



The production of Norwegian lexical pitch accents by multilingual non-native speakers

Journal:	<i>International Journal of Bilingualism</i>
Manuscript ID	IJB-16-0042.R2
Manuscript Type:	Original Article
Keywords:	Non-native tones, Norwegian, Swahili, Lingala, Multilingualism, Spontaneous speech, Second language acquisition
Abstract:	<p>Aim and objectives/purpose/research questions: The aim of this study is to examine the extent to which multilingual second language speakers of Norwegian manage to produce lexical pitch accents (L*H⁻ or H*LH⁻) as expected in natural spontaneous speech. Using native speech as a reference, we analyze realizations of multilingual speakers whose respective dominant languages are Lingala, a lexical tone language, and Swahili, a non-tonal language with fixed stress, and hypothesize that this difference might be reflected in the speakers' competence in the East Norwegian tone system.</p> <p>Design/methodology/approach: We examined a corpus of spontaneous speech produced by eight L2 speakers and two native speakers of East Norwegian. Acoustic analysis was performed to collect fundamental frequency (f0) contours of 60 accentual phrases per speaker.</p> <p>Data and analysis: For LH and HLH tonal patterns, measuring points were defined for quantitative evaluation of f0 values. Relevant aspects investigated were (a) pattern consistency, (b) f0 dynamic range and (c) rate of f0 change. Pattern consistency data were statistically evaluated using chi-square. Dynamic range and rate of f0 change data were explored through to linear mixed effects models.</p> <p>Findings/conclusions: We found no really substantial differences between the speaker groups in the parameters we examined, neither between the L2 speakers and the Norwegian natives nor between the Lingala and Swahili speakers.</p> <p>Originality and significance/implications: This study is a contribution to the scarcely explored area of L2 acquisition of tones. It is concerned with languages that have received little or no attention in the field: Norwegian, Lingala and Swahili. Participants are multilinguals who have extensive language learning experience. Further, the study is based on a corpus of spontaneous speech.</p>

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



SCHOLARONE™
Manuscripts

For Peer Review

THE PRODUCTION OF NORWEGIAN LEXICAL PITCH ACCENTS BY MULTILINGUAL NON-NATIVE SPEAKERS

Introduction

It is generally assumed that tones and intonation are difficult to acquire in a second language (L2), here defined as a language that is acquired after one or more other languages. It has even been claimed that intonational characteristics are among the last properties of a phonological system learners are able to master (Mennen, 2015). Possibly due to methodological issues, second language acquisition of tones and intonation is an understudied field and no model is available that generates predictions regarding the difficulties learners with different linguistic backgrounds will encounter when acquiring new tonal systems (Broselow & Kang, 2013; Mennen, 2004, 2015). In this article, we examine tonal patterns in spontaneous speech produced by non-native speakers of East Norwegian, in which every non-clitic word is associated with one of two lexical pitch accents (/L*H⁻/ and /H*LH⁻/).

The present speakers are eight multilingual Congolese immigrants in Norway, who have acquired and used several languages in different periods of their lives. Many learners of European languages are migrants from areas where multilingualism is common. Studying the linguistic competence of this kind of speakers contributes to get a picture of the advantages or challenges speakers with complex language backgrounds might encounter in acquiring a new language. The speakers can be grouped into two categories according their dominant languages. For four of them it is Swahili, while the other four are Lingala-dominant. Different from Lingala, Swahili has no lexical tones, a difference we hypothesized might be reflected in the speakers' competence in Norwegian, specifically the extent to which they master the Norwegian tone system.

Background

Among the factors that determine how successful a learner is in acquiring L2 phonology (see Piske, MacKay & Flege, 2001 for a review) *linguistic experience* is undoubtedly important (So & Best, 2010). Influential models of perception of second language segments, such as Flege's (1995) Speech Learning Model (SLM) and the Perceptual Assimilation Model (PAM; Best & Tyler, 2007), are, for instance, based on the assumption that the L1 is decisive in how learners perceive and categorize non-native contrasts.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

At present, studies on non-native production of tones are relatively scarce (Hao, 2012). To the best of our knowledge, there are no publications involving L2 learners of Norwegian, but a series of investigations looked into the production of Swedish lexical pitch accents by speakers with different language backgrounds. The Central Swedish tonal system is similar to the Norwegian one, although the tonal configurations are slightly different. Depending on focus conditions, Central Swedish accent 1 may have the tonal patterns L*H or HL*, while accent 2 may have the patterns H*LH or H*L (Riad, 2014). Some studies indicate that learners of Swedish without lexical tones in their L1 have difficulty producing lexical pitch accents (Kaiser, 2011; Schmid, 1986), and that L1 speakers of tone languages perform better than speakers of non-tonal languages (Eliasson, 1997; Hed, 2014). However, evidence has also been presented that speaking an L1 with tonal contrast is not an advantage in the acquisition of the Swedish pitch accents (Tronnier & Zetterholm, 2013, 2014).

It is often assumed that correct perception of L2 categories is a prerequisite for appropriate production. Some studies show that the presence of tone in the speakers' L1 is advantageous in the perception of non-native tones (Caldwell-Harris, Lancaster, Ladd, Dediu & Christiansen, 2015; Wayland & Guion, 2004). Such an advantage is also attested in one of the few studies of Norwegian as an L2. Van Dommelen and Husby (2009) compared Chinese and German listeners' perception of Norwegian lexical pitch accents and found that the former group outperformed the latter.

