Rhythm analysis in Arabic L2 speech

Ghania Droua-Hamdani[1], Sid-Ahmed Selouani[2], Yousef A. Alotaibi[3]

[1]Speech Processing Laboratory, CRSTDLA, Algiers, Algeria
[2]LARIHS Laboratory, University of Moncton, Shippagan, Canada
[3]College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

gh.droua@post.com, sid-ahmed.selouani@umoncton.ca, yaalotaibi@ksu.edu.sa

Abstract

This paper investigates rhythm speech metrics in Modern Standard Arabic. Recordings of native (L1) and non-native (L2) speakers were obtained from the West Point corpus. The experiment examined the rhythm metric properties of L2 speech using Pairwise Variability Indices and Interval Measures. The application of the Control/Compensation Index to the corpus is also described. Variations in rhythm metrics are detailed by focusing on between-speaker differences such as gender of speakers. L1 and L2 speakers (females and males) were compared through measurement of the duration of short and long vowels.

Index Terms: rhythm metrics, Control/Compensation Index, L2 speakers, gender of speakers, Modern Standard Arabic, English

1. Introduction

Recent studies have developed a battery of metrics to show rhythm differences and similarities between languages. These same metrics have also been used to study second language acquisition, by examining the impact of the first language (L1) on the rhythm of the second language (L2). A large number of studies have investigated the effect of L1 use on L2 production: Korean English [1]; Singapore English [2]; German L2 influenced by Chinese, English, French, Italian and Romanian L1 [3]; Norwegian as L2 [4]; English, Spanish, Dutch and French [5].

Rhythm refers to the temporal organization of speech. The rhythm models developed are based on the acoustic durations of vocalic and consonantal intervals in vocal signals. The most popular algorithms performed are: Interval Measures (IM) and Pairwise Variability Indices (PVI). The IM approach involves computing three separate measures from the segmentation of speech signals (global utterance) into vocalic (V) and consonantal (C) units (ΔV, ΔC and %V) [6]. The time-normalized metric measures (VarcoV/C) were introduced when it was observed that the consonantal interval measure is inversely proportional to speech rate [7]. The PVI algorithm differs from IM and VarcoV/C models in that it focuses on the temporal succession of the vocalic and consonantal intervals instead of the global utterance [8]. The model suggests that the rPVI should be used for the consonantal intervals, while the nPVI (normalized Pairwise Variability Index) should be used for the vocalic intervals, which are more prone to be affected by speech rate.

More recently, a new proposal of rhythm metrics has been put forward [9]. The model is inspired by previous studies on syllable compensation, which state that controlling languages (syllable-timed) show low levels of compensation at intra- and inter-syllabic levels in comparison to compensating languages (stressed-timed) that are thought to show higher levels of compensation. The rhythm metric suggested is the Compensation and Control Index (CCI) that may be used for computing the level of compression (lengthening or shortening) allowed in a language according to the context. In order to show the intra-syllabic comportment, the CCI takes into account all the segments composing each vocalic and consonantal interval. The formula used in CCI computation consists of a modification of the rPVI algorithm where each vocalic or consonantal interval is divided by the number of phonological segments that are included in this interval. A new version of the CCI formula has been developed [10], but for the purpose of this experiment, we used the original formula, shown below, where *m* stands for V or C intervals, *d* for duration, and *n* for number of segments within the relevant interval.

$$cci = \frac{100}{(m-1)} \sum_{k-1}^{m-1} \left| \frac{d_k}{n_k} - \frac{d_{k+1}}{n_{k+1}} \right|$$

The model aims to offer a better representation of the rhythmic tendencies of natural languages, suggesting that geminate consonants and long vowels separately count for two different segments. According to the model, the controlling languages should present similar tendencies of C and V local durational fluctuations scattered along the bisecting line. Conversely, compensating languages should fluctuate more in the V than in the C segments. They should be clustered under the bisector as shown in Figure 1 [10].

The experiment, described in this paper, focused on the rhythm properties of Arabic L2 speech by gauging the influence of L1 on L2 rhythm. A comparison of Arabic L2 with different languages is also given. The rhythm metrics that were examined include: three interval measures (%V, ΔV, and ΔC), two time-normalized indices (VarcoV, VarcoC), two pairwise variability indices (nPVI-V, rPVI-C), and two compensation and control indices (CCI-V, CCI-C).

