ICCSPA20 1570619802

Extended Rhythm-Based Investigation of Saudi Dialects Using the Saudi Accented Arabic Voice Bank Corpus

Soha B. Sandouka
Computer Engineering Department
King Saud University
Riyadh, Saudi Arabia
439204384@student.ksu.edu.sa

Yousef A. Alotaibi
Computer Engineering Department
King Saud University
Riyadh, Saudi Arabia
yaalotaibi@ksu.edu.sa

***Abstract*—In this study, we investigate the speech rhythm of five dialectal regions of the Kingdom of Saudi Arabia (KSA). The relations between different Interval Measures (IMs) and their capabilities of classifying the native origins of speakers, their gender (male/female), and type of sentence (question or statement) are studied. The Saudi Accented Arabic Voice Bank (SAAVB) corpus was used for all experiments. It contains a set of Arabic speech that represents native Arabic speakers from all the cities around Saudi Arabia. A number of rhythm metrics, namely, ΔV, ΔC, and %V, are calculated and measured in detail. The results show that ΔV metric can differentiate between question and statement sentences. Moreover, the %V detects a significant difference between male and female speakers. The results of the study demonstrate the uniqueness of the Riyadh dialect. In other words, the Riyadh dialect is different from all other investigated dialects. Moreover, a high similarity is observed between the dialects of Tabuk and Northern Border (NB) regions. These two regions are located in the north of Saudi Arabia and border Jordan and Iraq international states, which may explain the reason behind this similarity. Finally, a conclusion and general discussion on speech rhythm metrics is provided.**

***Index Terms*—Arabic, speech, rhythm, dialect, classification.**

I. Introduction

Studying speech rhythm in the Arabic language is an important area of research that has seen little advancement compared with other languages such as English, French, and Japanese. Many speech rhythm researchers have focused on the cross-linguistic variation and durational contrast between stressed and unstressed syllables. Rhythm is derived from the repetition of elements perceived as similar. In speech signal, these elements are syllables or stressed syllables in particular.

Speech rhythm is an approach of objectively referring to the temporal distribution of linguistic information in a language. It can represent the degree of closeness or relatedness among speech units at different levels. Languages are generally classified as stress-, syllable-, or mora-timed. Stress-timed languages such as Arabic, English, and Swedish have a near-equal interval between stresses, syllable-timed languages such as Spanish, French, and Italian have a near-equal interval between syllables, and mora-timed languages such as Japanese have equal spacing between mora [1].

Ramus et al. [2] proposed a new measurement method that can be used to classify any language or dialect into one of the three classes mentioned above. Their method is based on the duration of adjacent consonants/vowels and the intervals between them. The speech of eight languages is segmented into vocalic and consonantal intervals, and rhythm metrics such as Interval Measures (IMs) are computed. IMs are defined as ΔC, the standard deviation of the duration of consonantal intervals; ΔV, the standard deviation of the duration of vocalic intervals; and %V, the percentage of total vocalic intervals to the entire actual speech duration in the sentence. Ramus et al. have shown that stress-timed languages have a relatively low %V and high ΔC, while mora-timed languages have a high %V and low ΔC. However, the %V–ΔC values of syllable-timed languages are between these two extremes. In other related studies, Low et al. [3] introduced another metric known as the Pairwise Variability Index (PVI) to account for local differences in speaking rates. It measures the durational variability between successive pairs of vocalic/consonantal intervals in a given utterance.

Based on the corpus being used [6], there are nine main dialectal regions in the Kingdom of Saudi Arabia (KSA). Riyadh region is located at the center of KSA and on the Najd plateau; its capital is Riyadh, which is also the capital of KSA. Another important region is Al Qassim, which is located in the north of the Najd plateau, and its capital is Buraidah. Makkah region is the most popular region; it is located in the western region of KSA and has an extended coastline. Its capital is Makkah and its largest city is Jeddah, which is a coastal city on the Red sea. Tabuk region is also one of the most important regions of KSA; it is located along the north-west coast of KSA and its capital is Tabuk. The Northern Borders (NB) Region is the least populated region of KSA and its capital is Arar. Riyadh, Qassim, Makkah, Tabuk, and NB have different Arabic dialects that can easily be differentiated on hearing a short sentence in their normal spontaneous conversations.

The rest of this paper is structured as follows: in Section

978-1-7281-6535-6/20/$31.00 ©2020 IEEE

II, we present the selected relevant work; in Section III, we present a description of the rhythm metric experiment in the SAAVB corpus. The experiments for the significance of rhythm metrics in the five different regions are reported and analyzed in Section IV. In Section V, we present the results and general discussion of the study. Finally, in Section VI, we present the conclusions of our research.