Other studies suggest that the interplay between learners' L1 and L2 is more complex than just a question of whether the listeners have experience with lexical tones or not. The inventory of intonational categories (Braun, Galts, & Kabak, 2014; Francis, Ciocca, Ma & Fenn, 2008; So & Best, 2014), the types of tones (Gandour, 1983; Peng et al., 2010; So & Best, 2010; Wu, Munro & Wang 2014), and the domain of specification of tonal contrasts (Schaefer & Darcy, 2014) are among characteristics of the L1 that determine how well listeners perceive non-native tonal contrasts. More studies, in particular in the domain of production, are needed to get a deeper understanding of the factors that determine how successful different learners may be in the acquisition of L2 tones. Further, most studies share the same restricted number of target languages (e.g. Cantonese, Mandarin, Thai), and often English or another well-studied European language is used as L1. Therefore, this study aims at broadening our horizon and deals with speakers of less investigated languages (Lingala and Swahili among others), and with East Norwegian as a target language.

Moreover, studies of speakers with more complex linguistic backgrounds are rare. In cases of the acquisition of a third language (L3), the learner already has second language learning experience, a factor which may be beneficial in the process independently of the linguistic structure of the languages involved (Gut, 2010; Hammarberg, 2001). This is rarely discussed, even in studies where the speakers have acquired more than one language before learning the target language. For instance, in So and Best (2014), French listeners performed better than Australian English listeners in perceiving Mandarin tones, a finding the authors explain by the lack of stress in French. Although they mention that the French speakers also speak English, they do not discuss the potential effect of language learning experience on their results. In the same vein, Caldwell-Harris et al. (2015) show that American English speakers who have experience with Asian languages performed better than monolinguals in the perception of tones, but

they do not raise the issue of whether it is previous experience with tones in itself or just general benefits of multilingualism that may explain this difference.

East Norwegian lexical pitch accents

There is important regional variation in Norwegian prosody (Fintoft, 1970) but the speakers of this study all live in regions where the variety referred to as East Norwegian (EN) is spoken. We will therefore assume that EN is the main target norm for our speakers and concentrate solely on this variety.

From a typological point of view, EN can be classified as a subtype of lexical tone languages, where tones interact with stress. In her recent typology, Jun (2014) refers to languages such as EN as languages with *lexical pitch accent*, a term we will adopt here.

EN has two lexical pitch accents, usually referred to as accent 1 and accent 2 (Kristoffersen, 2000, 2006). Their tonal patterns are respectively L^*H^- and H^*LH^- . Every non-clitic word is associated with one of these accents, which distinguish segmentally similar words, such as the pairs ¹*bønder* ('farmers') and ²*bønner* ('beans'), both pronounced [*bønər*]. Some generalizations can be made as regards the distribution of the accents. For instance, monosyllabic words and words with stress on the last syllable have accent 1 (for more examples, see Kristoffersen, 2006). However, some monosyllabic stems can get accent 2 when inflected.

The first tonal targets L^* or H^* are linked to the stressed syllable. On the surface level, stress generally falls on the first syllable of polysyllabic words, but there are several exceptions (for a more theoretical interpretation of the EN stress system, see Kristoffersen, 2000). For instance, words with the prefix *be-* systematically receive stress on the second syllable (*behandle*, 'to treat'). Moreover, some speakers of EN stress the originally stressed syllable in loanwords from, e.g., Latin or Greek (*universitet*, 'university'), but there are also speakers of this variety who produce such words with initial stress and accent 2 (*universitet*). According to Kristoffersen (2006), the L of accent 2 is linked to the first post-stress syllable and spreads to all the remaining toneless syllables before the next tonal target. He postulates that the same spreading rule also applies to the L^* of accent 1. The last H^- tone of both accents is a boundary tone associated with the right edge of an accentual phrase (AP) (the diacritic $-$ is here used to mark AP-final boundary tones), which is the domain of realization of the full melodies (L^*H^- and H^*LH^-). Words that are utterance-initial or occur just after a boundary tone H^- of the preceding AP are in a position of generating a new AP, which consists of a stressed syllable and all the following unstressed syllables. Hence, the isomorphism between syntactic constituency and AP boundaries is limited. For instance, clitics belong to the same AP as the preceding word and not the following one, although they belong to the same syntactic constituent as the latter.

All tonal targets of Norwegian tend to be realized independently of the position of the word in the utterance and of most pragmatic constraints. Focus is, for instance, signaled by extended pitch range, i.e., the H^- of a focal AP has a higher target point than the other H^- APs of the utterance (see Fretheim & Dommelen, 2012 for a discussion), but the overall melody is not altered. However, the H^- does not necessarily surface in an intonation phrase-final AP that ends with a monosyllabic word. Further, post-focal APs tend to be realized with a flat contour.

1
2
3 From a learner perspective, the Norwegian intonation system is, on the one hand,
4 a quite predictable one, as the distribution of tones is first and foremost determined by the
5 lexical pitch accents associated to the words in an utterance. So, compared to languages
6 such as English (Pierrehumbert, 1980), the choice of tunes is restricted. On the other
7 hand, which accent to associate with a word cannot necessarily be deduced from any of
8 its other properties. Therefore, the successful learner of Norwegian needs to store tonal
9 information as a property of at least some entries in the mental lexicon. Given these
10 conditions, it seems natural to expect that mastering tonal aspects of the language will be
11 a challenge for the learner.
12
13

14 **Prosodic systems of participants' other languages**

15
16 From a prosodic perspective, Swahili and Lingala differ in that the latter but not the
17 former has lexical tones. Swahili has fixed stress on the penultimate syllable of
18 polysyllabic words. Both intensity and duration are mentioned as acoustic correlates of
19 stress in this language (Ashton 1947), but high pitch is the main correlate (Ashton, 1947;
20 Lodhi, 2004; Maw & Kelly, 1975; Mpiranya, 1995; Polomé, 1967). Dependent on its
21 position in the utterance, the stressed syllable can have both rising and falling pitch.
22 However, the highest melodic peak of the word systematically occurs on the stressed
23 syllable (Maw & Kelly, 1975). Variations in fundamental frequency (f_0) are also used to
24 mark edges of prosodic constituents, and signal focus and other pragmatic meanings.
25
26

