Available online at www.sciencedirect.com
ScienceDirect
Computer Speech & Language xxx (2017) xxx-xxx
www.elsevier.com/locate/csl

Application of the pairwise variability index of speech rhythm with particle swarm optimization to the classification of native and non-native accents☆

Q2 Soumaya Gharsellaoui[a], Sid-Ahmed Selouani[b,*], Wladyslaw Cichocki[c], Yousef Alotaibi[d], Adel Omar Dahmane[a]

[a] Department of Electrical and Computer Engineering, Université du Québec, Trois-Rivières, Québec, Canada
Q3 [b] Université de Moncton, Campus de Shippagan, Shippagan, New Brunswick, Canada
[c] Department of French, University of New Brunswick, Fredericton, New Brunswick, Canada
[d] Department of Computer Engineering, King Saud University, Riyadh, Saudi Arabia

Received 18 October 2015; received in revised form 15 January 2017; accepted 16 October 2017
Available online xxx

Abstract

This paper presents a technique that applies the pairwise variability index (PVI), a rhythm metric that quantifies variability in speech rhythm, to the classification of speech varieties. The technique combines the Particle Swarm Optimization (PSO) algorithm with a generalization of several rhythm metrics that are based on the PVI. The performance of this optimization-oriented classification is compared with classification that uses conventional (both PVI-based and interval-based) rhythm metrics. Application is made to the classification of native and non-native Arabic speech using data are from the West Point Arabic Speech Corpus; experiments are based on segmental durations and use Support Vector Machine (SVM) classification. Results show that the optimization-oriented classification provides a better discrimination between native and non-native speech varieties than classification based of the conventional rhythm metrics. When added to different combinations of these conventional metrics, the optimization-oriented procedure consistently improves classification rates.

Keywords: Speech rhythm; Rhythm metrics; Pairwise variability index; Classification; Modern standard Arabic; Particle swarm optimization; Non-native accent

1. Introduction

Since the advent of speech rhythm metrics, numerous studies have used these measures to classify languages, to identify different accents and other language features, and to capture the timing properties of speech. There is a general intuition that perceived differences in rhythm can be measured acoustically, and that rhythm metrics can offer

☆ This paper has been recommended for acceptance by Roger K. Moore.

* Corresponding author.
E-mail address: soumaya.gharsellaoul@umoncton.ca (S. Gharsellaoui), sid-ahmed.selouani@umoncton.ca (S.-A. Selouani), cicho@unb.ca (W. Cichocki), yaalotaibi@ksu.edu.sa (Y. Alotaibi), dahmane@uqtr.ca (A.O. Dahmane).

http://dx.doi.org/10.1016/j.csl.2017.10.006
0885-2308/ 2017 Elsevier Ltd. All rights reserved.

an objective and practical approach to this question. Studies of speech processing systems have applied rhythm metrics in order to identify languages and language varieties (Rouas et al., 2005). Related areas that make use of these metrics include research on second language phonology (Gut, 2012; Obin and Polyanskaya, 2015), on dialects (Ferragne and Pellegrino, 2008; Tan and Low, 2014) and on motor speech disorders (Liss et al., 2009; Lowit, 2014). Studies of musical rhythm analysis (Patel and Daniele, 2003; Toussaint, 2012) also find that these metrics are useful for describing musical rhythm complexity and for distinguishing musical genres.

Underlying these applications of rhythm metrics is the notion that rhythm is based on durational variability. Speech rhythm is defined as patterned variation in the duration of intervals such as vowels and consonants. It is noteworthy that this is only one view of rhythm. As discussed in an overview article by Turk and Shattuck-Hufnagel (2013), there is a wide range of definitions for this term; these include a perceptual concept based on regular prominences that we perceive, a surface periodicity in the occurrence of some specific linguistic constituent in the speech signal, and an abstract structuring that reflects grouping and prominence structure. An important advantage of the view of rhythm as patterns of durational variability is that it provides a clear statement that identifies specific units that can be measured objectively and compared.

There are two main families of rhythm metrics, and both use measures of speech timing that are based on the segmentation of the signal speech into vocalic and consonantal intervals. Pairwise variability indices (PVI) measure the difference in duration between immediately consecutive intervals and average these differences over an utterance (Grabe and Low, 2002). On the other hand, interval-based metrics quantify variation in individual vocalic or consonantal interval durations (Ramus et al., 1999) by looking at their standard deviations (deltaV and deltaC) over an utterance. A related measure looks at the proportion of vowel durations (%V) over the utterance. The essential difference between the two families is that PVIs measure durational contrast while interval-based metrics measure overall variability with no attention to the order of intervals in an utterance. Both types of indices have normalized measures that reduce the effect that speech tempo can have on raw measures. For the PVI-based measures, normalization is based on the mean duration of the two intervals being compared; researchers tend to use raw durations when measuring consonantal effects (rPVI-C) but normalized measures for vowels (nPVI-V). For interval-based measures, the rate-normalized indices of vowel and consonant intervals (VarcoV and VarcoC) are calculated by dividing the standard deviations of the interval durations by their mean (Dellwo, 2006). A description of the metrics used in this paper is given below in Section 2.

Researchers have noticed that using normalized metrics can improve discrimination among speech varieties. Wiget et al. (2010) find that normalized measures of variability of vocalic intervals perform better in discriminating between languages than consonantal or non-normalized measures; they recommend that a combination of at least two measures be used. In their study of the classification of five languages, Loukina et al. (2011) report that no single rhythm metric or set of metrics provides an optimal separation across all pairs of these languages. The authors find that three measures of rhythm metrics are necessary to achieve such a separation; they also report that normalized vocalic measures perform above chance in correctly sorting all five languages. Liss et al. (2009) find that variants of PVI and Varco metrics successfully discriminate healthy control speakers from speakers with dysarthria, although the best distinguishing metric depends on the type of disorder under investigation.

The focus of our research is on how to improve rhythm-based classification of speech varieties, with attention to PVI-based rhythm metrics. Some of the issues that researchers have raised concerning PVI metrics deal with normalization. Bertinetto and Bertini (2008) propose the Control/Compensation Index (CCI), which takes into account the degree of phonotactic complexity found in natural languages. The CCI metric consists of "relativizing" (that is, normalizing) the PVI algorithm to the number of segments (or phonemes) that make up each vocalic or consonantal interval. This metric offers a classification of languages that differs somewhat from traditional stress-timed vs. syllable-timed language classes. In their applications of PVI to music studies, London and Jones (2011) find that the nPVI metric discriminates among most families of musical rhythms; however, they observe that the normalization may in fact be distorting effects such as *accelerandi* or *ritardandi* because dramatic shifts in durational pacing may only result in very small changes in nPVI. They suggest that some form of "denormalizing" the nPVI formula may help to capture distinctions in timing that are musically salient.

In this paper, we propose a generalization of PVI-based metrics that allows for several kinds of normalization of these metrics. This generalization has the form of an index called the optimization-oriented pairwise-variability index (O-PVI). Combined with the particle swarm optimization algorithm this index provides a technique that can improve discrimination among speech varieties. From the perspective of applications, we apply this optimization-