (2005) allow for three separate representations of a bitonal contour in a bimoraic syllable: the initial target can be linked with the left edge of the first mora, the right edge of the first mora, or the right edge of the second mora. While the current study provides evidence for a contrast between early and late alignment (Section 4.1), it equally provides evidence of limitations to the production and perception of such tone patterns (Section 4.2). For example, the f0 fall involved in the realization of the Fall on a bimoraic CVVP syllable is only 0.2 ST above the glissando threshold. If a third falling pattern were to be fitted into the same domain, it would be below the glissando threshold consistently. These limitations suggest that an inventory involving three contours, of the same shape but diverging in alignment, is beyond the range of phenomena a theory of tonal alignment needs to be able to accommodate.

4.3.2 Representing alignment on the tonal tier

Apart from the proposal to associate tone targets with mora edges, autosegmental theory does not offer a way to represent a contrast involving contours of the same shape (cf. Pierrehumbert & Beckman 1988:159, cited above). For example, contrasts distinguished through the starredness convention (e.g. H* + L vs. H + L*) involve a sequence of tone targets that are realized over consecutive syllables, rather than two targets within the same syllable domain (Beckman & Pierrehumbert 1986:257). Other proposals relating to the phonological representation of tone targets cannot serve this purpose either. Ladd (1983) proposes a feature [+/- delayed peak], but this distinguishes high or falling accents from rising accents (Ladd 1983:730), rather than contours of the same shape. Akinlabi & Liberman (2001) argue for a tonal complex with two slots, each of which can accommodate a tone target. All of these
proposals are in line with the null hypothesis, whereby languages can have only one falling contour within the syllable.

The investigations into contrastive alignment in Section 2 and 3 of this paper indicate that alignment can be distinctive in contour tones. At the same time, the phonetic limitations considered in section 4.2 suggest that such contrasts are maximally binary. Given so, contrastive alignment can be represented adequately by means of a binary feature. I invoke \([± \text{ late-aligned}]\), as an alignment feature on the tonal tier, as in (16). The content of \([± \text{ late-aligned}]\) refers to the alignment of the first target within a bitonal configuration associated with a single syllable. This feature is associated with a bitonal sequence or contour, rather than with an individual target. That is, given that alignment cannot be contrastive in level tones, \([± \text{ late-aligned}]\) is not specified for single-tone specifications.

\[(16) \quad \text{a. Low}^{\text{Fall}}: \quad (HL)[−\text{late-aligned}]\quad \text{b. Fall:} \quad (HL)[+ \text{late-aligned}]\]

The feature \([± \text{ late-aligned}]\) is descriptively adequate to represent the distinction between \(\text{Low}^{\text{Fall}}\) and Fall in the surface phonology of Dinka. Below I will argue that it is also adequate in an explanatory sense, because the boundary between its feature values in Dinka is in line with a hypothesized quantal threshold.

### 4.3.2 A quantal grounding for \([± \text{ late-aligned}]\)

In quantal theory (Stevens 1989, Stevens & Keyser 2010), phonological contrasts are explained in terms of discontinuities between speech production and speech perception, whereby, as a parameter is varied continuously in speech production, there is an abrupt change in the resulting percept. Stevens hypothesizes (1989:41) that “quantal relations at the articulatory or auditory level underlie all features”. There is evidence that the \([± \text{ late-aligned}]\) feature hinges on such a quantal threshold. Across the dataset investigated in Section 3, the mean values for peak alignment for the \(\text{Low}^{\text{Fall}}\) and the Fall are 2.3 ms and 41.0 ms into the vowel, respectively, in each case with a standard deviations of 15 ms. The category boundary for peak alignment, therefore, is in the region of 15-25 ms into the vowel. This location of the category boundary is

House carried out experiments aimed at determining the influence of the timing of f0 changes on pitch perception. The results indicate that an f0 contour will only be optimally perceived as a pitch movement if the beginning of the contour is aligned in a particular way: the f0 change needs to set in beyond the region where rapid shifts or transitions in the vowel formants convey the nature of the preceding onset consonant (cf. Blumstein & Stevens 1980). In the resulting model, in order to be optimally perceived as a falling pitch movement, “the tonal movement must be synchronized with vowel onset so that the beginning of the fall or rise occurs some 30-50 [ms] into the vowel.” (House 1990:134). This is illustrated by the black trace in Figure 12.