10.21437/SpeechProsody.2014-190

The paper is organized as follows: Section 2 summarizes the main characteristics of the Arabic language. Section 3 gives an overview of the speech data and speakers used in the study. Section 4 describes the results and discussion about different rhythm experiments on Arabic L2. Section 5 concludes this work.

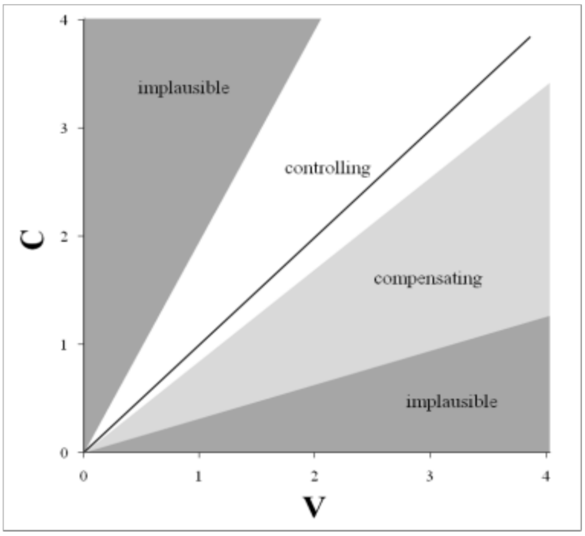

Figure 1: *Schematic representation of the major rhythmic types according to the CCI model [10].*

2. Modern Standard Arabic

Arabic is part of the Semitic language family, and spoken by more than 340 million speakers from the Middle East to North Africa. Modern Standard Arabic (MSA) is used in most written documents as well as in formal spoken interactions in all Arabic countries.

MSA uses a distinct phonological set that includes 28 consonants and six vowels: three short vowels /a/, /u/ and /i/ and three long vowels /a:/, /u:/ and /i:/, that are the counterparts of the short vowels. Regarding the number of vowels, Arabic is much less complex than English, which has twelve vowels. English and Arabic rely on different types of phonological contrasts in both vowels'quality and quantity [11].

There are two distinct consonant classes in MSA: pharyngeal and emphatic consonants. In addition to these, the language is characterized by two distinctive features that are fundamental in avoiding semantic ambiguity: long vowels and gemination. Gemination, or consonant elongation, occurs when a consonant is pronounced for a perceptibly longer period of time than a short consonant. All Arabic consonants may be geminated.

The Arabic language consists of two kinds of syllables: open syllables (CV and CV:) —and closed syllables (CVC, CV:C and CVCC). Arabic vowels never occur initially. Every vowel must be preceded by a consonant (which may include the glottal stop [ʔ]). All Arabic syllables must contain at least one vowel [12].

3. Corpus description and speakers

Speech material used for the study was taken from the West Point corpus, which was collected and processed by the Department of Foreign Languages at the United States Military Academy at West Point and the Center for Technology-Enhanced Language Learning (CTELL) (Linguistic Data Consortium (LDC) [13]). West Point speech files were recorded from 110 speakers. The corpus consists of collections of four main Arabic scripts including a total of 258 sentences. Script 1 is spoken by 75 L1 speakers of Arabic (41 males). Scripts 2–4 are read by 35 L2 speakers (25 males). The L2 speakers' recordings were captured at a sampling rate of 16 bit at 22.05 kHz. The recordings were collected at normal speech rate. The age of the speakers is not mentioned.

To reduce factors that can compromise the reliability of rhythm analysis, a large sample of speakers was used and the measurement effects across speech materials were controlled for. Speech material from 29 speakers reading either five sentences from scripts 1 or 2 was used. In total, 145 recordings were used in the analysis. Table 1 shows the number and gender of speakers in the sample.

Table 1. *Distribution of native and non-native speakers in Modern Standard Arabic (MSA) corpus.*

Native speakers (L1)		Non-native speakers (L2)	
male	female	male	Female
5	10	6	8
15		14	

To avoid variability due to the segmentation procedure, one researcher manually carried out all segmentations for the speech corpus. Recordings were analyzed using Praat software. All vowels and consonants were segmented by inspection of speech waveforms and wideband spectrograms by one researcher. Vowel and consonant durations were extracted using a customized script on the boundary label files. Rhythm metrics were computed for each sentence for each speaker.