II. Related Work and Objectines

Altuwaim et al. [4] investigated the efficacy of rhythm metric measurements in classifying only two Saudi dialects. Various metrics such as ΔV, ΔC, and %V were calculated, and the results showed that speech rate could be clearly identified using ΔC. In addition, the hypothesized rhythm classes could be classified using ΔC and %V metrics. ΔC, %V, and ΔV could differentiate between question and statement sentences. In [5], Alotaibi et al. studied the ability of rhythm IMs, namely, ΔV, ΔC, and %V, to classify speakers based on their gender and/or social environments. The results showed that the ΔV metric was able to classify five classes of speakers.

In this study, we aim to investigate the speech rhythm of the five dialectal regions of KSA outlined in SAAVB [6], the relation between different IMs and their capabilities of classifying the speakers' native origin, gender (male/female), and type of sentence. In SAAVB, we have two types of sentences, the first is type 42, which is the statement type and the second is type 41, which is a question sentence.

III. Experimental Setup

A. SAAVB Corpus Analysis

The Saudi Accented Arabic Voice Bank (SAAVB) corpus [6] was used for all experiments in this study. This corpus was designed and built by the Institute of Computer and Electronics Research at King Abdulaziz City for Science and Technology (KACST) in Saudi Arabia. The SAAVB corpus is a set of Arabic speech audio files and their related materials such as scripts and transcriptions that represent native Arabic Saudi speakers from all the cities around Saudi Arabia. It is rich in terms of both speech sound content and speaker diversity within Saudi Arabia.

SAAVB is a collection of audio files vocalized by 1033 speakers with a Saudi accent, but the read sentences were created based on Modern Standard Arabic (MSA) rules. The SAAVB corpus is composed of two types of sentences, type 41, a question sentence and type 42, a statement sentence. The number of utterances by each speaker is 59. There are 333 audio files uttered by speakers from Riyadh region, 84 male speakers and 83 female speakers for type 41, and 86 male speakers and 80 female speakers for type 42. There are 305 utterances in the Makkah region dialect, 76 male speakers each for type 41 and type 42, 79 female speakers for type 41, and 74 female speakers for type 42.

There are 159 audio files vocalized by speakers from Qassim region, 40 male speakers each for type 41 and 42, 40 female speakers for type 41, and 39 female speakers for type 42. For Tabuk region, there are 107 audio files, 24 female speakers each for type 41 and 42, 29 male speakers for type 41, and 30 male speakers for type 42. Finally, there are 112 audio files for the Northern Borders (NB) Region, 28 male speakers each for type 41 and type 42, and 28 female speakers each for type 41 and type 42. The same was applied to all remaining four dialectal regions. The content of SAAVB was verified internally and externally by IBM Cairo [6]. The size of the sample taken for this study is 15 audio files of male and female speakers for each sentence type and for all the five regions under consideration.

B. Rhythm Metrics (RMs)

Many previous studies have proposed a number of rhythm metrics to identify differences in syllable structure, vowel reduction, and stress-based lengthening. These metrics consist of IMs including ΔV, ΔC, and %V [2]; PVIs including nPVI-V, rPVI-C, nPVI-VC, and rPVI-VC; VarcoV, VarcoC, and VarcoVC [3].

The IMs are based on the segmentation of the speech signal into vocalic and consonantal intervals. All neighboring consonants/vowels are considered as one interval even if they belong to different syllables [3]. The speech signal is segmented, and then ΔV, ΔC, and %V are calculated. The definitions of these metrics are given as follows [8]:

- %V (percentage of vocalic intervals): the total duration of all vocalic sequences divided by the total duration of the utterance.
- ΔV: the standard deviation of individual vocalic sequence duration within each sentence.
- ΔC: the standard deviation of consonantal sequence duration within each sentence.

C. Proposed Experiments

In previous studies, ΔC, ΔV, and %V were significantly useful in differentiating between languages, dialects, speakers, emotions, and genders. In this study, we used the SAAVB corpus of five dialectal regions, Riyadh, Makkah, Qassim, Tabuk, and NB, to investigate the relationship between rhythm IM metrics, namely, ΔV, ΔC, and %V, with the aim of correctly classifying the regions, speakers, gender, and sentence types.

IV. Analysis and Results

Tables 1, 2, and 3 present the average percentage of vocalic intervals (%V), average standard deviation of vocalic intervals (ΔV), and average standard deviation of consonantal intervals (ΔC) in all five dialectal regions under investigation with respect to gender and the two sentence types. All the time units are given in milliseconds.