![Figure 12. Schematic representation of two f0 falls. The black dotted trace gives rise to falling movement perception according to House (1990:133ff); the grey trace does not.](image)

If an f0 change that is identical in direction and excursion size sets in during the onset consonant or at the beginning of the vowel, as in the grey trace in Figure 12, the resulting percept would be categorically different: it would be perceived in terms of level targets, with the end target likely to predominate.12

According to this interpretation, the perceptual space for tonal alignment is not homogeneous, with the distance in timing between two tonal configurations of the same shape within a syllable being the only factor determining whether the difference between them is perceptually salient. Instead, it is argued that a contrast in alignment within a syllable will be discriminated more easily if the constituent patterns are separated by the threshold. There is independent evidence that the perception of f0 changes is indeed different across the syllable. Verhoeven (1994) carried out discrimination tests involving continua of equidistant alignment patterns. In one type of continuum the subjects compared stimuli in which the f0 change took place in the onset or at the start of the vowel (early continuum); in other type of continuum the f0 change
set in further along in the syllable, in the course of the vowel (late continuum). Comparable continua were created for falling and for rising f0 changes. For rises and falls alike, discrimination was significantly more accurate to pairs of stimuli from the late continua.

House’s finding that the alignment of f0 changes triggers a categorical shift in perception is supported by the intuitions of specialists on tone and intonation. For example, Hyman (2010:203) discusses a rising f0 trace on a disyllabic word as follows: “[T]he second syllable seems to have a continuous rise in it, suggesting maybe a L-LH transcription. Listening to it, however, it was clear that it was perceptually [L-H]. To have been L-LH, the transition from L to H would have had to take place later in the [second] syllable.” Hyman continues to make the same argument for a falling f0 trace, again contrasting the acoustic shape as an f0 change with a level pitch impression. Hyman’s account is in line with House’s hypothesis that it takes late alignment for an acoustic contour to be saliently perceived as a pitch movement.

The findings on contrasting falling contours in Dinka support the case for a quantal threshold in tonal alignment, because the category boundary is in the region which House (1990) hypothesized to be critical in determining the pitch percept. In this way, the feature [± late-aligned] is adequate in an explanatory sense.

5. Conclusion

The results of this investigation into falling tone contours in Dinka indicate that a language can contrast two falling contours, diverging primarily in the alignment of the f0 change within a syllable. This finding contradicts the widely-held assumption that tonal alignment is not contrastive within the syllable (Odden 1995:450, Silverman 1997:479-480, Yip 2002:29).

The time-pressure study also indicates that this alignment contrast between falling contours is difficult to maintain in the face of phonetic limitations. First, the f0 change of the late-aligned pattern (Fall) is below the glissando threshold when the amount of sonorous space is maximally restricted, and in some other contexts the f0 change is only just above this threshold. This means that the Fall toneme encompasses a level pitch percept within its allophonic range. The interpretation that the Fall would be perceived as a level tone pattern is borne out in Bor South, where this toneme has been reinterpreted as a Mid toneme on short vowels, i.e., when time pressure is greater (cf. Ohala 1989). Second, the acoustic difference in alignment between the
two falling contours is at the bottom end of the range of values reported for sensitivity to timing differences in perception studies (House 2004). The relevance of time pressure in Dinka is in line with cross-linguistic observations regarding the dependency of contour tone realization on the availability of sonorous space (Zhang 2001, 2002).

The evidence of phonetic limitations is relevant to the question of phonological representation: given the difficulty to produce and perceive binary contrast in tonal alignment between contour tones, it is unlikely that a system of contrasts involving three or more levels of tonal alignment could be maintained. Based on this consideration, I have argued for a representation of tonal alignment through a feature ([± late-aligned]), i.e., a representation that inherently imposes a maximum of two levels. This consideration is also part of the argument against a representation with reference to mora edges, which allows for three or more levels of contrast (e.g. Prieto et al. 2005). Finally, the location of the category boundary in Dinka, within the first 30 ms of the vowel, is in line with the model of House (1990). This suggests that the alignment feature invoked for Dinka may exploit a quantal threshold (Stevens 1989, Stevens & Keyser 2010).

Aside from contrastive tonal alignment, the current study also presents evidence of another suprasegmental phenomenon that is typologically unusual. Three-level vowel length in Dinka was first reported in Andersen (1987), and has been corroborated on the basis of phonetic evidence (Remijsen & Gilley 2008). The current study strengthens the case for three-level vowel length in the surface phonology of Dinka in two ways. First, it widens the range of dialects for which three-level vowel length has been reported. There is now evidence of three-level vowel length in three of the four dialect clusters, namely Agar (Andersen 1987, 1993), Rek (Remijsen & Gilley 2008), and Bor (this study). Second, the results provide additional phonetic evidence, based on sets for three-level vowel length in which one is kept constant.