27
28 Lingala is a lexical tone language, in which every syllable is associated with a
29 tone. The language has two lexical tones, L and H (Clements, 1979; Guthrie, 1940;
30 Kukanda, 1983). Word melodies are variable, and any combinations of tones within the
31 word are possible. Trisyllabic words can for instance have the patterns HHL, LHH, LLL,
32 LLH, HLL, etc. Tshiluba is also a tone language. As Lingala it has two tones, L and H,
33 and maximal tonal density (Spaandonck, 2010).
34

35 We lack detailed description of the prosodic systems of all the other African
36 languages in the speakers' repertoires (see Table 1). However, as they are all Bantu
37 languages (cf. ethnologue.com), we can assume that they have lexical tones (Kissebirth
38 & Odden, 2003).
39

40 **Participants**

41
42 This study is based on data collected for a project on multilingualism at the University of
43 Oslo focusing on the Congolese community in Norway. Fieldwork was conducted during
44 autumn 2013 and spring 2014. Thirteen speakers in all participated in the project. For this
45 study, we selected the eight speakers whose recordings were of the best acoustic quality.
46 In the following section, we will briefly present the speakers' linguistic biographies
47 (Pavlenko, 2007). The information is based on the speakers' narratives that were elicited
48 during focus group interviews where the researcher (one of the authors of this paper), the
49 speaker and members of the speakers' family were present.
50

51
52 There are some common characteristics in the linguistic biographies of all the
53 speakers: i) they acquired French at school, and have received tuition solely in French at
54 least from secondary school on, ii) they have studied English as school subject, but most
55 of them have rarely used it, and iii) they acquired Norwegian as adults when they came to
56 the country as refugees. They all attended Norwegian classes for about one year. It is not
57
58
59
60

clear whether the use of lexical pitch accents was explicitly taught during Norwegian classes, but from the experiences of the authors of this paper (a former teacher of Norwegian as an L2 and an L2 learner of Norwegian, respectively), this is an aspect of pronunciation that tends to be neglected. We divide the speakers in two groups according to their dominant language which coincide with where they grew up in the Democratic Republic of Congo. Four speakers are from Western Congo, in regions where Lingala is the dominant language, while the four others are from Swahili-speaking regions in Eastern Congo. As can be seen from Table 1, three Swahili speakers have some knowledge of Lingala. The fourth speaker of Swahili is to a certain degree familiar with two further African languages. All of them might therefore, at least to some extent, be acquainted with tones. The codes we use represent the speakers' main language (S = Swahili, L = Lingala), gender (M or F) and age (see Table 1).

Table 1 about here

Data and method

Speech recordings

This study is based on recordings of conversations between each speaker and a research assistant who is a native speaker of East Norwegian, and who does not share any of the speakers' other languages, except for English which is weaker than Norwegian for all the speakers and thus not a natural choice for the conversations. The aim was to put the speakers close to a monolingual mode, i.e. a context that favor the activation of one of the languages in the multilingual speaker's repertoire (Grosjean, 2011).

We aimed at creating a context that was as natural as possible in order to enhance the ecological validity of the data, *i.e.* to capture, to the extent to which such is possible, how the speakers would express themselves in Norwegian in a non-controlled setting. The conversations took place in the speakers' homes. The speakers were thus in a familiar environment, which presumably made the context relatively natural. The research assistant was instructed to start the conversation asking the speakers to tell about some topic such as their lives in Congo, differences between Congo and Norway and their experiences as refugees. She was further asked to let the conversation flow naturally once she had found a topic that was of interest to the speaker. Each conversation lasted for about 30 minutes. For this study, we chose excerpts of five minutes of conversation for each speaker. We selected parts where there were few interruptions from the interlocutor, where the quality of the signal allowed reliable f0 analysis and the speaker seemed confident and fluent.

The excerpts were transcribed orthographically in Praat (Boersma & Weenink, 2015) and segmented manually into words, syllables and phonemes. Speakers' realizations of phonemes were transcribed using SAMPA where appropriate.

Control data

In addition to the conversation with the Congolese speakers, we have studied five minutes of spontaneous conversation between native speakers of Norwegian (Spilková,

2014). Two different conversations were used for this study, one with a female main speaker (NOR_F) and one with a male main speaker (NOR_M). Both speakers have lived their whole life in parts of Norway where EN is spoken, and can thus be considered representative for the target norm to which the Congolese speakers are exposed. As the conversations with the Congolese speakers, the recordings were orthographically transcribed and manually segmented into words, syllables and phonemes.

Annotation, selection and identification of APs

Since the AP is the domain of realization of the two Norwegian pitch accents, it is the tonal content of this unit that is of interest for this study. We adopted an annotation system that does not represent the actual realization of the AP, but the pattern that is expected in EN. Each AP was annotated using the labels LH (representing accent 1) or HLH (representing accent 2) according to the accent of its first word in standard EN (Vanvik, 1985). We used these annotations as a reference for determining whether the speakers had realized the AP with the expected pattern or not and as anchor points in our instrumental analysis of f0 contours.

Realized APs in the corpora were identified in the following way. As a departure point, we selected content words that were in a position of generating an AP. That is, words that are utterance-initial or occur just after a pitch peak that could be interpreted as the boundary tone H^- of the preceding AP. The syllable of the word that is stressed in EN was considered to be the left edge of the AP. The relative pitch of this syllable was our reference for determining whether the AP had accent 1 or accent 2. Note that we did not examine any other correlates of stress than variations in f0 in this study.

The right edge of an AP was identified as follows. The first syllable with a pitch peak following the stressed syllable within the same breath-group, i.e., a sequence of speech without pauses, was considered to carry a H^- boundary tone marking the right edge of the AP.