4. Results & Discussion

4.1. IM & PVI metrics for L1 and L2 speakers

The first experiment studied rhythm metrics of L2 speakers in comparison to L1 speakers by computing different IM and PVI metrics. Table 2 reports the average values of each of the seven rhythm metrics applied to the data. Figure 2 reveals that all the vocalic metrics: vocalic interval measure, vocalic time-normalized interval, vocalic percentage, vocalic pairwise variability index (ΔV, VarcoV, %V and nPVI-V) are higher for L1 speakers in comparison to L2 rhythm values. As can be seen from the same plot, the consonantal interval measure and the consonantal pairwise variability index (ΔC and rPVI-C) for L1 speakers also present higher scores than L2 metric values. However, the VarcoC value for L2 speakers is close to the L1 measure.

Table 2. *Modern Standard Arabic (MSA) rhythm metrics (ms) for L1/L2 speakers.*

Metrics	L1 spk.	L2 spk.
%V	42.41	40.12
ΔV	49.42	45.58
ΔC	53.29	43.90
VarcoV	65.14	58.64
VarcoC	50.87	49.79
rPVI-C	74.84	68.71
nPVI-V	56.22	52.17

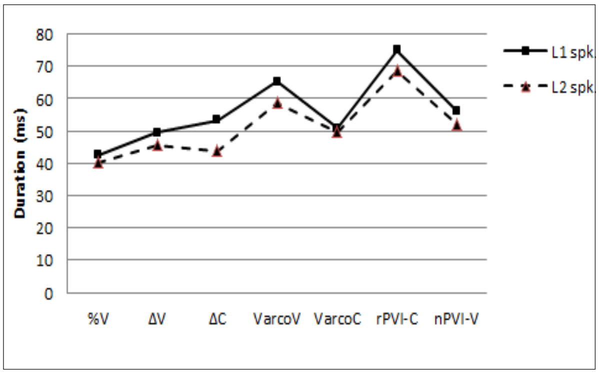

Figure 2: *Mean rhythm metrics of Modern Standard Arabic (MSA) L1 and L2.*

Statistical analysis (One-way ANOVA) shows significant effects between L1 and L2 rhythm values for three metrics: %V, ΔC and ΔV (F (1.143)=13.17, p<0.001; F (1.143)=15.79, p<0.001; F (1.143)=8.71, p=.004; 95% confidence intervals). These results show that L1 speakers present higher vocalic proportions compared to L2 speakers. This fluctuation in vocalic intervals suggests that non-native speakers reduce the vocalic intervals. Likewise, consonantal intervals of L2 speakers are less lengthened than their L1 counterparts. This finding suggests that the pronunciation of MSA consonants by non-native speakers decreases in duration compared to that of native speakers. Considering all the metrics employed, it seems that this reduction, especially in vocalic duration, helps L2 speakers to sustain syllable structures when producing MSA.

4.2. CCI metrics of L1 and L2

The CCI algorithm aims to describe the intra-syllabic behavior of languages. In the second experiment, CCI metrics were computed (vocalic/consonantal compensation and control Index; CCI-V and CCI-C) for each sentence of MSA produced by L1 and L2 speakers. Statistical analysis (One-way ANOVA) shows significant differences in CCI-C and CCI-V values between L1 and L2 rhythm (F (1,143)=4.75, p=0.031; F (1,143)=29.78, p< 0.001).

Regarding the CCI chart, the average values of each of CCI-C and CCI-V were computed for both native (L1) and non-native (L2) speakers. Comparison of the vocalic/consonantal CCI metrics shows that L1 and L2 are separated by the bisector. As shown in Figure 3, L2 Arabic falls close to the bisecting line in the controlling area, whereas L1 Arabic is below the bisector in the compensating part.

Arabic is a stress-timed language, like English and German [14]. According to the CCI model, stress-timed languages should show higher levels of compensation between vowel and consonant segments, whereas controlling languages should present similar tendencies of C and V local durational fluctuations. However, the results reveal that Arabic L1 presents more compensating CCI while the opposite happens to Arabic L2 production, which appears to be more controlling then compensating. MSA L1 fluctuates more in the V than C segments compared to MSA L2, where the variation tendency is similar for vowels and consonants.