The average metric values for each region for the speakers of both genders and type 41/42 were calculated and plotted for easier evaluation in Figs. 1, 2, and 3.

By considering the three used IM parameters, we can obtain all extreme readings to draw the best conclusions from research experiments. Regarding percentage of vocalic intervals, %V, the difference between type 41 (question type) and type 42 (statement type) is very clear, as can be seen from

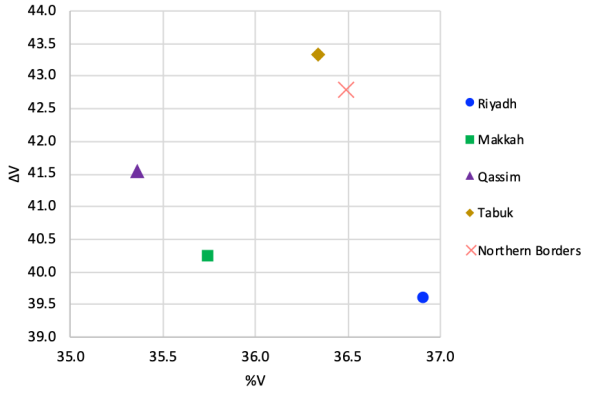

Fig. 1. Distribution of dialects along the %V and ΔV dimensions.

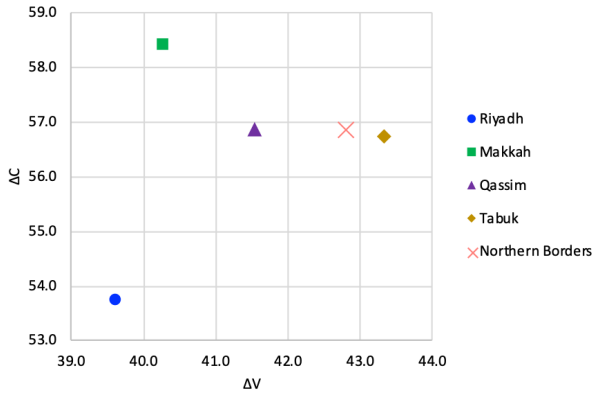

Fig. 2. Distribution of dialects along the ΔV and ΔC dimensions.

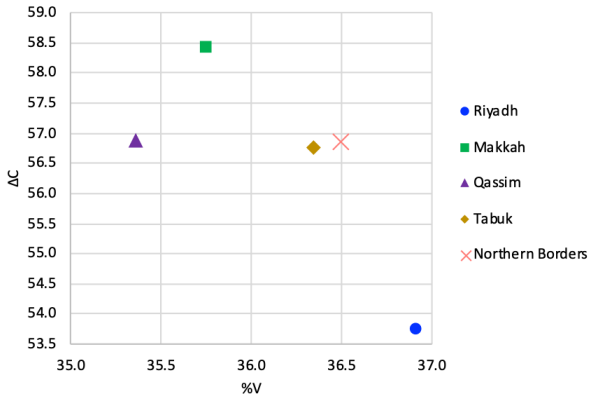

Fig. 3. Distribution of dialects along the %V and ΔC dimensions.

TABLE I
%V METRIC VALUES OF FIVE DIALECTAL REGIONS (ms)

Region	Type 41		Type 42	
	Male	Female	Male	Female
Riyadh	34.54	36.83	37.44	39.22
Makkah	34.69	36.55	39.72	39.26
Qassim	34.90	34.66	35.00	36.85
Tabuk	34.29	36.06	37.03	38.08
NB	33.27	37.57	35.82	39.29

TABLE II
ΔV METRIC VALUES OF FIVE DIALECTAL REGIONS (ms)

Region	Type 41		Type 42	
	Male	Female	Male	Female
Riyadh	27.68	30.42	54.27	53.37
Makkah	33.95	33.61	52.26	59.59
Qassim	31.27	30.53	53.17	51.17
Tabuk	31.08	32.73	53.98	55.58
NB	27.98	33.43	52.79	56.97

the tables and figures. This is caused by the language content of the two sentences. In other words, the type 42 sentence contains more vowels than that of type 41.

Also, the contribution of the parameter %V is clear regarding both gender and dialectal region. There are huge differences in the averages of %V regarding gender types and regions. There is a significant difference between male and female speakers; where male speakers have a lower vowel percentage than female speakers. This may be viewed as either female speakers trying to extend durations in pronouncing vowels or trying to compress them for consonants.

As a specific outcome, Qassim male and female speakers have almost the same vowel duration percentage, %V, for type 41. Similarly, Makkah male and female speakers have almost the same average %V for type 42. The average %V values for each region for all speakers, i.e., both males and females and type 41/42 show that Qassim region has the shortest %V but Riyadh has a greater average. Also, %V averages show a noticeable closeness in Tabuk and NB dialects.