In a pilot study, we identified the 100 first APs of each transcribed excerpt and annotated their tonal pattern by inspection of the f0 contours (Steien & Hansen 2015). However, as the recordings were done in the speakers' homes, the quality of the recordings varied and not all of the APs occurred in contexts where the quality of the signal was sufficient to undertake detailed acoustic analysis. Therefore, the final number of APs investigated was 60 per speaker. As we gave priority to the acoustic quality, we were not able to fully balance the number of words belonging to each (see Table 2).

Table 2 about here

Acoustic measures and statistical evaluation

Analysis was performed as follows. Praat (Boersma & Weenink, 2015) was used to extract f0 values from the acoustic waveform. Special care was taken to obtain appropriate values by adjusting analysis settings for individual speakers. Due to varying settings, f0 frame rates varied between 5 ms and 15 ms. For each segmental interval, extracted f0 values were used to determine f0 minimum, f0 maximum, and mean f0. Annotated intervals were the segment of the stressed syllable and the following ones until

1
2
3 the last syllable of the AP. As a rule, those segments were vowels; only a limited number
4 of them were nasals. The number of measured segmental intervals in an AP varied
5 between two and four for both LH and HLH patterns. An example of an LH token
6 featuring three measuring intervals is given in Figure 1. For f₀ pattern evaluation,
7 measuring intervals were represented by their respective mean f₀ values expressed in
8 semitones re 100 Hz (in the figure's example 13.2 st, 13.7 st and 16.0 st respectively).
9 The only exception to the use of mean f₀ values in pattern evaluation was the occurrence
10 of HLH patterns with only two measured intervals in disyllabic words exhausting an AP.
11 In these cases, the first interval's maximum and minimum f₀ were taken to calculate the
12 fall from H to L.
13
14
15

16
17 -----
18 Figure 1 about here
19 -----
20

21 Data were analyzed using chi-square testing and the R program's package lme4 (R Core
22 Team, 2015) to calculate Linear Mixed Effects Models (LMEM). Rejection level was at α
23 = 0.05.
24

25 Findings

26 Three aspects of each AP realization were analyzed: pattern consistency, f₀ dynamic
27 range, and rate of f₀ change. Pattern consistency was defined as a criterion requiring
28 segmental intervals' f₀ values to be in accordance with phonological patterns LH and
29 HLH, respectively. To investigate the acoustic prominence of H vs. L realizations,
30 corresponding f₀ differences were calculated and will be referred to as f₀ dynamic range.
31 Finally, the rate of f₀ change measure specified steepness of f₀ rise in LH, and steepness
32 of f₀ fall and rise in HLH patterns. In each of the following three sections, our evaluation
33 procedure will be described in more detail.
34
35
36
37

38 **Pattern consistency**

39 The first issue we shall address is the question to which degree speakers produced tonal
40 patterns in accordance with expected phonological specifications. Whereas accent 1
41 patterns have two phonologically relevant anchor points (L and H), accent 2 patterns can
42 be evaluated by investigating three such points (H, L, and H). Let us start with evaluation
43 of accent 1 pattern, LH. Since measurements not always involved only two mean f₀
44 values (one for L and H each) but also three or four, evaluation was performed following
45 two different criteria. The most stringent evaluation criterion required a monotonic rise of
46 mean f₀ over all measured intervals, be it two, three or four. In Table 3 this criterion is
47 entitled *all-rising*. From the table it can be seen that a relatively large amount of tonal
48 patterns did not comply with that requirement. For Norwegian natives an amount of
49 85.2 % of the patterns were found to be consistent. They outperformed both Swahili
50 (78.3 %) and Lingala (74.3 %) speakers, though not to a large degree. Even the largest
51 between-group difference (corresponding to 10.9 %) did not reach statistical significance
52 ($\chi^2(1) = 2.49$; $p = 0.114$). The other criterion used in data evaluation merely required
53 mean f₀ in the AP-initial segmental interval (representing L) to be lower than mean f₀ in
54
55
56
57
58
59
60

the final interval, regardless of the pattern found for intervals in between. Adopting this less stringent evaluation criterion causes a substantial rise in scores considered consistent, the increase being largest for the Norwegian natives. For the latter group, the amount of consisting patterns (98.1 %) is significantly higher than the percentages found for the other two groups (Lingala: 86.7 %; $\chi^2(1) = 5.50$, $p = 0.019$; Swahili: 85.9%; $\chi^2(1) = 5.92$; $p = 0.015$).

Due to the presence of three phonologically relevant anchor points, evaluation of HLH patterns was more complicated. Here, three different criteria were used. The first one was the default applying to tokens with three f_0 measurement intervals, requiring a contour in congruence with its phonological specification, i.e. being phonetically high-low-high. For tokens involving four measured intervals, the following strategy was chosen. The first mean f_0 value had to be higher than the second one and at the same time, the last value higher than preceding value number three. Finally, a number of accent 2 patterns involved only two f_0 measurement intervals. The criterion for this situation was a falling f_0 contour during the first measured interval, while the mean f_0 value belonging to the second measured interval had to be higher than the minimum value of the first one. Mean f_0 of the second interval rather than a rising contour in that interval was chosen considering often very short segments, rendering f_0 contours unrealistic. Since results emerging for each of the three above-mentioned evaluation criteria were basically the same, Table 3 presents pooled results under the heading *falling-rising*. Although also here scores for Norwegian natives were higher than for the two groups of L2 users, none of the between-group differences reached statistical significance.

 Table 3 about here

We then looked at the tokens that did not have the expected pattern. We were particularly interested to see if there were systematic differences between native and non-native productions in this respect.

Norwegian natives. In general, phonological specification seemed to be reflected in production, but in some cases the alignment between text and tune differed slightly from our expectations. Some particularly early or late realizations of peaks (*H) or valleys (L*) were thus not captured by our measures.

