Word-level Accent in Zarma

Dine Mamadou

[1] Rutgers University

Word-level Accent in Zarma

1 Introduction

The question investigated in this paper is whether a lexical tone (F0) language with contrastive vowel duration can also bear a word-level accent; and if so, how is that accent realized in a context where (most of) the traditional acoustic correlates of accent are already used contrastively in the language? More specifically, this study aims not only at finding phonetic evidence for a word accent in Zarma - as impressionistically claimed by Hamani (1982) - but also to test the validity of the Functional Load Hypothesis, which makes testable predictions on the acoustic realizations of co-existing prosodic systems, like *lexical tone* and *accent* systems. Zarma is interesting in this regard as it is a language with lexical tone, contrastive vowel duration and constrastive consonant gemination, which are respectively realized with pitch (F0), vowel duration and consonant closure duration, all potential cues for accent (Gordon, 2011; Gordon & Roettger, 2017). Although Hamani's claim singled out intensity (a cue that was not used contrastively elsewhere in the language) as the positive cue for accent in Zarma, this claim still needs to be experimentally investigated, which is one of the goals of the present study.

For this purpose I adopt Downing (2010)'s general defition of *accent* as a prominence asymmetry that makes a syllable the most sailiant within a domain *by enhancing some combination of phonetic properties* (Also see van der Hulst 1999; Downing 2013). This definition has the advantage of making explicit the need for the *accent* to be realized phonetically. Another use of *accent* defines it as an abstract property of a word or morpheme, with no specific reference to how it is (phonetically) realized (Hyman 1977; Fox 2002; and others). Under this tradition, as van der Hulst (2010) puts it, compounds of the form *X-accent* are adopted, where *X* represents the cue that correlates with the *accent* (e.g: in *pitch-accent* languages, the *accent* is realized with pitch as the main cue). That use is not adopted here.

The results of the production study presented here show that Zarma does in fact have a word-level *accent* that is realized on the *peninitial* syllable with an *increased intensity* and a *centralization* of the vowel. As such, Zarma exemplifies an exceptional case of vowel centralization in a prominent position, unlike the cross-linguistic tendency to have more peripheral vowels in prominent positions. The results of the experiment further suggest that the Functional Load Hypothesis (FLH; Vogel et al. 2015, and many others) needs to be weakened because the same cues are used to mark more than one contrast in Zarma. To account for this, we propose a Relativized Functional Load Hypothesis (RFLH), which can reset in each one of the following three categories of the Prosodic Hierarchy (Selkirk 1978): the *Rhythmic*, *Pre-interface* and *Interface* categories (Itô & Mester, 2012).

The paper is organized as follows: *Section 2* presents the relevant background on Zarma and the accent related claims in the language; *Section 3* introduces the details of the experiment, while *Section 4* presents its results, along with the discussion of some minor points. The major findings of the study and their impplications for thoeries of prosody are discussed in *Section 5*. *Section 6* sketches the connections of the results to other phonological phenomena in the language and *Section 7* concludes.

2 Background

This section presents a snapshot of the Zarma language, its phonology and prosody as discussed in previous work along with an overview of the claims that Zarma has an *accent*.

2.1 Overview of the Zarma Phonology

Zarma is a Nilo-Saharan language of the Southern Sonray-Zarma group (Bendor 1997, Nicolaï 1980). It is a lexical tone language with two level tones, H(igh) and L(ow), and two contour tones, falling and rising as shown in ((2)) (Hamani, 1982b; Tersis, 1972). It has a five vowel system, /i, u, e, o, a/ with their long counterparts, meaning vowel length is contrastive in Zarma (as shown in (3)). Another noteworthy characteristics of the phonology of Zarma is that it has contrastive geminate consonants as shown in example (1)[1]. The attested syllable structures are V, CV, CVV/CV: and CVC.

(1) Contrastive Geminate

a. zìmà “priest” vs zìmmà “hit hard”

b. kùrú “graze” vs kùrrú “drag”

(2) Lexical Tone

a. bí “cotton thread”

b. bì “wound”

c. bî “black”

d. bǐ “yesterday”

(3) Contrastive Length

a. kòmá “run!” vs kò:má “bump”

[1]The examples given here are from Tersis (1972), except (1-b) and (3-b), which are from Bernard and Kaba (1994).

b. kùrú “graze” vs kù:rú

2.2 On Accent in Zarma

Most previous work on Zarma were limited, as far as its prosody is concerned, to tone and intonation. Hamani (1982) is one of the few to introduce *accent* into the discussion of Zarma prosody by claiming that Zarma words have an intensity accent on their second syllable, and that the accent is only present on some but not all words. According to Hamani, this accent “highlights the element that is at the center of the communication” (Hamani, 1982); that is, words with a special syntactic and/or pragmatic status. Furthermore, Hamani argued that this accent is not contrastive in Zarma, although it can be used to discriminate between some complex nouns and their morphologically equivalent *Noun + Adjective* counterparts as in **háw bî** “black cow” and **háw-bî** “buffalo”. In the former, the accent can either be on **háw** “cow” or **bî** “black”, while it is obligatorily on the first syllable in **‘háw-bî** “buffalo”[2].

To summarize, Hamani made the following three, experimentally testable, points:

a. Zarma has an intensity accent
b. The accent is realized on the second syllable
c. The accent only shows up on words with a special status in the sentence, e.g. Topic, Focus, etc. In other words, the accent is *phrasal* rather than *lexical* or *grammatical*.

Prior to Hamani, Nicolaï (1980) argued that the prosodic system of Southern Sonray-Zarma languages is undergoing a transition brought about in contexts where speakers are immersed in other language environments. The source of this transition in Zarma, he argued, is the large number of loanwords it borrowed from Tamasheq, an areal Berber language of the Afro-Asiatic group, meaning Zarma is acquiring a Tamasheq-like prosody. Although Nicolaï did not give any specifics about what exactly is changing in the prosody of Zarma, Tamasheq however, has been described as a language with contrastive stress in the stative aspect of verbs (Sudlow, 2001). In this sense, to acquire a Tamasheq-like prosody translates into aquiring a stress-like prosody.

In fact, the idea of a stress system in the prosody of Sonray-Zarma languages is not new, as shown by Nicolaï (1980)“s extensive study of the language continuum of the Sonray-Zarma group. Nicolaï’s study revealed the existence of three different prosodic

[2]Hamani noted that compound words are accented on their first syllable, unlike simple words. That point will not be particularly investigated in the current study which only focuses on morphologically simple words.

systems in this continuum: fully lexical tone systems (e.g. Zarma), stress systems (e.g. Tadaksahak) and lexical tone systems with a lost of contrast on certain vowels (Tasawaq) (also see Alidou 1988; Wolff and Alidou 2001 and Kossman 2007 for discussions on Tasawaq's ‘mixed” status). Nicolaï argued however, that this ongoing transition does not mean that there are more than one prosodic system in any one of these languages at any particular point in time, which leaves unanswered the questions about the evidence for arguing that Sonray-Zarma languages in general, and Zarma in particular, are acquiring a stress-like prosody in the first place. The experiment presented below answers precisely that question.

3 Experiment

To answer the questions raised in the sections above, a production experiment was conducted to collect speech data from native speakers of Zarma. The subsections below present details about the design of the experiment.

3.1 Targets

The target words used in this experiment are made of trisyllabic words (24 out of the 31) of the type C<u>V</u>C<u>**V**</u>CV and disyllabic words of the type C<u>V</u>C<u>**V**</u> (7). The first and second (underlined) were compared to see if the second syllable (bolded) is accented as hypothesized in (3.5) below. Several controls (segmental, prosodic, and lexical) were implemented to make sure the measurements are not biased by aspects of the language that interact with the potential accent cues we are interested in, namely intensity, duration, F0, and vowel F1 & F2 (Shih, 2016; Liberman, 1960; Gordon, 2011; Gordon & Roettger, 2017).

On the segmental level, a subset of the target words have the form OVXVSV, where O=obstruents ([p, t, k, f, s]), T=voiceless oral stop [p, t, k] and S=sonorants [l, r, m], and another subset with the form OVTVOV. The former made up 15 counts of the trisyllabic words while the latter were only 9 words. As far as disyllabic words are concerned, they are of the form OVSV (4 words) and OVOV (3 words). These controls make it easy to identify vowel boundaries while keeping segmental effects on pitch and vowel duration constant **?**. Short vowels, namely [u], [a], [o] and [i] are used as syllable nuclei. None of the target words contained the vowel [e] due to lexical gaps given the controls implemented.

On the prosodic level, only words with H tones were used. To avoid potential H tone boosts when adjacent to L tone (Akinlabi and Liberman, 1995), words with homogenous tones (H) were used (e.g: kúkúsí “to rince”, fátámá “louse”, etc). Having homogenous tones make it possible to reliably consider f0 as a potential cue for the accent under study. Finally on the lexical level, all the trisyllabic words were made of single morphemes except when they are nouns, in which case their final vowel may be turned into /ô/ to mark definiteness. Note that the final H level tone would, in those cases, be replaced by a falling tone. This should not be a problem because contour tones are analyzed as a sequence of two level tones in Zarma (Hamani, 1982; Nicolaï, 1980). As such, the tone of the second syllable in trisyllabic words is still technically followed by a H tone, just like in words with all H tones. In addition, due to lexical gap, we could not limit the target words to one grammatical category, say only nouns or verbs. So, the stimuli contained Zarma verbs, nouns and adjectives, and no *nonce* words.

3.2 Frame sentences

Two frame sentences were used to serve as natural environments for the target words. The first frame sentence accommodated the target words in a focused position, thus emphasized position while the second one placed it in a regular (unfocused) position. Participants were asked to imagine themselves being asked by one of their sibblings the word that their classmate Rákìyá wrote on her notebook two days earlier and whether Rákìyá wrote the same word on her table then. Each target word will then be the word that Rákìyá wrote on her notebook and at the same time the word she did not write on her table. As such, the word will be (narrowly) focused in the first frame sentence but unfocused in the second frame sentence. The sentences are as follows:

Focused: bìí fó [target] nô Rákìyá hàntúm à tírà rà.

Gloss: yesterday one [target] TOP Rákìyá write her notebook inside

Meaning: ‘Two days ago, it is [target] that Rákìyá wrote in her notebook.’

unFocused: Rákìyá mǎn [target] hàntúm téébùró bòŋ.

Gloss: Rákìyá NEG [target] write table-Déf on

Meaning: ‘Rákìyá did not write [target] on the table.’

Note that the syllables of the frame sentence words that are adjacent to the target word position have H tones except for the initial syllable of *hàntúm*. This will not pose any major problems since only the first syllable and the second syllables will be measured and compared. The frame sentence type is included as a fixed effect variable in the statistical analyses presented below.

3.3 Participants

The data collection happened in two phases resulting in the participation of a total of 13 people. The first phase consisted of putting together a list of mostly trisyllabic Zarma words and creating a scenario in order to generate two frame sentences that would place the target words in focused and unfocused positions. Two native speakers of Zarma (1 female, 1 male) - judged on the basis of the background information they provided - helped preselect the target words and fix the scenario described above. The second phase happened in Niamey, Niger where 11 people were recruited. One of them (Male, 44) who had a formal training in linguistics served as a consultant and helped get a final version of the target words and frame sentences. The remaining 10 (6 female, 4 male; aged 23-47) were recorded in the main experiment. The results reported in this paper are from 3 of the 10 participants recorded (2 male and 1 female).

Demographic and linguistic information were collected through a questionnaire that the participants filled out. Despite the variability in the places of birth of the 3 participants reported here (Niamey, Kollo and Birni, respectively), they all moved to Niamey as children and have been speaking Zarma from birth and still speak it 50-60% daily in non-formal environments (e.g: home). It's worth mentioning that the female and one of the male participants were each fluent in at least 2 other languages, namely Hausa and French. These demographic information suggest that the participants have a non-atrophied Zarma L1 (c.f. Fremed, 2003), thus meet the requirements to participate in the present study.