Some studies claim that rhythm metrics do not manage to clearly distinguish between L1 and L2 speakers [15]. However, the main findings of this experiment show that the deviation is noticeably observed in the case of Arabic L2 and L1 speakers when CCI metrics are used.

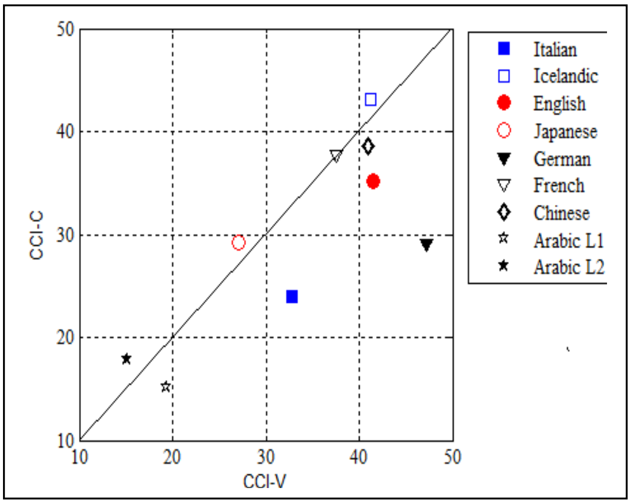

Figure 3: *MSA L1 and L2 among controlling and compensating languages.*

Figure 3 illustrates a comparison of Arabic L1 and L2 with a set of languages including English, French, Italian, Japanese, Chinese, Icelandic and German [16]. The graph shows that Arabic L2 falls in the controlling area, similar to French, which is considered one of the controlling languages. In contrast, Arabic L1 is grouped with the compensating languages, like English and German, which are clustered under the bisector. These results show that in the case of MSA, it is possible to distinguish L2 from L1, and Arabic L1 from other languages.

In this study, the L1 of non-native Arabic speakers was English. When L1 and L2 share the same rhythm typology, it may be difficult to distinguish between L1 and L2 speakers [15]. However, findings reported in Figure 2 show that distinction between L1 and L2 speakers is possible using rhythm metrics, even although Arabic and English are both stress-timed languages.

4.3. Vowel duration in L2

In light of the substantial differences between English and Arabic vowel systems (in terms of quantity) and IM, PVI, VarcoV/C and CCI–V findings, an investigation of the acoustic characteristics of MSA vowels produced by English speakers was conducted. The purpose was to show how durations fluctuate in the vocalic rhythm metric of L2 speakers. One-way ANOVA was conducted to test the effect of the origin of speakers on vowel duration values. The results show a highly significant effect ($F (1, 1347)=29.41$, $p<0.001$). Table 3 presents the comparison of the mean duration values in milliseconds of short (v) and long (V) vowels of L1 and L2 speakers.

Table 3. *Average durations of short (v) and long (V) MSA vowels by L1/L2 speakers (standard deviations are given in parentheses).*

	v (ms)	V (ms)	V/v
L1 spk.	86.62 (36.46)	170.29 (43.27)	1.97
L2 spk.	83.95 (36.41)	139.39 (43.45)	1.66

The results indicate that durations of L1/L2 short vowel values are close to each other. For long vowels, the data do not exhibit a unified pattern for both groups. L1 speakers have longer durations compared to L2 individuals. The results also show that the vowel length contrast (the ratio of long-to-short vowel durations) is greatest in the L1 group, while L2 speakers have a lower ratio. ANOVA was conducted to test the effect of vowel length (short/long) on L1/L2 speakers. The results show a significant effect of origin of speaker on vowel length [$F (1, 1345)=32.42$, $p<0.001$]. These findings suggest that the difference in duration of long vowels is mainly related to the L1 of speakers. The fluctuation observed in vocalic rhythm metric values of L2 compared with L1 speakers is due to the reduction of long vowels rather than short vowels. These results suggest that L2 speakers struggle to produce the long vowels as produced by L1 speakers.

To further examine between-speaker differences according to gender, a comparison of the mean duration of short/long vowels of female and male speakers for both L1 and L2 was performed. Results are summarized in Table 4.