Non-native speakers. Different from what was found in native productions, variation in timing was not the reason why some of the APs produced by the Congolese speakers did not have the expected pattern. The unexpected patterns were in most cases due to realization of accent 1 as accent 2 and *vice versa*. However, no realizations that do not exist in the Norwegian tone system were attested.

f₀ dynamic range

The second aspect of the tonal patterns we looked into was the tonal dynamics of the LH and HLH realizations in the APs that were realized with the expected patterns. Even if there is an appropriate f_0 movement as defined by the criteria mentioned above, the size of the movement, and so its perceptual relevance, might differ between realizations.

Ranges for LH patterns were always calculated as the difference between f_0 minimum in

the initial and f_0 maximum in the final segmental interval, both for the *all-rising* and the *initial < final* condition (see Fig. 1). Data presented in Table 3 revealed that all three speaker groups produced f_0 movements sized about 4-5 semitones. As was the case with consistency of pattern production, non-native and native values were similar. There was a tendency for Swahili speakers to have generally larger movements but according to statistical analysis this tendency appeared not to be significant (LMEM analyses with L1 as a fixed factor with by-subject random intercepts showed that range differences for LH were non-significant; $F(2, 199) < 1$ and $F(2, 227) < 1$).

Ranges presented for the *falling-rising* HLH conditions were always calculated as the average of HL fall and LH rise. Ranges for patterns containing two measuring intervals were defined as the average of fall from f_0 maximum to f_0 minimum in the first interval and rise from f_0 minimum to f_0 maximum in the second interval. For tokens containing three intervals, relevant f_0 values were the maxima in the first and third interval, respectively, and the minimum in the second one. Finally, ranges for HLH productions with four measuring intervals were calculated as the average of fall from f_0 maximum in the first to f_0 minimum in the second interval and rise from f_0 minimum in the third to f_0 maximum in the last interval.

As is obvious from the table, pooled range data for HLH were in accordance with those for LH patterns. Generally, values were somewhat smaller but, importantly, non-native and native values were similar (according to LMEMs similar to the analyses mentioned above, the effect of L1 never reached significance).

Rate of f_0 change

To further investigate possible differences in tonal productions related to differences in linguistic experience between the Lingala and the Swahili speakers, we analyzed rate of f_0 change during LH and HLH patterns. For LH realizations, this measure was always f_0 change across the time interval from f_0 minimum in the initial interval to f_0 maximum in the final measuring interval (see Fig. 1). Similarly, rate of f_0 change in HLH tokens was calculated as the average of the changes across the fall and the rise. Measuring intervals for falls and rises were the same as used for the analysis described above in section Pattern consistency. To facilitate phonetic interpretation, measures were expressed in semitones per ds (100 ms).

Data are presented in Table 4. Generally, the Lingala speakers had somewhat lower f_0 change rates than the Swahili speakers. A similar picture was found their corresponding range values, which seems to suggest that the Swahili speakers actually showed moderately larger f_0 dynamics (Table 3). Their larger range values resulted from steeper falls and rises rather than from more shallow f_0 movements across longer time spans. At the same time, there are no systematic differences between results for natives and non-natives. According to LMEM analysis none of the differences in f_0 change rates did reach statistical significance. On the whole, it seems safe to conclude that non-native and native behavior was similar as regards f_0 dynamics.

 Table 4 about here

Discussion and conclusion

Despite some typical L2 features, such as interchanging the two Norwegian pitch accents, the overall picture that emerges from our data is that the non-native speakers have high proficiency in the EN tonal system. We should underline that several aspects of the speakers' prosody have not been examined here. We have not measured any other acoustic aspects of the phonologically stressed syllable in EN syllable than its tonal identity. The fact the speakers are able to associate tones with the right syllable does not imply that they master other aspects of the stress system. However, the aim of this study was not to evaluate the speakers' pronunciation as such, but to examine to what extent they produce lexical pitch accents according to the EN target norm in natural spontaneous speech.

Given that prosody is an L2 aspect that tends to be acquired late or never (cf. Mennen 2015), it is reasonable to state that the speakers have overcome a main challenge of the L2 acquisition of EN. The question is then why all of the speakers perform so well, in spite of differences in age of onset, length of stay and linguistic background? It is obviously difficult to weigh the impact of all the different factors that play a role in second language acquisition, but a series of observations allow us to put forward some ideas for further research. First, all of the speakers had been in Norway for at least six years at the moment of the recordings. During the focus groups interviews, they all told that they use Norwegian regularly in different kinds of interactions, for instance at the workplace, in church or in other social activities. Regular and longstanding practice of the languages is obviously a factor that has contributed to successful learning outcome (Miller & Kubota, 2013). Second, all the speakers expressed the need for learning Norwegian to be able to find a job, and to participate in the Norwegian society. Consequently, we can assume that they have been motivated to learn and hence have put effort into the acquisition of Norwegian (Gardner, 2001; Gass, 2013; Pavlenko & Lantolf 2000; Selinker & Gass, 2008). Factors such as motivation, exposure and practice are obviously conditions necessary for successful L2 acquisition, although they do not guarantee target-like proficiency (see for example Mayo & Soler, 2013).

The fact that the Swahili speakers performed equally well as the Lingala speakers indicate that speaking a language with tonal contrasts as one's dominant language is not a decisive factor, at least not at an advanced level of proficiency. If experience with tone languages is *per se* an advantage, this could explain that there were few differences between the two groups because the Swahili speakers grew up in an environment where tone languages were spoken. That said, it is not clear to what extent the Swahili speakers master(ed) the tone system of languages they were exposed to as children. Moreover, as we saw above, the interaction between L1 prosody and L2 tones seems to be more complex than just a question of whether the speaker has experience with tone language or not.