References


**Abbreviations**

<table>
<thead>
<tr>
<th>DC</th>
<th>declarative</th>
<th>PASS</th>
<th>passive</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>plural</td>
<td>S</td>
<td>singular</td>
</tr>
<tr>
<td>PREP</td>
<td>preposition</td>
<td>U</td>
<td>uncountable</td>
</tr>
</tbody>
</table>

**Acknowledgements**

I gratefully acknowledge discussions with Johanneke Caspers, Ingo Hertrich, Larry Hyman, Bob Ladd, and Alice Turk. I thank Yi Xu, for sharing the algorithm to trim spikes in f0 traces, and Mike Allerhand and Mateo Obregón, for support with statistics. Earlier versions were presented at the ANU Tone Workshop (Canberra, 4-16 of December 2011), and at the 20th Manchester Phonology conference (Manchester, 24-26 of May 2012). The final version has been improved on the basis of an evaluation by Kie Zuraw, the associate editor, and by two anonymous reviewers. I thank them for their valuable feedback.

The speech data on which this research is based were recorded from John Penn de Ngong, Peter Garang Nyarjok, Abraham Pach Alier, Job Anyang Aluong, James Maker Riak, and Mary Agotich Buol for Bor South, and from Daniel Akoy Deng, Leek Deng Mawut, Leek Deng Lueth, Akol Kongoor Reech, Abraham Duot de Khueer, Aluel Ajaang Jibil, Abraham Leek Makuei, Emmanuel Deng Jogaak, and Simon Yak Deng for Bor North. The last three also took part in the perception study, as did Elizabeth Achol Ajuet, Peter Malek Ayuel, and David Dau Arop. I gratefully acknowledge their involvement, and in particular John Penn de Ngong, who provided assistance with data elicitation.

All of the data were collected in Juba (South Sudan), in the context of research trips sponsored by SIL Sudan (3 trips) and by The British Council (1 trip). I thank them, and also the community at the SIL compound in Juba, for hospitality there. Finally I gratefully acknowledge the Arts & Humanities Research Council (United Kingdom), which funded this research through the project ‘Metre and Melody in Dinka Speech and Song’, as part of the Beyond Text initiative.
Notes

1 The abbreviations in the glosses are explained at the end of the paper.
2 F0 is expressed using the Equivalent Rectangular Bandwidth (ERB). This scale provides a more accurate representation of human pitch perception than the hertz scale (Hermes & van Gestel 1991, Nolan 2003).
3 I speculate that the ‘harsh’ and ‘hollow’ voice qualities reported for Dinka in Edmondson & Esling (2006) on the basis of fibroscopic evidence correspond to these two allophonic variations, respectively.
4 The control set involved the verb forms /ā-wèel/ ‘DC\S-turn.\over\1S’ and /ā-wèel/ ‘DC\S-turn.\over\3S’.
5 The Twic dialect in the Rek cluster is different from the Twic dialect in the Bor cluster.
6 The combination of the Fall on an overlong vowel is found among others in long verbs inflected for spatial deixis. In pilot data collection, it became clear that these inflections are not available for all verbs, and that they are difficult to elicit in a controlled recording session. This is why I did not include this combination in the design.
7 For the sake of comparison, the mean duration for the Mid toneme on a short vowel in Bor South is 73.9 ms.
8 The effect of Prefix length on the excursion size of the fall may be due to the influence of Prefix length on the duration of the vowel of the stem syllable. That is, mean stem vowel duration is slightly shorter after a long prefix vowel than after a short prefix vowel – the mean values are 117.5 and 115.6. As seen from Table III, this difference is not significant.
9 If D is 100 ms (0.1 second), G is 16 ST per second, or 1.6 ST over 100 ms. If D is 50 ms (0.05 second), then G is 64 ST per second, or 3.2 ST over 50 ms. Hence a change of 1.83 ST is above G for a duration of 100 ms, but below G for a duration of 50 ms.
10 Bruce’s stimuli involve falls from 160 to 100 Hz (8 ST) over 40/60/80 ms. In contrast, the fastest speaker (out of 33 subjects) in the study by Xu & Sun (2002:1407) takes at least 97 ms to produce an f0 change of this size.
11 Xu (2001) argues for an alternative theory, in which f0 slopes may also reflect tone targets.
12 In the perception of the pitch level of frequency changes, the influence of the end frequency predominates (Nábělek, Nábělek & Hirsh 1970:538, Rossi 1971).