The average ΔV values show that the speakers in Riyadh region have the least vocalic intervals among all the other dialectal regions. On the other hand, those in Tabuk have the highest vocalic intervals. Additionally, the averages of ΔV show a high similarity between Tabuk and NB dialects. Based on the percentage of vocalic intervals (%V) and standard devi-

TABLE III
ΔC METRIC VALUES OF FIVE DIALECTAL REGIONS (ms)

Region	Type 41		Type 42	
	Male	Female	Male	Female
Riyadh	50.09	59.51	55.09	45.18
Makkah	59.85	65.29	49.56	53.28
Qassim	63.68	61.03	53.88	48.98
Tabuk	60.24	67.89	50.54	48.38
NB	61.67	64.26	55.37	46.16

3

ation of vocalic intervals ΔV, there is a significant dissimilarity between Type 41 and Type 42. For Riyadh, Makkah, and Tabuk dialects, there is a significant difference between the male and female speakers of type 41. Moreover, a similarity between the male and female speakers of type 41 from Qassim region is seen. For Makkah dialect, the male and female speakers of type 41 have close vocalic intervals. However, for type 42, there is a significant difference between them.

As shown in Fig. 3, %V–ΔC plot shows that Riyadh dialect is isolated from all other dialects and has a low ΔC, which means that its speakers are less spread around the average. In comparison, Riyadh has the lowest ΔC value (least spreading consonantal intervals around the average) and Makkah has the highest (widely spreading consonantal intervals around the average). Moreover, Qassim, Tabuk, and NB have similar ΔC. Additionally, ΔC values show a significant difference between type 41 and type 42 as well as between male and female speakers.

The average metric values for all male and female speakers in the five dialects and of type 41/42 are calculated and plotted for easier evaluation in Figs. 4, 5, and 6.

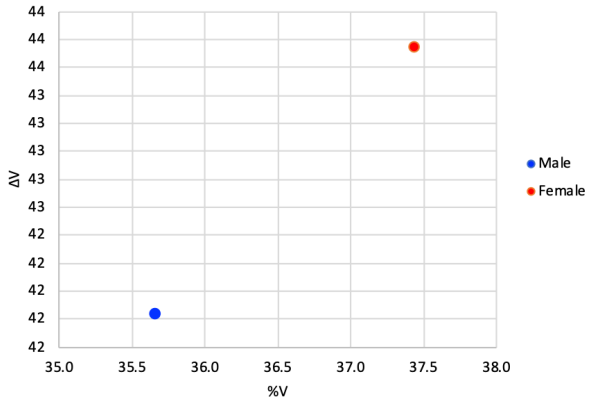

Fig. 4. Distribution of dialects along the %V and ΔV dimensions.

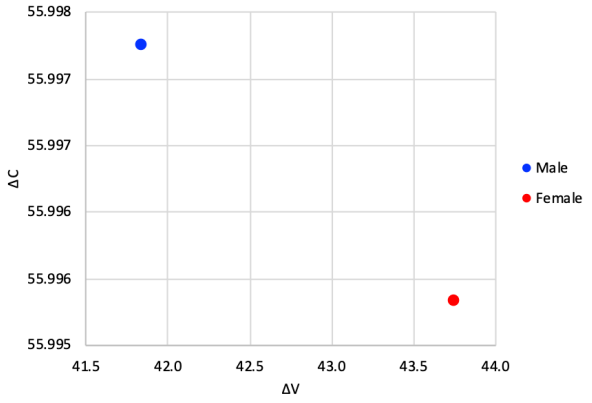

Fig. 5. Distribution of dialects along the ΔV and ΔC dimensions.

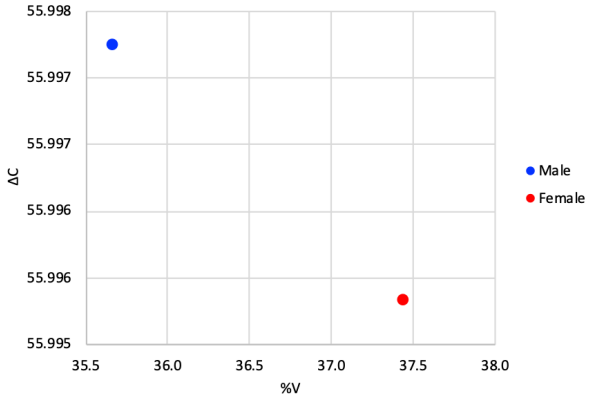

Fig. 6. Distribution of dialects along the %V and ΔC dimensions.