3.4 Recording sessions

The participants were recorded in a quiet room with an AKG C420 microphone mounted to their head to keep the microphone steady and maintain distance from the mouth constant throughout the recording. The microphone is then connected to a Steady State MARANTZ PMD670 recorder that recorded at 44.1k Hz sampling rate and 16-bit quantizing resolution in mono. The participants were instructed to speak at normal conversational rate. The stimuli appeared on a laptop screen along with an image portraying a girl (Rákìyá) writing in her notebook. The participants were instructed to press the forward key to advance to the subsequent word.

Next to the computer was a sheet of paper displaying the two frame sentences to help the participants remember. Target words (31) and fillers (32) were semi-randomized and counterbalanced. Each of the words was pronounced in the focused frame sentences and then in the unfocused one. Every word was repeated 3 times during a single recording session (per participant), with each repetition of all the words lasting about 20 minutes. The repetitions were seperated with short pauses (5-8min). All three sets of repetitions started with 6 fillers in order to correct any hyper- and/or mis-articulation from the participant at the beginning of each session. The 63 target x 2 frames x 3 repetitions yielded a total of 378 tokens per participant. Training through the repetition of 10 separate, non-experiment words pronounced with the two frame sentences was provided to all the participants.

3.5 Hypotheses

The following hypotheses are based on Hamani (1982)'s claims about accent in Zarma:

Hypothesis 1: The second syllable is accented.

[Prediction 1]: The second syllable of trisyllabic and disyllabic words will have a higher intensity than the first syllable.

Hypothesis 2: Only focused words are accented.

[Prediction 2]: The second syllable of Focused words will have a significantly higher intensity than the first one, but that asymmetry will be absent in unFocused words.

Hypothesis 3: Each acoustic cue is used to express only *one* contrast in the language.

*[Prediction 3]: Acoustic cues like *duration*, *F1&F2*, and *F0* will not vary as a function of vowel position or focus status of the target word; only *intensity* may.*

3.6 Data Processing

The soundfiles that resulted from the recording sessions were imported into Praat where TextGrids were created and annotated to extract Intensity, Duration, F0 (pitch), F1 (first formant) and F2 (second formant) data. The annotation consisted of marking interval boundaries for the first and second vowels in both disyllabic and trisyllabic target words. The left boundary was marked at the nearest zero crossing of the first well-formed periodic waveform coinciding with a clearly defined first formant onset in the spectrogram (Peterson and Lehiste, 1960). As for the right boundary, it was placed at the nearest zero crossing following a dip in the intensity tracker. When a clear intensity dip was not observed, the boundary was placed at the last clearly formed second formant striation in the spectrogram (Turk et al., 2006).

Each delimited interval is labeled to contain information about the *vowel* (i, u, o, a) in question, the *position* of that vowel in the word (either fist or second), the frame sentence/*context* (either Focused or Unfocused), the *repetition* (first, second or third) and the *class* of the segment that follows the second syllable (sonorant or obstruent). For example, **r1s7a1** encodes the following: the vowel is from the *first repetition* of a target word with a *sonorant* consonant following the second syllable and the vowel in question is [a] in the first syllable position, pronounced within a *focused* frame sentence. The digit *7* uniquely identifies the target word. These factors were used in the fixed and random effect structures of the statistical models. A modified version of the praat script SpeCT (Lennes, 2011) was used to collect the above-mentioned acoustic cues. For each boundary, the script returned the duration in ms, the mean Intensity in dB, F1 & F2 in Hz, and the maximum F0 (formerly F0) in Hz.

The resulting data was imported into R (Pinheiro et al., 2017) for the statistical analysis. The *lmer()* function in the lme4 package (Bates et al., 2015) was used to fit nested omnibus linear mixed effect models for the random effects on the one hand and the fixed effects on the other. The measurements of the five cues (intensity, duration, f0, f1 & f2) were normalized (z-score) to allow for between-subject comparisons despite individual differences such as speech rate and difference in F0 ranges as a function of gender (female vs male).

All random effect models containing *position*, *gender*, *word* and segment *type* as random slopes did not converge. As a result, the random effect structure only consisted of *word* and *participant* as random intercepts. No effect of these two intercepts was observed throughout the data. As for the fixed effect models, the acoustic measurements of Intensity, Duration, F0, F1 & F2 were each used as a dependent variables in a set of models with either *position* or *frame* (depending on the data subset that is being used), *type* and *repetition*. Since this study is not interested in the effects of gender per se, the dependent variables were grouped by gender and normalized to allow for inter-subject comparisons. Doing so made it possible to remove the variable *gender* from the predictors.

In order to test for significance, the nested models were compared via a Likelihood Ratio Test using the *anova()* function in R. The residuals of the models were visually inspected to make sure the model assumptions hold, namely the assumption of normality and homoskedasticity (equal variation of the data). These assumptions hold for the most part; cases where they are violated are mentioned and include small subsets of the data (e.g: F0 of the vowel [a] in unfocused words).

4 Results

The results of the study are organized in the following way: in order to test *Hypothesis 1* and *Hypothesis 2*, the first vowel of trisyllabic words was compared to their second vowel in the Focused context on one hand and in the unFocused context on the other hand. Next, the

Focused context was compared to the Unfocused one in the first syllable of these same trisyllabic words and then in the second syllable position to test *Hypothesis 3*. This comparison did not include the third syllable of trisyllabic words (c.f. Vogel et al. 2015 for a similar methodological choice). Given the limited number of the disyllabic words (7 words), they are used only to verify *Hypothesis 1* and *Hypothesis 2*. That is, their first (initial) syllable will be compared to their second (final) syllable in the unfocused context only, in order to determine the edge of the word marked by the accent under investigation (*Demarcativity*; Hyman 1977; Downing 2010). Due to the different controls implimented on the selection of the disyllabic words, only [i] and [u] are available in the disyllabic words whose results are reported here. As far as the trisyllabic words are concerned, the results presented below are based on four of the five Zarma short vowels, namely [i], [u], [a] and [o]. The vowel [e] was not included due both to the segmental and prosodic controls and to the vowel [e]'s limited distribution in those positions (**?**).

4.1 Trisyllabic Words

In this section, I present the results of the comparison between the first and second syllables of the trisyllabic words.

4.1.1 First vs Second syllable. Below, the first syllable of trisyllabic words is compared to their second syllable, first in the Focused frame sentence and then in the Unfocused one. *Position* stands for the **first** or **second** syllable positions of the vowel, *Frame* refers to the frame sentences, thus, to whether the target word is in a **focused** or **unfocused** frame sentence. As far as *Type* is concerned, it represents the natural class of the segment that follows the second syllable (either **obstruent** or **sonorant)**. The raw values were normalized (z-score) for all five cues for the purposes of the statistical tests and to make comparisons possible between the different acoustic cues. Importantly, in cases where more than one fixed effect variable was found, we checked for potential interactions between these fixed effect variables but no interaction was found.

4.1.1.1 Intensity. The statistics for the intensity data presented here is based on the two male speakers and did not include the production of the female speaker due to inconsistency in her production of the different cues across and within her repetitions of the target words. However, as the participant-based intensity plot in *Figure 2* shows, the female participant's data still patterns like the rest of participants. In fact, her data were more consistant with the pooled data than that of the participant *M01* whose intensity realization in the second syllable was not so different from the first one due to his intensity being very high overall. Due to the fact that her intensity patterns with regard to the position of the vowel paralleled that of the pooled data, her data was included in the dB values given in parentheses below and in the summary table in *Table 1*.

That being said, in Focused words, the linear mixed effect models for the mean intensity did not show any effect of *position* ($X^2(1) = 1.79$, $p = 0.18$). That is, intensity in the second syllable was 1.1 dB higher than in the first syllable of focused words. There was rather a main effect of *type* ($X^2(1) = 4.39$, $p=0.04$). More specifically, vowels followed by sonorants were 0.32 standard units +/- 0.15 (roughly 1.5 dB) higher in intensity than the vowels followed by osbtruents ($df = 22.06$, $t = 2.20$, $p = 0.04$). With Unfocused words, however, only *position* had a main effect on vowel intensity ($X^2(1) = 8.9$, $p = 0.0029$); that is, the mean intensity value of vowels in the second syllable were 0.18 standard units +/- 0.06 (about 1.3 dB) higher than that of vowels in the first syllable ($df = 246.91$, $t = 3.01$, $p = 0.0029$).

The effect of position supports the hypothesis that the second syllable is more prominent than the first one and that intensity is (at least one of) the correlate(s) of that prominence as shown in *Figure 1*. The Just Noticeable Difference (JND; Klatt 1976) for intensity, i.e the minimum difference threshold in intensity that can be picked up by human ears, is 1 dB (Harris, 1963; Shih, 2016). In Zarma, the intensity difference between the first and second syllables passed that threshold (both in Focused and Unfocused words), confirming that the accent referred to by Hamani (1982) is in fact realized through intensity but contradicts the idea that only Focused words have accent. The fact that Zarma's intensity difference passes that threshold also makes it learnable to children acquiring the language. The lack of statistical difference between first and second syllable positions is certainly due to a 'ceiling effect', which in the present case reduces the contrast between the two positions without neutralizing it (compare Focused [1st.Syll = 78dB, 2nd.Syll = 79.1sB] to Unfocused [1st.Syll = 74.4dB, 2nd.Syll = 75.7dB]).

The pooled intensity results in the *Figure 1* reflects individual results as shown in the per-participant results in *Figure 2*.

4.1.1.2 Duration. Although duration was not predicted to vary as a function of the syllable *position*, it was examined nonetheless to see if it constitutes a secondary cue for prominence/accent realization in addition to Intensity. As expected, the duration models in focused words showed an effect of *type* ($X^2(1) = 10.03$, $p = 0.0015$) but not that of *position* ($X^2(1) = 0.04$, $p = 0.83$), such that vowels that were followed by sonorants were 0.10 standard units +/- 0.073 (roughly 8.8 ms) longer than the ones that are followed by

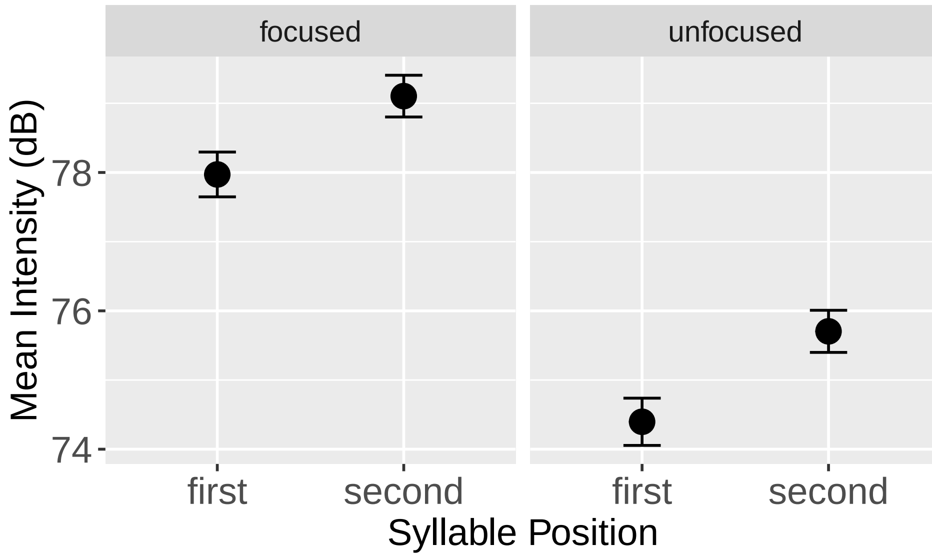


Figure 1. Intensity in dB as a function of the focus status of the target word and of syllable position for all three participants.

obstruents (df = 403.41, t = 1.366, p = 0.17) as shown in the table below. The same effect of type was observed in unfocused words as well ($X^2(1) = 14.89$, p = 0.00011) with pre-sonorant vowels being about 7.8 ms longer than pre-obstruent vowels. Additionally, *position* was found to be marginally significant ($X^2(1) = 3.89$, p = 0.048), such that the second syllables were 0.13 standard units +/- 0.07 (about 3 ms) longer than the first syllable (df = 382.68, t = 1.96, p = 0.05).