What is certain, though, is that the present speakers had extensive language learning experience at the moment when they started to learn Norwegian. What all the speakers have in common is that they are multilingual from birth on and have continued learning and using different languages across the lifespan. Given what is known about second language acquisition by multilinguals (e.g. Hammarberg, 2010), we can hypothesize that the fact that the speakers have acquired and alternate daily between several typological distant prosodic systems might have been beneficial for the

1
2
3 acquisition of a new prosodic system. Hopefully, we will see more studies involving
4 speakers with complex language backgrounds in order to get a more thorough
5 understanding of potential advantages of multilingualism in the acquisition of non-native
6 tones.
7

8 9 **Acknowledgements**

10
11 This work was partly supported by the Research Council of Norway through its Centers
12 of Excellence funding scheme, project number 223265. We would also like to express
13 our gratitude to the two anonymous reviewers of this paper for their useful comments and
14 to the speakers who accepted to participate in this study.
15

16 17 **References**

- 18
19 Ashton, E. O. (1947). *Swahili grammar, including intonation*. London: Longman.
- 20 Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception:
21 Commonalities and complementarities. In O.S Bohn, & M. J. Munro (Eds.),
22 *Language experience in second language speech learning: In honor of James*
23 *Emil Flege* (pp. 13-34). Amsterdam/Philadelphia: John Benjamins Publishing
24 Company.
25
- 26 Boersma, P., & Weenink, D. (2015). *Praat: doing phonetics by computer. Version 6.0.05*.
27 Retrieved from www.praat.org.
- 28 Braun, B., Galts, T., & Kabak, B. (2014). Lexical encoding of L2 tones: The role of L1
29 stress, pitch accent and intonation. *Second Language Research, 30*, 323-350.
- 30 Broselow, E., & Kang, Y. (2013). Second language phonology and speech. In J.
31 Herschensohn & M. Young-Scholten (Eds), *The Cambridge handbook of second*
32 *language acquisition* (pp. 529-554). Cambridge: Cambridge University Press.
- 33 Caldwell-Harris, C. L., Lancaster, A., Ladd, D. R., Dediu, D., & Christiansen, M. H.
34 (2015). Factors influencing sensitivity to lexical tone in an artificial language.
35 *Studies in Second Language Acquisition, 37*, 335-357.
36
- 37 Clements, G. N. (1979). The description of terraced-level tone languages. *Language, 55*,
38 536-558.
- 39 Eliasson, S. (1997). Tone in second language acquisition. *Trends in Linguistic Studies*
40 *and Monographs, 101*, 1273-1290.
- 41 Fintoft, K. (1970). *Acoustical analysis and perception of tonemes in some Norwegian*
42 *dialects*. Oslo: Universitetsforlaget.
- 43 Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In
44 W. Strange (Ed), *Speech perception and linguistic experience: Issues in cross-*
45 *language research* (pp. 233-277). York: York Press.
- 46 Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese
47 lexical tones by tone and non-tone language speakers. *Journal of Phonetics, 36*,
48 268-294.
- 49 Fretheim, T., & van Dommelen, W. A. (2012). A pragmatic perspective on the
50 phonological values of utterance-final boundary tones in East Norwegian
51 intonation. *The Linguistic Review, 29*, 663-667.
- 52 Gandour, J. (1983). Tone perception in far eastern-languages. *Journal of Phonetics, 11*,
53 149-175.
54
55
56
57
58
59
60

- 1
2
3 Gardner, R. C. (2001). Language learning motivation: The student, the teacher, and the
4 researcher. *Texas Papers in Foreign Language Education*, 6(1), 1-18.
- 5 Gass, S. M., & Selinker, L. (2008). *Second language acquisition: An introductory course*
6 (Third edition). New York and London: Routledge.
- 7
8 Grosjean, F. (2011). An attempt to isolate, and then differentiate, transfer and
9 interference. *International Journal of Bilingualism*, 16, 11-21.
- 10 Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge
11 University Press.
- 12
13 Gut, U. (2010). Cross-linguistic influence in L3 phonological acquisition. *International*
14 *Journal of Multilingualism*, 7(1), 19-38.
- 15 Guthrie, M. (1940). Tone ranges in a two-tone language (Lingala). *Bulletin of the School*
16 *of Oriental and African Studies*, 10, 469-478.
- 17
18 Hammarberg, B. (2010). The languages of the multilingual: Some conceptual and
19 terminological issues. *IRAL-International Review of Applied Linguistics in*
20 *Language Teaching*, 48(2-3), 91-104.
- 21 Hammarberg, B. (2001). Roles of L1 and L2 in L3 production and acquisition. In J.
22 Cenoz, B. Hufeisen, & U. Jessner (Eds.), *Cross-linguistic influence in third*
23 *language acquisition: Psycholinguistic perspectives* (pp. 21-41).
24 Clevedon/Buffalo/Toronto/Sydney: Multilingual Matters LTD.
- 25
26 Hao, Y.-C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and
27 non-tonal language speakers. *Journal of Phonetics*, 40, 269-279.
- 28
29 Hed, A. (2014). Perception and production of Swedish word accents by Somali L1
30 speakers. *Proceedings from FONETIK 2014*, 105-110.
- 31 Jun, S.-A. (2014). *Prosodic typology II: The phonology of intonation and phrasing*.
32 Oxford: Oxford University Press.
- 33
34 Kaiser, R. (2011). Do Germans produce and perceive the Swedish word accent contrast?
35 A cross-language analysis. *TMH-QPSR*, 51, 93-96.
- 36
37 Kissebirth, C., & Odden, D. (2003). Tone. In D. Nurse, & G. Philippson (Eds.), *The*
38 *Bantu Languages* (pp. 59-70). London: Routledge.
- 39
40 Kristoffersen, G. (2000). *The phonology of Norwegian*. Oxford: Oxford University Press.
- 41
42 Kristoffersen, G. (2006). Markedness in urban East Norwegian tonal accent. *Nordic*
43 *Journal of Linguistics*, 29, 95-135.
- 44
45 Kukanda, V. (1983). *L'emprunt français en lingála de Kinshasa: quelques aspects de son*
46 *intégration phonétique, morphologique, sémantique et lexicale*. Tübingen: Gunter
47 Narr Verlag.
- 48
49 Lodhi, A. Y. (2004). Strategies of emphasis and intensity in Swahili. *Africa & Asia*, 4,
50 142-150.
- 51
52 Macken, M. A. (2008). Factors affecting the production of Vietnamese tones: A study of
53 American learners. *Studies in Second Language Acquisition*, 30, 49-77.
- 54
55 Maw, J., & Kelly, J. (1975). *Intonation in Swahili*. London: University of London,
56 School of Oriental and African Studies.
- 57
58 Mayo, G., & Soler, E. A. (2013). Negotiated input and output/interaction. In J.
59 Herschensohn, & M. Young-Scholten (Eds.), *The Cambridge handbook of second*
60 *language acquisition* (pp. 209-229). Cambridge: Cambridge University Press.
- Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of
Greek. *Journal of Phonetics*, 32, 543-563.