V. General Discussion

This study investigated the speech rhythm of five major Saudi dialectal regions, the relation between their IM metrics, and their capability of classifying speakers based on their native dialectal region, gender, and type of vocalized sentence. Generally, for all the three metrics, there was a large difference between question (type 41) and statement (type 42) sentences and between male and female speakers. In particular, we can conclude from all the results that Tabuk and NB dialects are close. In addition, these two dialects have low similarities when compared to all the other three dialects. All plots and tables verify this outcome in addition to the fact that these two dialects are geographically neighbors and are both located in the north of Saudi Arabia. Tabuk and NB regions are also adjacent to the states of Jordan and Iraq, and their dialects may be influenced by these international dialects. Another main outcome is the relatively high similarity between Makkah and Qassim dialects. These two dialects are geographically not far from each other. Additionally, many people from Qassim have moved to Makkah as it is where the places of worship, Hajj and Omrah, are located. In addition to the above, we noted that the Riyadh dialect is unique compared to the other four dialects. From all the figures, the Riyadh dialectal pattern is not close to any of the other four dialects. This is probably because Riyadh is the capital of the KSA and a variety of people of different nationalities with different cultures, dialects, and other native languages work there.

Moreover, there is a large difference between male and female speakers; female speakers have high ΔV and %V values. This may be affected by vowel formant frequencies, which are located at higher frequencies [7]. There is a significant similarity in ΔV, ΔC, and %V between Tabuk and NB dialects.

VI. Conclusion and Future Work

This study investigated the rhythm metrics of five Saudi dialects using the SAAVB corpus. These metrics were proposed for measuring rhythmic differences between the five dialects.

From this study, the rhythm metrics captured the differences between the five dialects investigated, which were affected

by factors such as gender and sentence types. All rhythm metrics had significant differences between male and female speakers and also between type 41 and type 42 sentences. The measure of %V showed a similarity between Qassim male and female speakers of type 41 and between Makkah male and female speakers of type 42. The ΔV values showed a similarity between Qassim male and female speakers of type 41 and between Makkah male and female speakers of type 41. Moreover, the Riyadh dialect was distinct from the other four dialects; it had low ΔC, which means that it has a faster rate than other dialects. Finally, there was a significant similarly in the ΔV, ΔC, and %V of Tabuk and NB dialects.

In the future, we will investigate speech rhythm metrics for the nine Saudi dialects using the PVI and calculate ANOVAs using the SAAVB corpus.

ACKNOWLEDGMENT

The authors thank the Deanship of Scientific Research and RSSU at King Saud University for their technical support.

REFERENCES

[1] S. Sid-Ahmed et al., "A Rhythm-Based Analysis of Arabic Native and Non-Native Speaking Styles," Int. J. Signal Process. Syst., vol. 1, pp. 202–207, Dec. 2013.

[2] F. Ramus, M. Nespor, and J. Mehler, "Correlates of linguistic rhythm in the speech signal," Cognition, vol. 73, no. 3, pp. 265–292, Dec. 1999.

[3] L. E. Ling, E. Grabe, and F. Nolan, "Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English," Lang. Speech, vol. 43, no. 4, pp. 377–401, 2000.

[4] A. A. Altuwaim, Y. A. Alotaibi, and S.-A. Selouani, "Investigation into the speech rhythm of two Saudi dialects using the SAAVB corpus," in 2014 6th International Symposium on Communications, Control and Signal Processing (ISCCSP), 2014, pp. 632–635.

[5] Y. A. Alotaibi, A. H. Meftah, S.-A. Selouani, and Y. M. Seddiq, "Speaker environment classification using rhythm metrics in Levantine Arabic dialect," in 2014 9th International Symposium on Communication Systems, Networks Digital Sign (CSNDSP), 2014, pp. 706–709.

[6] M. Alghamdi, F. Alhargan, M. Alkanhal, A. Alkhairy, M. Eldesouki, and A. Alenazi, "Saudi Accented Arabic Voice Bank," J King Saud Univ Comput Inf Sci, vol. 20, pp. 45–64, Jan. 2008.

[7] S. P. Whiteside, "Sex-specific fundamental and formant frequency patterns in a cross-sectional study," J. Acoust. Soc. Am., vol. 110, no. 1, pp. 464–478, Jul. 2001.

[8] J. M. Liss et al., "Quantifying speech rhythm abnormalities in the dysarthrias," J. Speech Lang. Hear. Res. JSLHR, vol. 52, no. 5, pp. 1334–1352, Oct. 2009.