This effect of *position* on duration is marginal and only constitutes about 7% difference between the two positions (compared to the 20% Just-Noticeable-Difference (JND) proposed in Klatt (1976). Since this small overall duration difference is not reliable, individual speakers' data were inspected. Two of the participants (F01 and M01) had their second syllables about 2~6ms longer than their first syllable, while the third particpant (M02) had approximately the same duration values in both positions. Although the per-participant results do not suggest anything radically different from the pooled data, the fact that the differences are still below the JND undermines the importance of the statistical significance found with duration between the two positions. Another source of evidence for why this difference is unimportant comes from comparing it to the difference between phonemically short and long vowels in the language: short vowels are in average 46ms in the initial position while long vowels are about 120ms in the same position, yielding a 260% duration difference between short and long vowel (compared to the 7% difference between the two positions found here). Hence, this duration difference is best analyzed as a kind of

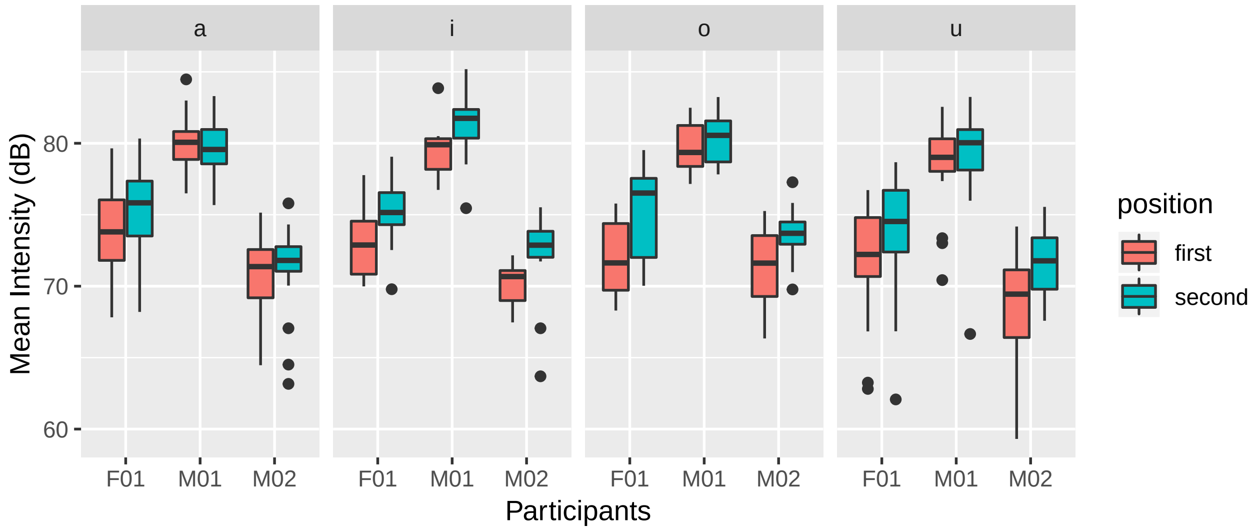


Figure 2. Intensity (in dB) in the first and second syllable positions of unfocused trisyllabic words for all three participants and the four vowels.

anticipatory shortening of the first syllable, following Remijsen (2002).

4.1.1.3 F1. *Table 2* summarizes the mean F1 value for high vowels [i] and [u], mid vowels [o] and low vowel [a]. F1 models were fitted on the basis of vowel height. [i] and [u] were thus fitted together but no main effect of vowel position was observed ($X^2(1) = 2.97$, p $= 0.08$) in focused words. The low vowel [a] showed a main effect of *position* ($X^2(1) = 4.08$, $p = 0.04$) as well as *type* ($X^2(1) = 5.39$, $p = 0.02$). More specifically, second syllable [a] is -0.15 standard units +/- 0.07 (about 14.67 Hz) higher than its first syllable counterpart (df = 178.12, t = -2.03, p = 0.04) whereas pre-sonorant [a] was 0.39 standard units +/- 0.15 (roughly 50.37 Hz) higher than the pre-obstruent one (df = 10.70, t = 2.66, p = 0.02).

In unfocused words, a main effect of position was observed for [i] and [u] ($X^2(1) =$ 12.28, $p = 0.00046$) on the one hand and [o] ($X^2(1) = 18.45$, $p = 1.741e\text{-}05$) on the other. That is, second syllable [i] and [u] were 0.38 standard units +/- 0.11 (roughly 49 Hz) lower than when they appear in the first syllable of the word (df = 17.1, t = 3.39, p = 0.00089). That difference was of 0.56 standard units +/- 0.12 (roughly 76.60 Hz) for [o] (df = 50, t = 4.7, p = 2.08e-05). In addition, [o] showed a main effect of *type* ($X^2(1) = 9.87$, $p = 0.0017$), such that when [o] is followed by a sonorant, it was 0.41 standard units +/- 0.13 (about 58.37 Hz) higher than when followed by an obstruent. No main effect was observed for [a].

These results suggest that vowels in Zarma are significantly lower in the second syllable position (higher F1) than they are in the first syllable: [i], [u] and [o] are all lower in the former position than in the latter, at least when they are unfocused as shown in *Figure 5.* In Tadaksahak (a Northern Sonray dialect) with lexical stress, Nicolaï (1980) reported that F1

Table 1
(#tab:loading scripts)Mean Duration (in ms) and standard deviation in the first and second syllables for focused and unfocused target words.

frame	position	Duration	Std.Dev
focused	first	50.75	14.68
focused	second	52.46	17.29
unfocused	first	45.45	12.70
unfocused	second	48.28	14.44

is correlated with stress, but it was unclear in what direction the correlation goes. In the present case, the data supports the view of a positive correlation between the prominent second syllable (by hypothesis) and the realization of F1, which is higher in that position. Hence, vowels are lower in the second syllable position than in the first[3]. The fact that Zarma (a lexical tone language) and Tadaksahak (a lexical stress language) use the same cue, regardless of the potential differences in how this cue is used, seems to lend additional support to Nicolaï's position that Zarma is acquiring a stress-like prosody (I return to this in the discussion).

4.1.1.4 F2. In focused words, only the vowel [i] showed a main effect of *position* ($X^2(1) = 6.04$, $p = 0.01$), such that [i] in the second syllable was -0.47 standard units +/- 0.18 (roughly 56 Hz) less fronted than in the first syllable (df = 38.4217, t = -2.55, p = 0.01). A marginal effect of *type* was also observed ($X^2(1) = 3.62$, $p = 0.06$). In the unfocused words, the likelihood ratio test for the back vowels [o] and [u] (fitted together) showed a main effect of vowel *position* ($X^2(1) = 9.30$, $p = 0.0022$), such that back vowels in the second syllable were 0.32 standard units +/- 0.11 (about 99 Hz) less fronted than in the first one (df = 157.66, t = 3.06, p = 0.0026). Additionally, [i] showed an effect of *position* as well ($X^2(1) = 6.66$, $p = 0.0099$) by being -0.35 standard units +/- 0.13 (about 94 Hz) less fronted in the second syllable than in the first (df = 55.79, t = -2.75, p = 0.008). Again, no main effect was observed with the vowel [a]. The F2 data patterned like the F1 data in that vowels tend to be more fronted when they are back vowels and less fronted when they are front vowels in the second syllable position than in the first syllable of unfocused words. When these F2 results are viewed together with the F1 results, it turns out that vowels are more centralized in the second syllable than in the first one, at least for unfocused words as shown in the rightmost plot of *Figure 3*.

[3]Crutially, this fact can not be attributed to the lack of balance in the distribution and frequency of vowels in the two positions (in the target words) because there are approximately the same number of high and low vowels (**V1**: 10 high vowels[i&u], 12 low vowels [a] and 3 mid vowels [o]; **V2**: 11 high vowels[i&u], 10 low vowels [a] and 3 mid vowels [o].

Table 2
(#tab:loading scripts)Mean F1 (in Hz) and standard deviation in the first and second syllables.

vowel	position	F1	Std.Dev
a	first	494.06	113.21
a	second	492.91	145.52
i	first	313.35	52.40
i	second	354.04	69.75
o	first	436.66	100.83
o	second	476.66	88.69
u	first	326.76	59.55
u	second	370.93	186.35

4.1.1.5 F0. The f0 results reported here are those of the 2 male speakers. The female speaker's data varied significantly between repetitions and was removed from the statistical models for the lack of consistancy it displayed, just like for the intensity data. In focused words, likelihood ratio tests did not show any main effects of *position* or *type* in any of the four vowels with f0 as the dependent variable. Similarly, no effect was observed in the unfocused words, except for a very marginal significance of *type* ($X^2(1) = 4.11$, $p = 0.043$) for the vowel [a]. An inspection of the [a] data residuals, however, suggests that the data violated the model assumption of homoskedasticity; that is, the data is categorical rather than continuous (see Appendix).

To summarize, all the cues examined to the sole exception of f0, showed some level of differences as a function of position and/or focus status of the target word. Intensity and Duration values were significantly higher in the second syllable position than they were in the first one. As for the formant structure, F1 values were higher for high vowels (in unfocused words) and lower for low vowels (in focused words) in the second syllable than they were in the first. F2 values were lower for front vowels (with focused words) and higher for back vowels (with unfocused words) in the second syllable than in the first, meaning that vowels are more centralized with regard to their F2 value in the second position. These results suggest that the second position is in fact a prominent one and that prominence is manifested through *Intensty*, *Duration* (to a neglectable extent) and, in an unexpected way, the formant structure (F1/F2) (see Discussion in *Section* 5).

4.1.2 Focused vs Unfocused words. This section presents the results of the comparison between the second syllable and the first syllable in the Focused frame sentence on one hand and in the Unfocused one. In this section, **Frame** refers to the frame sentence,

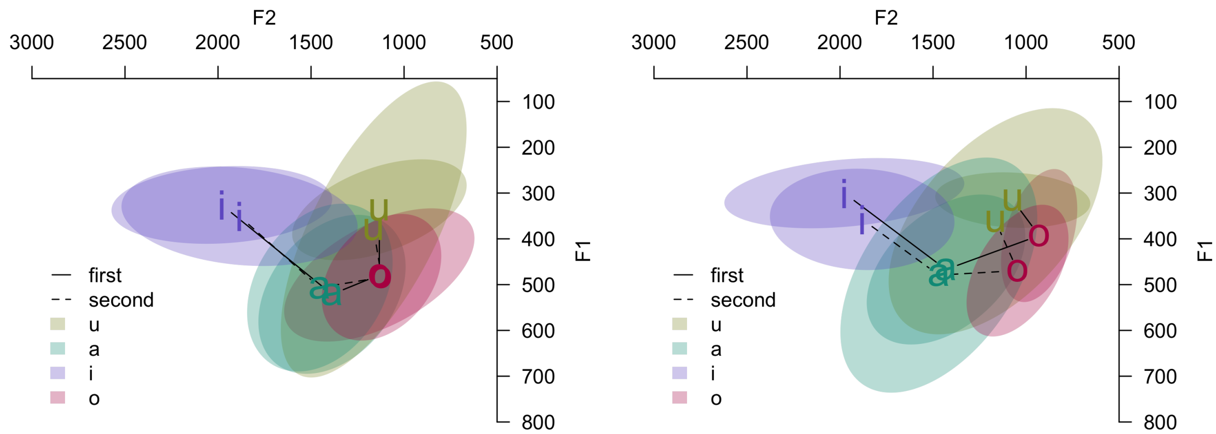


Figure 3. F1/F2 in Hz in the first and second syllable positions of focused words (left) and unfocused (right) words.

thus to whether the target word is *focused* or *unfocused*. As in the previous subsection, *type* represents the natural class of the segment that follows the second syllable (either *obstruent* or *sonorant)*. Likewise, the female participant's intensity data is not included in the statistical results presented below for the same reasons of inconsistancy across repetitions, although her data still patterned like the pooled data.

