Testing a Large Corpus of Natural Standard Arabic for Rhythm Class

Liz Dockendorf1, Dalal Almubayei1, Matthew Benton1

1 Department of Linguistics and TESOL, The University of Texas at Arlington, United States
liz1695@yahoo.com, almubayei@yahoo.com, matt@mjbenton.com

Abstract

Previous studies using acoustic correlates to measure speech rhythm have used small samples of audio and a limited number of speakers. Few have included standard Arabic in the analysis. This study uses Arabic news broadcast along with data output from an automatic speech recognizer time-aligned transcript to test over 50 minutes of speech by 46 speakers. The results show that Arabic, like English and Chinese news broadcasts, exhibits a range of individual speaker rhythm, but its average clearly distinguishes it from the other two on the rhythm continuum. A large number of speakers also makes the average more reliable.

Index Terms: speech timing patterns, Arabic rhythm, Language rhythm, stress-timing, Standard Arabic

1. Introduction

Many of the recently published speech rhythm (speech timing pattern) studies have attempted to measure speech rhythm using a few sentences of data from each language and measures such as standard deviation of consonant duration ($\Delta \overline{C}$) and vowel duration percentage of each sentence (%V) as in [1] or pairwise variability (PVI) [2] of segment duration. These two studies just mentioned are particularly relevant to the study conducted by [3]. The current study is a continuation of [3]. It differs from other speech rhythm studies by incorporating a larger number of speakers and subsequently more data from each language.

1.1. Previous research on Arabic rhythm

A previous study [4] examined speech timing and rhythm in Arabic (AR) dialects using the methods of [1], and found a distinction in the rhythm characteristics between the Eastern and Western dialects. It did not examine Modern Standard Arabic.

2. Methods and Results

This study attempts to complement the results of English (EN) and Chinese (CH) in [3] by adding calculations of acoustic measurements for AR (obtained from the output of speech recognition software [5]) to the previous research and by using more naturally occurring data from broadcast news. The AR news data consists of 46 (24M, 22 F) Arab speakers speaking modern standard AR (as determined by a native speaker) from Voice of America News Broadcast (out of Washington D.C.). All speakers are "reporters" (those who work for the broadcasting company).

2.1. Comparing Arabic to work on large corpora

As shown in Figure 1, the individual speaker points in the AR data (like that in [3] for CH and EN) occurs in a continuum-like pattern. The results of SPSS ANOVA show that there is a significant statistical difference between AR, CH, and EN for all values: %V, $\Delta \overline{V}$, $\Delta \overline{C}$, VnPVI, VrPVI, and CrPVI (p<.001 at 95% confidence level). An independent T-Test showed there was only significant difference between genders for AR reporters on $\Delta \overline{C}$.

3. Conclusion

This study argues (like [3]) that rhythmic distinctions hold when using large corpora of speech that include a large number of speakers. It also shows that Standard Broadcast AR has a continuous nature over individual speakers but a discrete class compared to American EN and Mandarin CH. Finally, the results from this test show that Standard AR is more likely stressed timed using the PVI metric.

4. References