Table 4. *Average durations of long and short MSA vowels by L1/L2 according to gender (standard deviations are given in parentheses).*

	v (ms)	V (ms)	VL1/VL2
L1 female	88.90 (36.99)	168.59 (45.05)	1.23
L2 female	88.09 (37.66)	136.58 (44.87)	
L1 male	82.19 (34.94)	174.10 (39.05)	1.22
L2 male	78.48 (34.02)	143.28 (44.87)	

The results show that L1 and L2 females have similar duration values for Arabic short vowels, while L1 males produce slightly longer durations than their L2 counterparts when producing short vowels. For long vowels, both L2 females and males show similarly reduced duration compared to the L1 females and males respectively (the ratio of VL1/VL2). Gender was the final independent variable tested in this experiment. Results of the ANOVA show a highly significant effect of gender on vowel length ($F (1, 1341)= 8.51$, $p=0.004$). L2 speakers struggle to produce long vowels that are similar to those of L1 speakers in terms of quantity (duration), and this is true for females and males. This can be explained by the influence of the L1 of non-native speakers, in this case the phonetic system of English on that of the L2, Arabic. The L1 of non-native speakers influences the acquisition and subsequently the production of Arabic, the L2.

5. Conclusions

This study examined variation in rhythm of L2 speakers of MSA. Several experiments were conducted to show rhythm properties of Arabic L2 using IM and PVI algorithms. CCI models were also used to describe L2 production at the intra-syllabic level. The main results show that rhythm metrics of L2 speakers present many differences in terms of phoneme duration (vowels and consonants) compared to L1 speakers. To explain the fluctuation in vocalic rhythm metrics, a comparison of vowel duration (short/long) between L1 and L2 speakers was performed. The results show that the fluctuation in vocalic rhythm metrics is due to a reduction in duration of long vowels for both male and female L2 speakers.

6. Acknowledgements

This work is supported by the NPST program at King Saud University, Project Number 10INF1325-02.

7. References

[1] Jang, T.Y., "Automatic assessment of non-native prosody using rhythm metrics: Focusing on Korean speakers' English pronunciation", In Proc. of the 2nd International Conference on East Asian Linguistics, 2009.

[2] Ling, L-E., Grabe, E. and Nolan, F., "Quantitative Characterizations of Speech Rhythm: Syllable-Timing in Singapore English", Language and speech, 43(4):377–401. 2000.

[3] Gut, U., "Non-native speech rhythm in German", In Proceedings of the ICPhS conference, 2437–2440, 2003.

[4] van Dommelen, W., "Temporal patterns in Norwegian as L2", Trends in linguistics studies and monographs, 186:121, 2007.

[5] White, L. and Mattys, S.L., "Calibrating rhythm: First language and second language studies", Journal of Phonetics, 35(4):501–522, 2007.

[6] Ramus, F., "Acoustic correlates of linguistic rhythm: Perspectives". In Speech prosody, 115–120. Citeseer, 2002.

[7] Dellwo. V., "Rhythm and speech rate: A variation coefficient for deltaC", Language and language processing, 231–241, 2006.

[8] Grabe, E. and Low, E.L., "Durational variability in speech and the rhythm class hypothesis", Papers in laboratory phonology, 7:515–546, 2002.

[9] Bertinetto, P.M. and Bertini, C., "On modelling the rhythm of natural languages", Proc. 4th International Conference on Speech Prosody, 427-430, Campinas, 2008.

[10] Zhi, N., Bertinetto, P.M., and Bertini, C., "Modelling the speech rhythm of Beijing Chinese in the CCI framework", Proc. 17th ICPhS, HongKong, 2011.

[11] Kopczynski, A. and Meliani, R., "The vowels of Arabic and English", Papers and studies in contrastive linguistics, 27:183–192, 1993.

[12] Alotaibi, Y. and Selouani, S. A., "Evaluating the MSA West Point speech corpus", International Journal of Computer Processing of Languages, 22(4): 285–304, 2009.

[13] Linguistic Data Consortium LDC. http://www.ldc.upenn.edu.

[14] Selouani, S-A., Alotaibi, Y. A. and Pan, L., "Comparing Arabic rhythm metrics among other languages", In Audio, Language and Image Processing (ICALIP), 2012 International Conference, 287–291, 2012.

[15] Gut. U., "Rhythm in L2 speech", Speech and Language Technology, 14/15, 2012.
[16] Mairano, P. and Romano, A., "Rhythm metrics for 21 languages", Proceedings of ICPhS XVII, Hong Kong, 17–21, 2011.