4.1.2.1 Intensity. The linear mixed effect models with the mean intensity as the dependent variable in the second position yielded a main effect of **frame** ($X^2(1) = 117.55$, p = 2.2e-16) and of **type** ($X^2(1) = 10.32$, p = 0.0013). In other words, intensity in the second syllable was 0.75 +/- 0.06 (roughly 3.5 dB) higher in focused words than in the unfocused words (df = 249.96, t = -12.23, p = 2e-16). Before sonorants, intensity was about 1.3 dB higher than it was before obstruents (df = 21.69986, t = 3.587, p = 0.0017). In the first syllable position, only a main effect of **frame** was observed ($X^2(1) = 175.28$, p = 2.2e-16), such that the first syllable of focused words was in average 0.86 standard units +/- 0.05 (roughly 4 dB) higher in intensity than vowels in the same position in focused words (df = 257.04, t = -15.83, p = 2e-16). As summarized in the *Table 3* below, focused words were found to have a higher intensity than unfocused words, regardless of the position of the syllable considered.

4.1.2.2 Duration. The comparison of the nested models with duration as their dependent variable determined that the second syllable in focused words was significantly different, though marginally so, from the second syllable in unfocused words ($X^2(1) = 9.25$, p = 0.0023), such that vowels in the second syllable of unfocused words were -0.21 standard units +/- 0.07 (roughly 4.2 ms) shorter than those in the second syllable of focused words

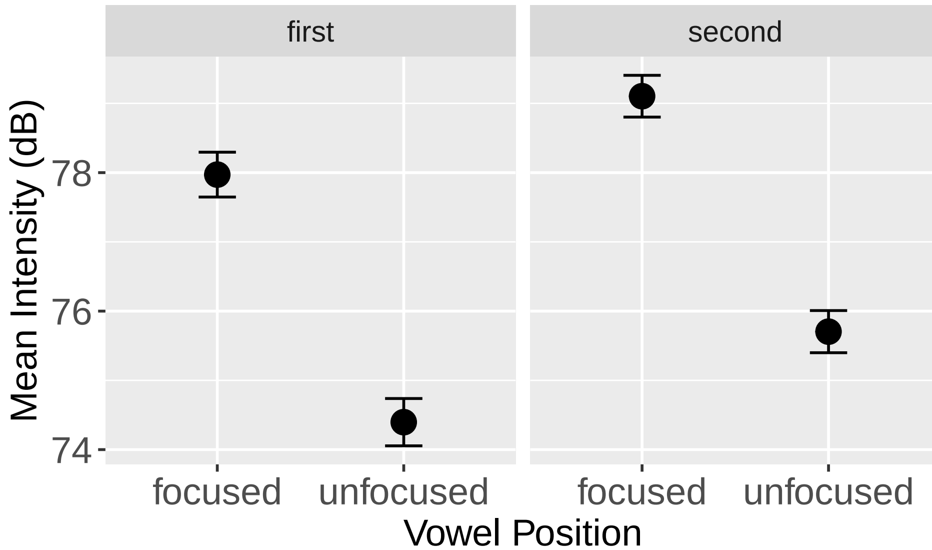


Figure 4. Intensity in dB per syllable position both in the Focused and Unfocused Frame sentences.

(df = 369.88, t = -3.06, p = 0.0024). Unsurprisingly, a fixed effect of *type* was also observed ($X^2(1) = 10.11$, p = 0.0015): vowels were 0.5 standard units +/- 0.14 (about 9.5 ms) longer when followed by a sonorant than when followed by an obstruent (df = 21.83, t = 3.56, p = 0.0018). The same effects of *frame* ($X^2(1) = 31.91$, p = 1.618e-08) and *type* (Xˆ2(1) = 10.37, p = 0.0013) were observed in the first position as well. That is, the first vowel of unfocused words were -0.33 standard units +/- 0.06 (roughly 5.3 ms) shorter than the first vowel of focused words; the difference reached statistical significance (df = 389.86, t = -5.73, p = 2.01e-08). Although the duration differences are under the 10ms JND, the results suggest that focus is not only correlated with higher intensity but also with vowel duration (although the duration difference is under *10ms*) and the effect of focus is not realized on a single syllable but on at least the first two syllables of the word and by extention on the whole word.

4.1.2.3 F1. The F1 models did not show any effects of *frame* in the second syllable for any of the vowels (e.g: for the high vowels [i] and [u], we had $X^2(1) = 0.17$ and p = 0.68). In other words, whether the word was focused or unfocused did not have a bearing on the F1 of all four vowels in the second syllable position of the target words. A marginal effect of *type* was observed with the vowel [a] ($X^2(1) = 2.95$, p = 0.022) and [o] ($X^2(1) = 3.52$, p = 0.06) in the second position. In the first syllable position, however, there was an effect of *frame* ($X^2(1) = 7.82$, p = 0.0052) with [i] and [u], where unfocused high vowels were -0.19 standard units +/- 0.07 (about 23 Hz) higher than their focused counterparts (df =

Table 3

(#tab:loading scripts)Mean intensity (in dB) and standard deviation in the first and second syllable positions for focused and unfocused target words.

frame	position	Intensity	Std.Dev
focused	first	77.97	4.76
focused	second	79.10	4.38
unfocused	first	74.40	4.89
unfocused	second	75.70	4.36

160.51, t = -2.78, p = 0.0061). The same effect was observed for the vowel [a] ($X^2(1) = 32.75$, p = 1.05e-08) and for the vowel [o] ($X^2(1) = 20.13$, p = 7.244e-06), such that [a] and [o] were respectively -0.41 standard units +/- 0.07 (approx. 54 Hz) and -0.62 standard units +/- 0.13 (approx. 86 Hz) higher in the unfocused words than they were in the focused ones. The fact that all four vowels were collectively higher in the unfocused first syllable suggests that focus tends to shift the entire vowel space down in Zarma as shown in *Figure 5*.

In sum, while vowels were significantly lower (less peripheral) in the first syllable of Focused words and less so in the first syllable of Unfocused words, no such effect was observed in the second position. However, as mentioned in subsection 4.1.1.3 above, considering that the second syllable is (by hypothesis) a prominent position in regular unfocused words of Zarma, the additional prominence brought about by focus then has little effect on the F1 of that prominent second position, which would be why no effect of F1 was observed in that position between focused and unfocused words. This result parallels the results (in 4.1.1.3) of the comparison between the second and the first syllable in focused words in the sense that no difference was found between the two positions (except with the vowel [a]). It is still an open question whether this lack of significance in the difference between focused and unfocused second syllable is conditioned by the "ceiling effect" (see Discussion).

4.1.2.4 F2. A pattern similar to the one oberved for F2 in 4.1.1.4 above was seen in the current F2 models, where there were neither an effect of *frame* nor an effect of *type* for all four vowels in the second syllable position. In the first syllable however, there was an effect of *frame* with the back vowels [a] and [o] ($X^2(1) = 6.04$, p = 0.014), such that in unfocused words, back vowels were -0.27 standard units +/- 0.11 (about 108 Hz) more back than in focused words, a difference that proved significant (df = 165.27, t = -2.53, p = 0.012). No such effect was observed for the vowel [a]. Vowel [i] showed an effect of *type* ($X^2(1) = 8.98$, p = 0.0027). However, the model assumption of homoskedasticity did not hold in the vowel [a] data (see Appendix) as the residuals are clustered in groups instead of the expected equal variation of the data.

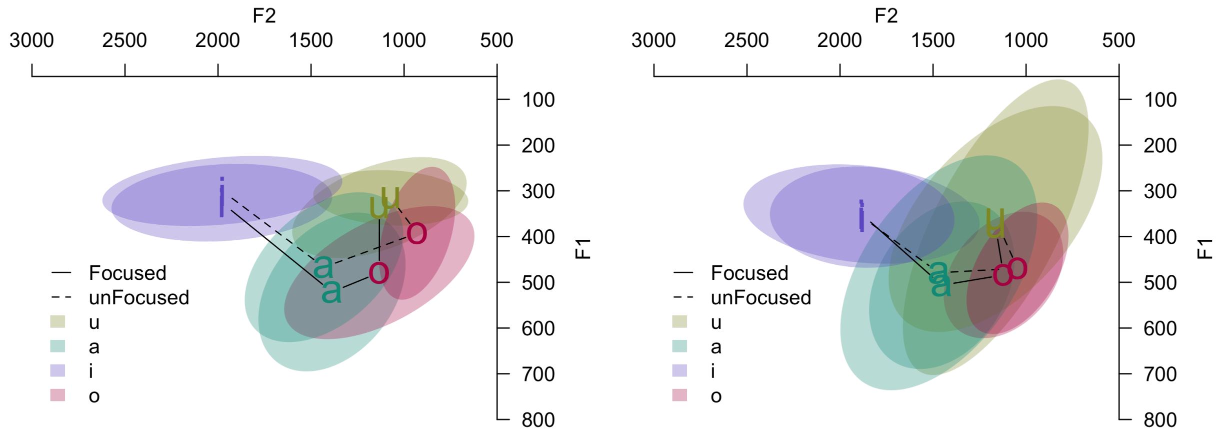


Figure 5. F1/F2 in Hz in the focused and unfocused frame sentences for the first syllable (left) and second syllable (right)

4.1.2.5 F0. Unlike in the comparison between positions in subsection 4.1.1.5 above, where no effect of position was found, there was a main effect of *frame* on all four vowels both in the second syllable on one hand and in the first syllable on the other. The F0 values of the high vowels ([i] and [u]), the low vowel [a] and the mid vowel [o] were significantly lower in the second syllable of unfocused words than they were in the second syllable of focused words: ($X^2(1) = 87.13$, p=2.2e-16), ($X^2(1) = 110.55$, p = 2.2e-16) and ($X^2(1) = 32.2$, p = 1.391e-08) respectively. More specifically, high vowels had -0.80 standard units +/- 0.07 (roughly 32 Hz) lower f0 (pitch) in unfocused than in focused words (df = 119, t = -11.33, p = 2e-16). The f0 of the vowel [a] was -0.73 +/- 0.06 (about 37 Hz) lower in unfocused words than in the focused ones (df = 176, t = -12.24, p = 2e-16); that difference was of -0.77 standard units +/- 0.12 (roughly 41 Hz) for the vowel [o] (df = 51, t = -6.68, p = 1.75e-08). No main effect of *type* was observed for any of the vowels in the second syllable.