- 1
2
3 Mennen, I. (2015). Beyond Segments: Towards a L2 Intonation Learning Theory. In E.
4 Delais-Roussaire, M. Avanzi, & S. Herment (Eds.), *Prosody and language in*
5 *contact* (pp. 171-188). Heidelberg: Springer.
- 6
7 Miller, E., & Kubota, R. (2013). Second language identity construction. *The Cambridge*
8 *handbook of second language acquisition*, 230-250.
- 9
10 Mpiranya, F. (1995). *Swahili phonology reconsidered in a diachronical perspective: the*
11 *impact of stress on morphonemics and syllable structure*. Köln: Rudiger Köppe
12 Verlag.
- 13
14 Pavlenko, A. (2007). Autobiographic narratives as data in applied linguistics. *Applied*
15 *Linguistics*, 28, 163-188.
- 16
17 Pavlenko, A. & Lantolf J. P. (2000). Second language learning as participation and the
18 (re) construction of selves. In J.P. Lantolf (Ed), *Sociocultural theory and second*
19 *language learning* (pp. 155-177). Oxford: Oxford University Press.
- 20
21 Peng, G., Zheng, H.-Y., Gong, T., Yang, R.-X., Kong, J.-P., & Wang, W. S.-Y. (2010).
22 The influence of language experience on categorical perception of pitch contours.
23 *Journal of Phonetics*, 38, 616-624.
- 24
25 Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*. Phd
26 Thesis. Massachusetts Institute of Technology.
- 27
28 Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign
29 accent in an L2: A review. *Journal of Phonetics*, 29, 191-215.
- 30
31 Polomé, E. C. (1967). *Swahili Language Handbook*. Washington: Center for applied Riad,
32 Riad, T. (2014). *The Phonology of Swedish*. Oxford: Oxford University Press.
- 33
34 Schaefer, V., & Darcy, I. (2014). Lexical function of pitch in the first language shapes
35 cross-linguistic perception of Thai tones. *Laboratory Phonology*, 5, 489-522.
- 36
37 Schmid, B. (1986). A comparative study of children's and adult's acquisition of tone
38 accents in Swedish. *Language Learning*, 36, 185-210.
- 39
40 So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts:
41 Effects of native phonological and phonetic influences. *Language and Speech*, 53,
42 273-293.
- 43
44 So, C. K., & Best, C. T. (2014). Phonetic influences on English and French listener's
45 assimilation of Mandarin tones to native prosodic categories. *Studies in Second*
46 *Language Acquisition*, 36, 195-221.
- 47
48 Spilková, H. (2014). *Phonetic reduction in spontaneous speech: an investigation of*
49 *native and non-native production*. Phd Thesis. Norwegian University of Science
50 and Technology, Trondheim.
- 51
52 Steien, G. B., & Hansen, P. (2015). Target-like distribution of Norwegian lexical pitch
53 accents in spontaneous speech produced by L2 speakers. In The Scottish
54 Consortium for ICPhS 2015 (Eds.), *Proceedings of the 18th International*
55 *Congress of Phonetic Sciences* (Paper number 0787). Glasgow: the University of
56 Glasgow.
- 57
58 R Core Team (2015). R: A language and environment for statistical computing. Retrieved
59 from <https://www.R-project.org/>
- 60
61 Tronnier, M., & Zetterholm, E. (2013). Tendencies of Swedish word accent production
62 by L2-learners with tonal and non-tonal L1. *Nordic prosody: proceedings of the*
63 *XIth conferece*, 391-400.

- 1
2
3 Tronnier, M., & Zetterholm, E. (2014). Appropriate tone accent production in L2-
4 Swedish by L1-speakers of Somali? *Proceedings of the international symposium*
5 *on the acquisition of second language speech. Concordia working papers in*
6 *applied linguistics, 5*, 722-736.
7
8 van Dommelen, W. A., & Husby, O. (2009). Perception of Norwegian word tones by
9 Chinese and German listeners. In M. A. Watkins, A. S. Rauber, & B. O. Baptista
10 (Eds.), *Recent research in second language phonetics/phonology: Perception and*
11 *production* (pp. 318-331). Cambridge: Cambridge Scholar Publishing.
12
13 van Spaandonck, M. (2010). On the so-called reversing tonal system of Ciluba: a Case
14 for restructuring. *Studies in African Linguistics, 2*, 131-144.
15
16 Vanvik, A. (1985). *Norsk uttaleordbok: A Norwegian pronouncing dictionary*. Oslo:
17 Universitetet i Oslo.
18
19 Wayland, R., & Guion, S. (2004). Training native English and native Chinese speakers to
20 perceive Thai tones. *Language Learning, 54*, 681-712.
21
22 Wu, X., Munro, M. J., & Wang, Y. (2014). Tone assimilation by Mandarin and Thai
23 listeners with and without L2 experience. *Journal of Phonetics, 46*, 86-100.
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1. Participant Characteristics (AoO= Age of Onset and LoR= Length of Residence in years).