In the first syllable, the same main effect of *frame* was observed for all vowels. [i] and [u] have a significantly lower F0 in unfocused words than in focused ones ($X^2(1) = 92.06$, p = 2.2e-16), such that F0 in unfocused words was -0.74 standard units +/- 0.07 (about 35 Hz) lower than in focused words (df = 167, t = -11.07, p = 2e-16). The linear models fitted for the vowel [a] and [o] also yielded an effect of *frame*, ($X^2(1) = 185.87$, p = 2.2e-16) and ($X^2(1) = 34.60$, p = 4.042e-09), respectively. That is, the F0 value for the vowel [a] in the unfocused words was -0.80 standard units +/- 0.05 (roughly 37 Hz) lower than in the focused position (df = 195, t = -17.55, p = 2e-16). That difference was of -0.77 standard units +/- 0.11 (roughly 39 Hz) for the vowel [o] (df = 50, t = -7.06, p = 4.77e-09). *Type* was also significant with the vowel [a] ($X^2(1) = 11.41$, $p = 0.00073$), such that the F0 of vowels that precede sonorants were -0.16 standard units +/- 0.05 (about 8 Hz) lower than that of vowels preceding obstruents ($df = 195$, $t = -3.41$, $p = 0.00078$). The plot in *Figure 6* below summarizes the F0 data in this subsection.

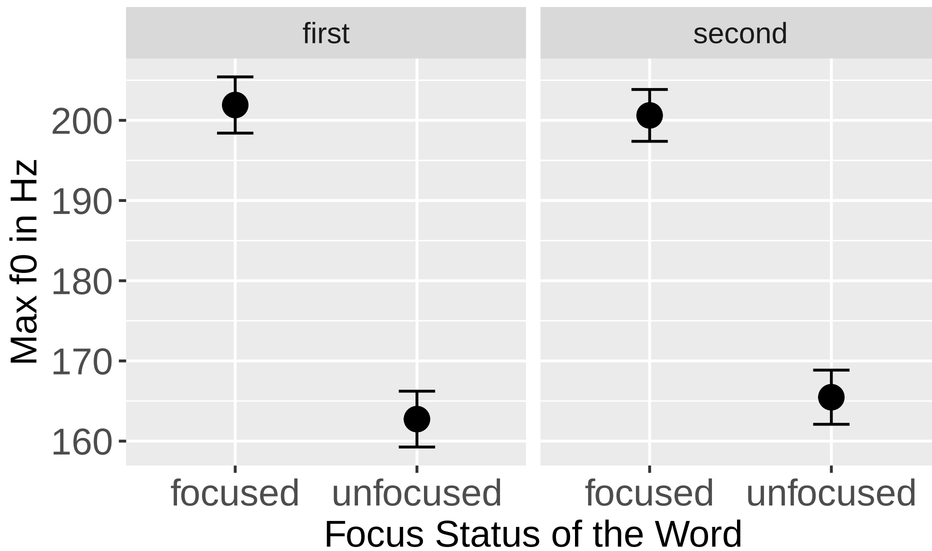


Figure 6. Maximum f0 in Hz as a function of the focus status of the target word and of syllable position.

It is worth noting that the assumption of homoskedasticity did not hold in any of the subsetted **f0** data (i.e vowel-based F0) used in the statistics behind the results presented in subsection 4.1.2.5 above. So, while the vowel-based **f0** results is to be taken with caution, it nonetheless suggests that **f0**, along with **intensity**, is a reliable cue for the realization of focus in Zarma, given the fact that both were consistantly different, and significantly so, between focused and unfocused words regardless of the position of the syllable.

In summary, **intensity**, **f0** and **duration** were significantly higher/longer in both the first and second syllables of focused words than of unfocused words. Although these cues show the effect of focus to varying degrees, they all indicate that focus spans over the entire word and does not fall on one specific syllable in the word. The formant structure on the other hand only shows sensitivity to the focus status of the word in the first syllable: all four vowels had a higher F1 in focused than in unfocused words but only the back vowels [a] and [o] had higher F2 in the first syllable of focused than in that of unfocused words, meaning they were more fronted in the former than in the latter. This behaviour of formants in focused words aligns well with the observed tendancy for vowels to be more centralized in prominent positions and words with the Focus prominence.

4.2 Disyllabic Words

In order to determine whether the same differences found between the first and second syllables of unfocused trisyllabic words hold in disyllabic words, the first (initial) syllable of unfocused disyllabic words was compared to their second (final) syllable. Unlike in the trisyllabic words, the fixed effects of disyllabic words only includes *position* and *gender*, while the random effects have *participant* and *word* as random intercepts (except in the F2 and Duration models where *participant* and *word* were respectively the only random effect variables due to the models' faliure to converge). Only data from [i] and [u] are available, thus presented here[4]. Also, the comparison between the two positions is limited to the unfocused words because, as the trisyllables show, unfocused words show no ceiling effects.

The **intensity** model yielded no effect of *position* ($X^2(1) = 1.03$, $p = 0.31$) and *gender* ($X^2(1) = 0.05$, $p = 0.81$). Despite the lack of statistical significance, the second syllable still has a higher intensity than the first syllable as shown in *Figure 7*. The **Duration** model on the other hand yielded both the effect of *position* ($X^2(1) = 14.05$, $p = 0.00018$) and that of *gender* ($X^2(1) = 5.65$, $p = 0.017$). More specifically, [i] and [u] are about -1.11 standard units +/- 0.29 (roughly 15 ms) shorter in the first than in the second syllable position (df = 105.77, t = -3.9, p = 0.00017) and about -0.71 standard units +/- 0.22 (roughly 10 ms) shorter for the Male speakers (M01 and M02) than for the Female speaker (F01) (df = 100.58, t = -3.16, p = 0.002). No interaction was found between *position* and *gender*.

As far as the vowel formants are concerned, there was a significant effect of vowel *position* on **F1** ($X^2(1) = 11.8$, $p = 0.0006$); that is [i] and [u] were -0.58 standard units +/- 0.16 (approx. 43 Hz) higher in the first position than in the second position (df = 105.17, t = -3.59, p = 0.00051). This effect parallels the effect of *position* on vowel F1 in the first syllable of unfocused trisyllabic words as shown in *Figure 8*.

No effect of gender was observed on F1. With **F2** however, a significant main effect of *gender* was observed for the vowel [i] ($X^2(1) = 11.3$, $p=0.0008$) on one hand and for the vowel [u] ($X^2(1) = 4.32$, $p = 0.04$) on the other. [i] was -0.88 standard units +/- 0.12 (approx. 394 Hz) less fronted for the male speakers than for the female speakers (df = 72, t = -6.99, p = 1.16e-09). Similarly, [u] was -0.52 standard units +/- 0.23 (approx. 247 Hz) less fronted for the male sppeakers than for the female speakers, meaning the female speaker's vowels are overall more fronted than those of the male speakers. *Position* only had a marginal, but not statistically significant, effect on **F2** for the vowel [i] ($X^2(1) = 3.59$, p =

[4]Only the high vowels ([i] and [u]) were present in both the second and final syllables of the disyllabic words available in the two dictionaries consulted after the segmental and prosodic controls were implemented, which is why only the two vowels are used here

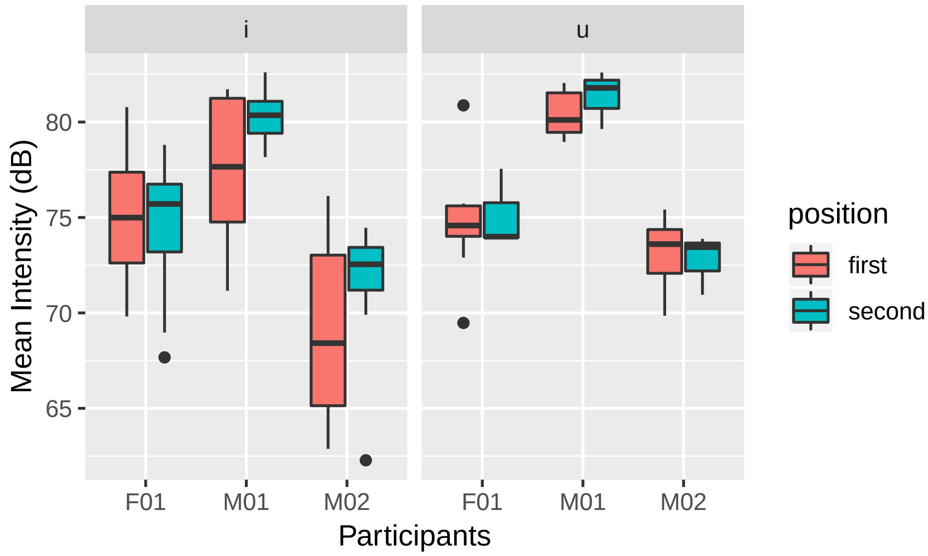


Figure 7. Intensity (in dB) in the first and second syllables of unfocused disyllabic words for all three participants.

0.058), making [i] slightly retracted. Finally, **F0** showed an effect of *gender* ($X^2(1) = 5.92$, $p = 0.015$). In other words, male speakers had -0.67 standard units +/- 0.15 (approx. 121 Hz) lower **F0** than the female speaker (df = 3, t = -4.32, p = 0.022). No effect of *position* was noted on **F0** ($X^2(1) = 0.93$, $p = 0.34$).

To summarize the results of the disyllabic words, **Intensity** showed neither an efffect of *position* nor *gender*, although the final syllable had a higher overall intensity than the second syllable ([u]: F01 = 0.5dB, M01=0.92dB, M02 = -0.35dB; [i]: F01 = -0.3dB, M01 = 2.82dB, M02 = 2.78dB). A source of explanation for the lack of significance in the difference in *Intensity* here may be due to the fact that the second syllable is in word final position, leading speakers to optionally realize the prominence on the first or the second syllable. Hyman (1977), citing Echeverrí & Contreras (1965), reported a similar variation in stress placement on disyllabic words for Araucanian. In this language, stress consistantly falls on the second syllable of *2+*-syllables words, but on disyllabic word, it variably falls on the first or the second syllables.

As far as **Duration** is concerned, it showed an effect of both *position* and *gender*, with longer vowels in the second position on one hand and in the female speaker's data on the other hand. A confounding factor in a reliable interpetation of this effect of position, is the fact that the second syllable coincides with the final position; so the (level of) significance in the length difference could be due to phrase final lengthening. Finally, vowels were also found to be significantly lower in the first than in the second syllable (i.e higher **F1** in the

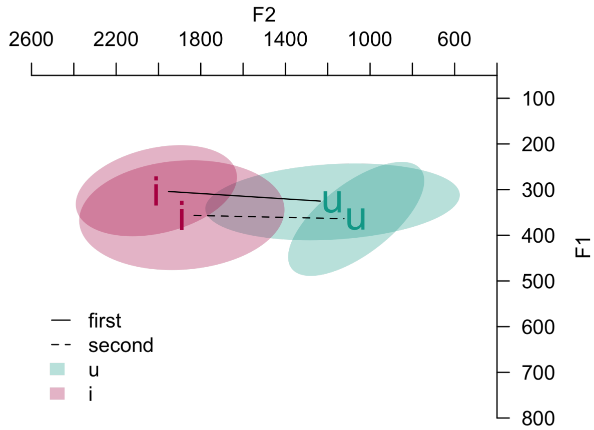


Figure 8. F1/F2 (Hz) in the first and second syllable positions of unfocused disyllabic words.

first syllable). As for **F2**, the female speaker produced significantly more fronted vowels than the male speaker in both positions. **F0** was also significantly higher for the female speaker than for the male speaker, as one would expect.

5 Discussion

In this section, I address the main aspects of the results as well as their meaning for theories of accent in lexical tone languages.

The results support the idea that Zarma has a prominent second (i.e peninitial) syllable, confirming Hypothesis 1, which holds that the second syllable is accented. The prominence is primarily realized with an increased intensity on the peninitial syllable, confirming Hamani (1982)'s claim that Zarma has an intensity accent. However, this increased intensity was mainly noted in words with no special status in the sentence (i.e in unfocused words), contra Hamani's claim that only words with a special syntactic and/or pragmatic status have an accented syllable. In fact, with regard to intensity, the prominence of the second syllable reached statistical significance only in unfocused words. The lack of that significance in focused words can be motivated by appealing to the ceiling effect (Johnson et al. 1993; and citations therein). That is, focused words having higher overall intensity values than unfocused words, the intensity asymmetry between the first and the second syllables becomes less robust - though not neutralized - due to the upper limit constraint on how high the intensity can be, a contraint that makes it possible for the first syllable intensity to raise closer to the second syllable's intensity without the latter being able to move up (compare Focused [1st.Syll = **78**dB, 2nd.Syll = **79.1**dB] to Unfocused

[1st.Syll = 74.4dB, 2nd.Syll = 75.7dB]). An additional support for the ceiling effect comes from the participant M01 who has a higher overall intensity than the other participants and whose first and second syllable intensity values for [a] were almost the same even in unfocused words, although the second syllable is still the the highest.

Intensity is by hypothesis and based on the results discussed here, a correlate of prominence in Zarma. Additionally, however, vowel formants were also found to be significantly different in prominent syllables, which seems to suggest that prominence in Zarma is expressed with multiple cues. More specifically, the prominent syllable (i.e the second syllable) was found to have more centralized vowels[5] than the non-prominent one (i.e the first syllable). This centralization effect, however, radically departs from the cross-linguististic tendancy for prominent positions to correlate with more peripheral vowels and/or fortified consonants (Fry, 1955; Crosswhite, 2004); though there are a few cases, like in Mokša Mordvin, where consonants optionally lenite in the onset position of stressed syllables that are not word-initial (Vaysman, 2009). Gordon (2011) termed this phenomenon the 'exceptional lenition', in order to highlight the rarety of weaker segmental materials in prominent positions. With the centralization of vowels in the peninitial syllable and in focused words, Zarma may be analyzed as examplifying an exceptional case of vowel centralization in prominent syllables.

A competing analysis to the one above is that the relative peripherality of the first syllable can be due to their being in a word-initial position. This analysis suggests that vowels will be less and less peripheral as one moves from the beginning of the word to its end; thus predicting that the third vowel in trisyllabic words will be less peripheral than the second one. Although we only examined the first two syllables of trisyllabic words in the current study, thus unable to verify that prediction, the behavior of formants in focused positions suggests that the centralization analysis is the best anlysis for Zarma. As shown in the *Figure 3*, in the second syllable, the Focused ones were significantly less peripheral than the Unfocused ones, suggesting that all prominent syllables in zarma (either 'accented' or focused) have their vowels centralized. This behavior of vowels in prominent positions in Zarma challenges the cross-linguistic generalization that prominent positions go with more peripheral vowels.

As far as F0 and Duration are concerned, the fact that the former does not correlate

[5]Roughly put, higher F1 (for high vowels) and F2 (for [u] and [i]) in prominent positions. It is important to note that this is not a shift in the vowel space, as shown in the plots. It is the case that high vowels are lowered and back vowels, usually to the exception of [a], are fronted, etc. (see the results section for more details).

with accent and the latter does so very marginally is best analyzed as an effect of the Functional Load Hypothesis (FLH) on acoustic cues that are already employed for other phonological contrasts in the language (Vogel et al. 2015; Remijsen 2002; and others). In other words, since F0 is already used for lexical tone and duration, for length contrasts, they are not employed in the realization of accent. Despite this cross-linguistically robust prediction of the FLH, the Zarma results showed that F0 and Intensity are both reliable cues for focus, in addition to being cues for lexical tone and word-level accent, respectively. This raises the question of why is focus acoustically realized (in addition to having a higher intensity) with a higher pitch (Cooper et al., 1985; Vogel et al., 2015) assuming that the functional load hypothesis holds in Zarma?

To answer this question I propose that when the FLH is active in a language, it only holds of cues associated with phonological phenomena that operate at the Phonological Phrase level or on lower levels of the prosodic hierarchy[6] (i.e. Syllable, Foot and Prosodic Word) but not at the Phonological Phrase, Intonational Phrase and the Utterance levels, unless it is already used on those levels for other contrasts as well. As such, because the word-level "accent", the lexical tone and the contrastive length reside on these lower levels of the prosodic hierarchy, their respective cues are unavailable for usage by other prosodic phenomena on those lower levels. At the Phonological Phrase, Intonational Phrase or Utterance level, however, the FLH resets to allow those cues to be used once again, either through an extention of the range (i.e. upward shift of the normal F0 or Intensity range) used on lower levels or a re-engeneering of its shape (e.g: different overall pitch contour, as in Hungarian). Contrastive focus is well established as an Intonational Phrase phenomenon (Vogel, 1990; Ladd, 1996; Gussenhoven et al., 1997; Frota, 2014), which would be why it is not affected by the FLH whose effects reset at the Intonational Phrase level in Zarma.

In fact, Vogel et al. (2015) had similar results for their study of prominence in Hungarian. Their findings showed that F0 was the strongest cue for both stress and focus, contra the prediction of the FLH that stress and focus should have different cues. Although Vogel et al. found a different pitch pattern on focused words, in addition to a higher overall F0, they rejected the strong form of their Functional Load Hypothesis, which predicted that the same cue can not be used to express more than one type of contrast. As it turns out, the limits of this strong form of the FLH are more clearly seen when languages with a (nearly) exhausted acoustic space, like Zarma, are studied. So, although it is possible that the resetting of the FLH in Zarma is forced by the exhaustion of the acoustic cue space[7], the

[6] The version of the Prosodic Hierarchy assumed here is that of Itô & Mester (2012).

[7] Zarma has contrastive length (Duration), lexical tone (F0), geminates (Stop closure duration), and as

Hungarian results - where the acoustic space is not exhausted due to the fact that Intensty is not used for any contrast whatsoever in the language - compels a revision of the FLH. I propose that the FLH be relativized to each of the three, independently motivated, categories of immidiately dominating levels of the prosodic hierarchy in the following way:

Relativized Functional Load Hypothesis:

Let $\Sigma 1$*,* $\Sigma 2$ *and* $\Sigma 3$ *be omnibus levels (categories) of the prosodic hierarchy, where* $\Sigma 1$ *includes the Mora, the Syllable and the Foot;* $\Sigma 2$*, the Prosodic Word and* $\Sigma 3$*, the Phonological Phrase, the Intonational Phrase and the Utterance. If a language uses a cue X to express a phonological contrast that pertains to a level in* $\Sigma 1$*, that cue is made available for another phonological contrast within* $\Sigma 2$ *and* $\Sigma 3$*, but not within* $\Sigma 1$*. The same principle holds for* $\Sigma 2$ *and* $\Sigma 3$*.*

The first category is the equivalent of the *Rhythmic Categories*, the third category is that of the *Interface Categories* (c.f. **?** and citations therein) and the second category is a category with a single level, the Prosodic Word (*Pre-interface Category*). Assuming these three independently motivated categories in the Prosodic Hierachy, while allowing the FLH to (optionally) reset in each, guarantees that the FLH will neither be too strong nor too weak and will accurately predict the Hungarian and Zarma results, among others.

Returning to the notion of *accent* in Zarma, while the word-level prominence was casually referred to as "accent" in the discussion above, following Hamani, there are theoretical grounds to believing that the term *accent* is in fact the right terminology for the word-level prominence in Zarma. According to Downing (2010), an *accent* is a prominence asymmetry realized with a combination of phonetic properties, and just like stress, it is Culminative, Demarcative and Obligatory (Hyman, 1977, 2006; Odden, 1988; van der Hulst, 1999). Prominence in Zarma fits this definition and presents these properties because (1) in addition to Intensity, there is at least one other cue for the word-level prominence (i.e. vowel centralization), and (2) the prominence is *Culminative* by virtue of correlating with the highest intensity in the domain of the word, *Demarcative* in that it is realized near the left edge of the word (*peninitial*, as motivated by the disyllabic words' data) and *Obligatory* because it occurs on all words, assuming the generalizations in the two and three syllable words discussed here extend to longer words. The word-level prominence in Zarma is therefore a true word-level accent in the sense of Downing (2010).

Additionally, although Downing's definition of accent is in broad terms (also see van der Hulst 1999) because her use of the "domain" of the accent is not restricted[8], for Zarma

shown in the current study, a word-level accent (Intensity).

[8]According to the definition, the domain can be the word, the phrase, etc.

to have a word-level accent, then, is indistinguishable from having a stress system. In that sense, Zarma is also a stress-accent language. Previous work have shown that it is not unusual for lexical tone languages to have stress: Dagbani and Mò:ré (Niger-Congo; respectively by Olawsky 1999, and Cann 1976, cited in Downing 2010) and Ma'ya (Remijsen, 2002) have been reported to have stress-accent independently of lexical tone. However, just like in Dagbani and Mò:reé, but unlike in Ma'ya, Zarma accent is not contrastive[9]. This lack of contrastiveness puts Zarma accent among what Hyman (1977) refers to as "grammatical accent", which he argued to be the accent system from which languages known as having a (contrastive) stress-accent system evolved from. This raises the question of whether the prosody of Zarma is undergoing a transition towards becoming a stress-accent language as seen for the Western and Northern Sonray-Zarma languages (Nicolaï and Zima, 1997; Heath, 1999; Nicolaï, 1980). Other questions this raises are as follows: what does it mean for a lexical tone language to have an accent? Are there phonological phenomena that priviledge the accented syllable?

In the section below, I will argue that Zarma accent is the realization of the head of a single left-aligned moraic iamb and briefly sketch how this analysis can give a unified account for a number of supposedly unrelated phonological phenomena in the language.

6 Implications

The word-level accent for which phonetic evidence was found in the experiment above can be analyzed as marking the head of a left-aligned moraic iamb in Zarma. Evidence for the existence of this foot structure in Zarma (see Akinlabi and Urua 2003 for a similar foot-based analysis for Ibibio) can be motivated independently based on the restriction on tone plateau on one hand and on the other from the distributional frequency of consonantal phonemes in the language. For the former, the analysis below will just be limited to a scketch and will not get into the specifics of how to actually implement such a proposal to account for issues in tone association in Zarma.

6.1 Foot-sensitive Tone Plateau in Zarma

In Zarma, when disyllabic words have a falling tone on their first syllable, their second syllable can only bear a H tone while disyllabic words with a rising tone on their first syllable can only have a L tone on their second syllable as shown in (4).

[9] Zarma accent is not contrastive in the sense that there are no minimal pairs that can be distinguished only based on the accent, altgough Hamani (1982) reported a near minimal pair with háw ˈbî '*black* cow' and ˈ**háw-bî** 'buffalo', where the accent is what distinguishes the two.

(4) Disyllabic Words

a. hàágù “to grill”
b. kàágà “chair”
c. sôrbó “horse”
d. sààrǎj “grave”
e. ɟàrâw “charge”
f. tàbâ “taste”
g. zàmû “genealogy”

(5) Trisyllabic Words

a. tàsàbâ “rosary”
b. gùnàkô “seer”
c. gòrkàsîn “companion, neighbor”
d. gùŋùrí “egg”
e. hánnándì “to clean”
f. búzùgú “stomac”
g. bógótì “bucket”

This restriction, was argued by Oumarou Yaro (1993) to holds even with trisyllabic words, as in (5), due to a prohibition of LHH and HLL sequences on the surface, while HHL and LLH are attested. Oumarou Yaro analyzed these restrictions by positing that closed and long syllables are bimoraic[10] and proposed an edge-in association (Yip, 1989; Chan, 1991), where the tones of a word melody associate in the following fashion: say we have an HL melody for a trisyllabic word (e.g: bogoti “bucket”), the first tone H attaches to the first syllable, the second tone L attaches to the last syllable, and the middle syllable copies the preceding tone. This account covers more empirical ground than the left-to-right or right-to-left association analyses because none of the two can capture (5-a) and (5-b), assuming that OCP prohibits sequences of the same tones underlyingly. However, while the edge-in analysis can capture the data in (4) and (5), it can not account for the data in (6) below without making additional assumptions[11].