Speaker	Dominant language (s)	Other languages Order = descending order of self-reported proficiency	AoO	LoR
S_M33	Swahili	French, English, Norwegian, Lingala, Kinyndu	25	8
S_M36	Swahili	Norwegian, French, Lingala, Mashi, Luganda, English	25	11
S_F43	Swahili (French)	Norwegian, Lingala, Kirwanda, Kinyarwanda, Mashi, English	30	13
S_M52	Swahili	French, Norwegian, Bemba, Taabwa, English	46	6
L_F33	Lingala	Norwegian, French, Ekonda, English	27	6
L_F50	Lingala	French, Norwegian, Tshiluba	37	13
L_M53	Lingala (Tshiluba)	French, Norwegian, Kishelele, English	40	13
L_M55	Lingala (French)	Norwegian, English, Tshiluba	52	13

Table 2. Distribution of AP types investigated for each of 10 speakers.

Speaker	Accent 1	Accent 2
S_M33	20	39
S_M36	19	41
S_F43	32	28
S_M52	21	38
L_F33	18	42
L_F50	28	31
L_M53	33	27
L_M55	34	26
NOR_F	23	37
NOR_M	31	29

Table 3. Percentage of f0 patterns consistent with phonological specifications as LH and HLH. All-rising= monotonically rising f0 values over all (2, 3 or 4) segmental intervals. Initial < final= f0 in initial interval lower than in final interval. Falling-rising: see text. LIN= Lingala, SWA= Swahili, NOR= Norwegian. Range= f0 difference L-H and H-L-H in semitones (see text). N= total number of observations (100 %).

		All-rising			Initial < final	
Tonal pattern	Group	N	Consistent (%)	Range (st)	Consistent (%)	Range (st)
LH	LIN	113	74.3	4.1	86.7	4.1
	SWA	92	78.3	5.1	85.9	5.1
	NOR	54	85.2	4.9	98.1	4.7

		Falling-rising			
Tonal pattern	Group	N	Consistent (%)	Range (st)	
HLH	LIN	126	82.5	2.8	
	SWA	147	77.6	3.4	
	NOR	66	86.4	3.0	

Table 4. Rate of f0 change in LH (rise) and HLH patterns (fall-rise) in semitones/ds (=100 ms). All-rising= monotonically rising f0 values over all (2, 3 or 4) segmental intervals. Initial < final= f0 in initial interval lower than in final interval. f0 change in LH= f0 difference in final – initial interval. Falling-rising: see text. LIN= Lingala, SWA= Swahili, NOR= Norwegian. n= number of observations out of total N.

Tonal pattern	Group	All-rising			Initial < final	
		N	Rise (st/ds)	n	Rise (st/ds)	n
LH	LIN	113	1.8	84	1.7	98
	SWA	92	2.4	72	2.3	79
	NOR	54	2.2	46	2.1	53

Tonal pattern	Group	Falling-rising		
		N	Fall-Rise (st/ds)	n
HLH	LIN	126	1.8	104
	SWA	147	2.3	114
	NOR	66	1.9	57

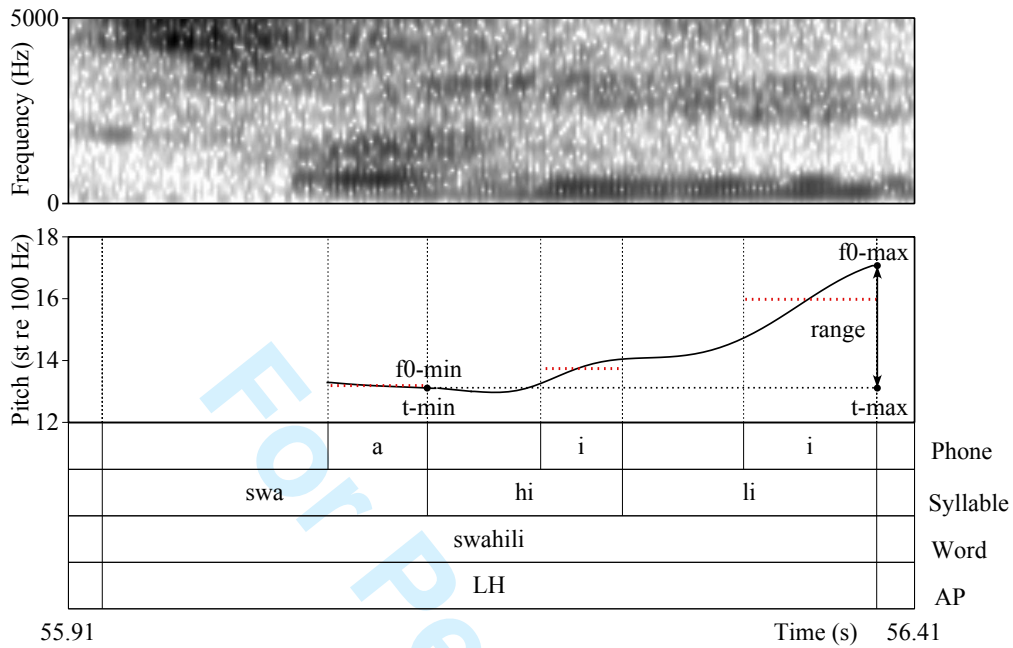


Figure 1. An LH token produced by speaker L_F50. Spectrogram and f0 contour. Mean f0 values in three segmental intervals are indicated by dotted lines. Range = f0-max – f0-min (indicated by vertical arrow). Rate of f0 change = range/(t-max – t-min).