[10] Oumarou Yaro (1993) referred to long and closed syllables as containing ’two other syllables’, the second of which is deficient. Our understanding of this account is that each of his two contained syllables translate to a mora (in moraic theory), with the exception that the morae have different importance.

[11] The fact that we can only have contours in bimoraic syllables suggest that the mora is the TBU. Crucially, under this analysis, the word-final shortening will be argued to occur after tone assignment. This will explain why the final short vowel can bear contour tones.

(6) (a), (b), (c) = /LHL/, (d) = /HLH/

a. lààsáábù "to think"

b. màsííbà "misfortune"

c. zùnúúbù "sin"

d. méèhámní "mustach"

To circumvent the challenges presented in (6), Oumarou Yaro assummed that heavy syllables have a deficient second mora, and that the edge-in association targets the non-deficient first mora and when there are no more tones to be associated, each remaining mora copies the tone of the preceding mora. With that extra assumption, the data in (6) is accounted for. However, this move has a number of issues. Not only does it distinguish between the morae of heavy syllables based on their linear order in a purely stipulative way, but also it implies that when there are as many tones in the melody as there are moraic TBUs, the association does not distinguish between deficient and non deficient morae but it is only when there are less tones than available TBUs that the association shows sensitivity to deficient morae (compare lààsáábù "to think" with kàmbàgàábà "span", both supposedly /LHL/ in their underlying form). These issues aside, the edge-in analysis, with the additional assumption, still does not account for the data in (7), where the edge-in account predicts the surface form [ìtààcí], rather than the attested [ìtáácí] for the word "four", assuming an underlying /LH/ melody. Also, note that the data in (7) suggests that a ban on HLL(L) and LHH(H) is not the right generalization for Zarma.

(7) (a),(b) = /LH/,(c), (d) = /HL/

a. ndúɲɲá "world"

b. ìtáácí "four"

c. íbèèrì "big"

d. íbòòbò "(?)"

Since none of the previous tone association mechanisms fully capture the observed data in Zarma, it is possible that the right generalization and tone association pattern is one that makes explicit reference to foot structure. That is, positing the metrical foot as the domain of tone association as argued for Maninka by Bamba (1991)[12]. As such the apparent ban on HLL and LHH can be re-analyzed as a ban on tone plateaux that span across foot

[12]The Zarma data can also be made consistent with a *tonal foot* analysis as argued by Leben (2002) for Hausa and Bambara, but crucially, this may not necessarily correspond with the metrical foot analysis adopted here.

boundaries, in a left-to-right association fashion, while (a) the foot itself does not have a plateau for its entire length and when (b) the melody to be associated has 1+ tones in it. (8) shows how (7) will be footed. We assume that the vowel [i] is extrametrical due to its optionality (e.g: táácí "four", bèèrì "big") and to the fact that its tone is predictable in many cases.

(8) (a), (b) = /LH/; (c), (d) = /HL/

a. ndúɲɲá "world"
b. ì(táá)cí "four"
c. í(bèè)rì "big"
d. í(bòò)bò "(?)"

In (8), there is a single H tone for the words which is why we get a plateau on the right (as in words like fúfúlé "heat"). This proposal is just a sketch and needs to be refined further.

6.2 Consonant Distributional Frequency

Another potential source of evidence for the foot structure is in the frequency of consonants distribution. While all consonants occur in all onset positions, their frequency in different onset positions, i.e in $C_1VC_2VC_3V$, is systematically different. In other words, certain consonant classes have a higher frequency in some onset positions than in others, although obstruents in general and stops in particular have a wide overall distribution.

6.2.1 Disyllabic words. In disyllabic words, obstruents are more frequent in C1 position than in C2 position while sonorants have the reversed frequency in the two positions as shown in *Figure 9*. Glides patterned like obstruents rather than sonorants. In these disyllabic words, C1 position coincides with the left edge of a foot and is thus predicted to have a 'full consonant contrasts' than C2 (Downing, 2010; Harris, 1999). In the current case, the notion of full contrasts is reinterpreted in terms of frequency since strict full consonant contrasts do not hold for Zarma due to the fact that all three onset positions examined here can accommodate all twenty consonantal phonemes[13]. About 83% of consonants in C1 position are obstruents while only about 50% of them in C2 position are obstruents. Sonorants on the other hand are thrice as much in C2 position (approx. 36%) than they are in C1 position (approx. 12%). While one may argue that the high frequency of sonorants in C2 position is conditioned by the intervocalic environment, the sonorant frequency in

[13]The only exception is /ŋ/, which was not found in any of the three onset positions in all 602 trisyllabic words examined.

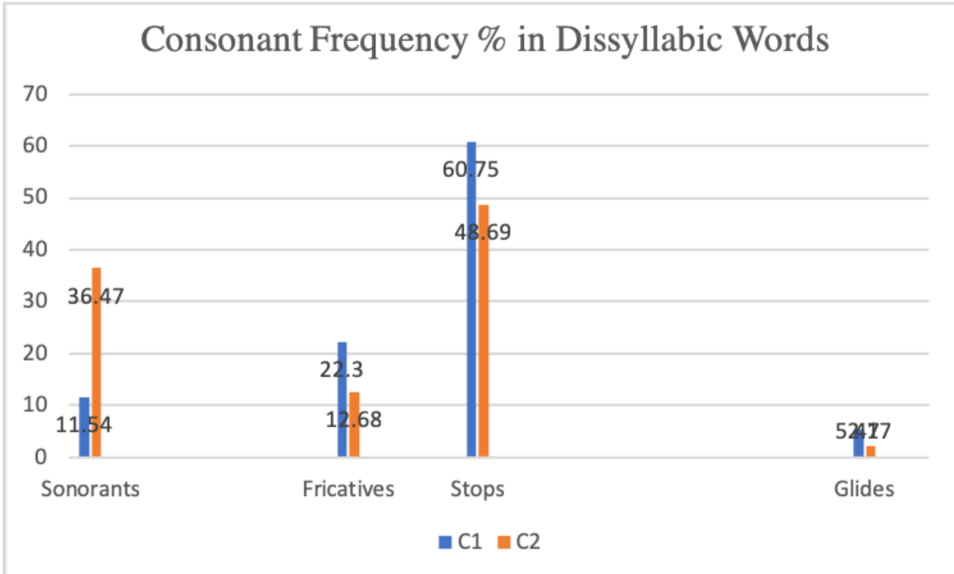


Figure 9. Consonant class frequency percentage in C1 and C2 positions from 1,933 disyllabic words.

trissylabic words shown in *Figure 10* indicates otherwise.

6.2.2 Trisyllabic words. Sonorants are more frequent in C3 position than in C2, unlike what one would expect if the intervocalic environment was the (only) reason for a higher frequency of sonorants in the C2 position of disyllabic words. That is, one would expect the C2 and C3 positions of trisyllabic words to have similar sonorant frequencies, which is not the case: only about 25% of consonants in C2 position are sonorants while there are twice as much sonorants in C3 position[14] than there are in C2 position. I propose to analyze the increased frequency of sonorants in C2 position of disyllabic words and C3 positions of trisyllabic words on one hand and the decreased frequency of obstruents in those same positions on the other hand as being conditioned by a foot structure. In other words, foot-medial positions (C2) favor sonorants, while the left edge of the foot (C1) favors obstruents. As for C3, it will be unfooted if we assume there is a single foot in Zarma words. As such, C3 will be in an even weaker position, which would further explain why there are more sonorant in that position (almost half of the sonorants found in Zarma are in C3 position). For a more comprehensive and convincing evidence of the foot structure in Zarma, segmental phenomena as well as tonal processes will need to more closely examined, a entreprise we leave for future work.

[14]The size of the trisyllabic words is admittedly smaller (602 words) than the disyllabic words (1,933) but those are the only available 2 and 3 syllable words in the most comprehensive Zarma dictionaries available (that is the 'Dictionnaire Zarma-Français' by Bernard and Kaba (1994).

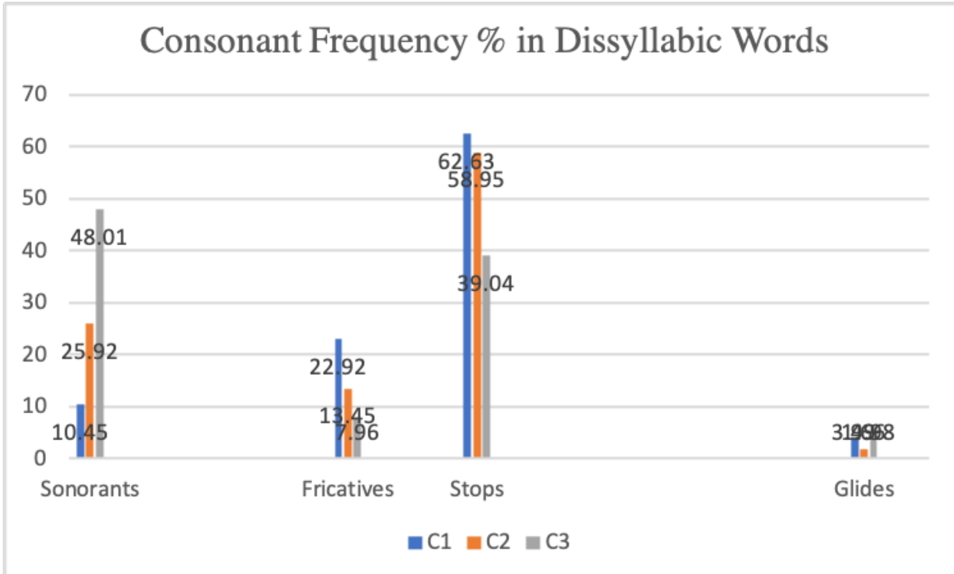


Figure 10. Consonant class frequency percentage in C1, C2 and C3 positions from 602 disyllabic words.

7 Conclusion

This paper provided an acoustic evidence for the existence of a word-level accent in Zarma, partially confirming a previous impressionistic desciption by Hamani (1982). The results of the production study examined trisyllabic and disyllabic words in Zarma. They indicated that there is a peninitial prominence, realized reliably with an increased intensity on all Zarma word, regardless of their syntactic status (focused or unfocused alike). This prominence was shown to present accentual properties in the sense of Downing (2010). As to the more general question of how accent is realized in a language with lexical tone, length contrast and geminate consonants, these findings suggest that the language relies on the available acoustic cue, which is consistent with the Functional Load Hypothesis. However, focus is found to be realized with an increased F0 and Intensity, meaning the same cues are used at both the Prosodic Word level (tone and accent) and at the Intonational Phrase level (focus). We proposed that the strong form of the Functional Load Hypothesis be relativized to partitions of levels on the prosodic hierarchy, such that cues that are used exclusively on lower levels are made available to be used on higher levels too and vice versa; a proposal that is constistent with recent findings in other languages.

The acoustic evidence for a word-level accent found in this paper raises the question of what it means for a language like Zarma to have an accent, given the fact that the accent in question is not contrastive (at least not in the way stress-accent is, in English for example)?

Are there independent phonological phenomena that are sensitive to that accent? To answer these questions, we briefly sketched a moraic binary metrical foot analysis for the accent, where the accented syllable correlates with the head of the moraic iamb, although the full analysis on the motivation for the existence of a foot structure in Zarma is still to be worked out. More work is in order on that end as well as a perceptual study to determine whether this accent can be perceived by speakers of Zarma, despite the fact that the intensity difference between an accented and an unaccented syllable passsed the Just Noticeable Difference threshold.

Appendix 1:

The target words used in this experiment are in the table below. The top table contains the three syllable words and the bottom one, the two syllable words.

Trisyllabic target words:

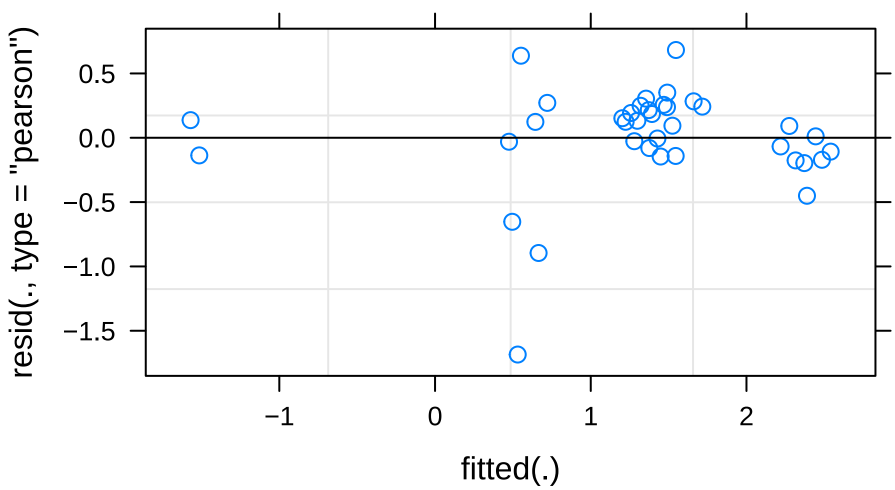
Word ID	Target Word	Meaning
S1	fúfúlé	heat
S2	kútúrú	type of plant
S3	sásáré	to align
S4	gútúláy	ungratefulness
O1	tútúbú	to grind
S5	tígíná	callus
S6	bákáráy	compassion
S7	kátárù	judiciary
S8	kókóró	weave
O2	kákátù	to sacrify
S9	sókólò	attic
S10	tígírí	to stand for long
S11	tátálí	to cajole
O3	kákásé	to have good taste
S12	cítílá	millstone
S13	kásámá	scabies
O4	súkútú	forehead
O5	tákúbâ	sword
O6	cákátí	trash
O7	kúkúsí	to clean
O8	kútíjí	threshold
O9	kókóbé	undust
S14	fátámá	chicken louse
S15	sákálí	wolf

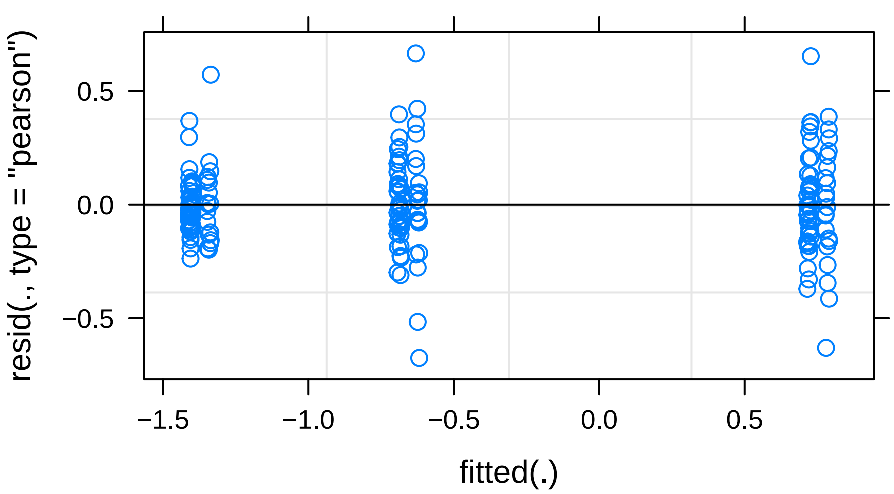
Disyllabic target words:

Word ID	Target word	Meaning
1	kúkú	calao
2	kútí	to punch
3	túrú	headdress
4	tísí	music instr.
5	kúrí	blood
6	tálí	green doum nuts
7	tírí	gizzard

Appendix 2:

The plots below show how the model assumption of homoskedasticity for the **F2** of the vowel [i] in the *first* position (Top) did not hold as the residuals are clustered instead of equally varying. Same for the assumption of homoskedasticity for the **F0** of the vowel [a] in the *unfocused* frame (Bottom).





References

Alidou, O. (1988). Tasawaq d'In-Gall: Esquisse linguistique d'une langue dite 'mixte'. *Université de Niamey*, MA Thesis.

Akinlabi, A. and Liberman, M. (1995). On the phonetic interpretation of the yoruba tonal system. In *Proceedings of the International Congress of Phonetic Sciences*, pages 42–45.

Akinlabi, A.; and Urua, E. (2003). Foot structure in the ibibio verb. *Journal of African languages and linguistics,*, 24(2):119–160

Bamba, M. (1991). De l'interaction entre tons et accent. *Doctoral dissertation, Université du Quebec à Montréal.*

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck, G., Eigen, C., and Rcpp, L. (2015). Package 'lme4'. *Convergence*, 12(1).

Bernard, Y. and Kaba, M. W. (1994). *Dictionnaire zarma-français (République du Niger).* Agence de coopération culturelle et technique.

Chan, M. K. M. (1991). Contour-tone spreading and tone sandhi in danyang chinese. *Phonology*, 8(02):237–259.

Cooper, W. E., Eady, S. J., and Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question–answer contexts. *The Journal of the Acoustical Society of America*, 77(6):2142–2156.

Crosswhite, K. (2004). Vowel reduction. *Phonetically based phonology*, pages 191–231.

de Lacy, P. (2002). The interaction of tone and stress in optimality theory. *Phonology*, 19(1):1–32.

Downing, L. J. (1990). Local and metrical tone shift in nguni. *Studies in African Linguistics*, 21(3):261.

Downing, L. J. (2010). Accent in african languages. *A survey of word accentual patterns in the languages of the world*, pages 381–427.

Downing, L. J. (2013). What African languages tell us about accent typology. *Universitätsbibliothek Johann Christian Senckenberg.*

Frota, S. (2014). The intonational phonology of european portuguese *. *Prosodic Typology II*, pages 6–42.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America*, 27(4):765–768.

Gordon, M. (2011). Stress: Phonotactic and phonetic evidence. *The Blackwell companion to phonology*, pages 1–25.

Gordon, M. and Roettger, T. (2017). Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard*, De Gruyter: 3(1).

Gussenhoven, C., Repp, B. H., Rietveld, A., Rump, H. H., and Terken, J. (1997). The perceptual prominence of fundamental frequency peaks. *The Journal of the Acoustical Society of America*, 102(5):3009–3022.

Hamani, A. (1982). *La structure grammaticale du zarma(essai de systématisation).* PhD thesis.

Hamani, A. (1982b). De l'oralité à l'écriture. *Le zarma s' écrit aussi, Niamey, INDRAP*, (158).

Harris, J. D. (1963). Loudness discrimination. *Journal of Speech & Hearing Disorders. Monograph Supplement.*

Harris, J. W. (1999). Nasal depalatalization no, morphological wellformedness sí; the structure of spanish word classes. *MIT Working Papers in Linguistics*, 33:47–82.

Heath, J. (1999). A grammar of Koyraboro (Koroboro) Senni: the songhay of Gao, Mali. *Köppe.*

Hyman, L. (2006). Word-prosodic typology. *Cambridge University Press*, Phonology 23(2):225–257.

Hyman, L. (1977). On the nature of linguistic stress. *Studies in stress and accent*, 4:37–82.

Itô, J & Mester, A. (2012). Recursive prosodic phrasing in Japanese. *Prosody matters: Essays in honor of Elisabeth Selkirk*, Equinox London: 280–303.

Johnson, K., Ladefoged, P., and Lindau, M. (1993). Individual differences in vowel production. *The Journal of the Acoustical Society of America*, 94(2):701–714.

Klatt, D. H. (1976). Linguistic uses of segmental duration in english: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, 59(5):1208–1221.

Kossman, M. (2007). Grammatical borrowing in Tasawaq. *EMPIRICAL APPROACHES TO LANGUAGE TYPOLOGY* Mouton de Gruyter: 38, 75p.

Ladd, D. R. (1996). Intonational phonology. *Cambridge: Cambridge University Press.*

Leben, W. R. (2002). Tonal feet. *Proceedings, typology of African prosodic systems*, 1:27–40.

Lennes, M. (2011). Spect-the speech corpus toolkit for praat. *former title: Mietta's scripts for the Praat program.*

Liberman, P. (1960). Some acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America*, ASA:32(4), pages 451–454.

Nicolaï, R. (1980). Le songhay septentrional: les parlers du groupe nomade. *Bulletin de l'Institut Fondamental d'Afrique Noire 41b2*, pages 304–70.

Nicolaï, R. (1997). Songhay. *Lincom Europa*

Odden, D. (1988). Predictable tone systems in Bantu. In Harry van der Hulst and Norval Smith (eds.), *Autosegmental Studies on Pitch Accent*, Dordrecht: Foris, 225-251.

O'Shaughnessy, D.; & Allen, J. (1983). Linguistic modality effects on fundamental frequency in speech. *The Journal of the Acoustical Society of America*, 74(4):1155–1171.

Oumarou Yaro, B. (1993). Elements de description du Zarma (Niger). *These de Doctorat.*

Peterson, G. E. and Lehiste, I. (1960). Duration of syllable nuclei in english. *The Journal of the Acoustical Society of America*, 32(6):693–703.

Pinheiro, J., Bates, D., DebRoy, S., and Sarkar, D. (2017). the r development core team. 2017. nlme: linear and nonlinear mixed effects models. r package version 3.1–131.

Remijsen, B. (2002). Lexically contrastive stress accent and lexical tone in ma'ya. *Laboratory phonology*, 7:585–614.

Shih, S.-h. (2016). Sonority-driven stress does not exist. In *Proceedings of the Annual Meetings on Phonology*, volume 3.

Sudlow, D. (2001). *The Tamasheq of North-East Burkina Faso: notes on grammar and syntax including a key vocabulary*, volume 1. R. Köppe Verlag.

Tersis, N. (1972). Le zarma: etude du parler djerma de dosso.

Turk, A., Nakai, S., and Sugahara, M. (2006). Acoustic segment durations in prosodic research: A practical guide. *Methods in empirical prosody research*, 3:1–28.

van der Hulst, H. (1999). Word accent. *Word Prosodic Systems in the Languages of Europe*, Berlin, Mouton de Gruyter, 3-115.

van der Hulst, H. (2010). Word accent: Terms, typologies and theories. *A survey of word accentual patterns in the languages of the world*, van der Hulst, Harry and Goedemans, Rob and van Zanten, Ellen, Walter de Gruyter, 3-53.

Vaysman, O. (2009). Segmental alternations and metrical theory. *Doctoral dissertation*.

Vogel, I., Athanasopoulou, A., and Pincus, N. (2015). Acoustic properties of prominence in hungarian and the functional load hypothesis. *Approaches to Hungarian*, 14:267–292.

Vogel, I.; & Kenesei, I. (1990). Syntax and semantics in phonology. *The phonology-syntax connection*, (339-363).

Wolff, E. and Alidou, O. (2001). On the Non-Linear Ancestry of Tasawaq (Niger), or: How 'Mixed' Can a Language Be?. In *Sprache und Geschichte in Afrika*, 16:42–45.

Yip, M. (1989). Contour tones. *Phonology*, 6(01):149–174.