Journal of King Saud University – Computer and Information Sciences (2013) 25, 43–49

King Saud University

Journal of King Saud University – Computer and Information Sciences

www.ksu.edu.sa
www.sciencedirect.com

Assessment of dysarthric speech through rhythm metrics

H. Dahmani [a,e,*], S.-A. Selouani [b], D. O'shaughnessy [a], M. Chetouani [c], N. Doghmane [d]

[a] *INRS-EMT, Université du Québec, Canada*
[b] *Université de Moncton, Campus de Shippaga, Canada*
[c] *Université Pierre and Marie Curie, France*
[d] *Université Mokhtar Badji, Annaba, Algeria*
[e] *Université M'sila, M'sila, Algeria*

Received 21 October 2011; revised 19 April 2012; accepted 29 May 2012
Available online 9 June 2012

KEYWORDS

Dysarthria;
Rhythm;
Pairwise variability index;
Acoustical analysis;
Timing;
Nemours database;
Dysarthric severity

Abstract This paper reports the results of acoustic investigation based on rhythmic classifications of speech from duration measurements carried out to distinguish dysarthric speech from healthy speech. The Nemours database of American dysarthric speakers is used throughout experiments conducted for this study. The speakers are eleven young adult males with dysarthria caused by cerebral palsy (CP) or head trauma (HT) and one non-dysarthric adult male. Eight different sentences for each speaker were segmented manually to vocalic and intervocalic segmentation (176 sentences). Seventy-four different sentences for each speaker were automatically segmented to voiced and non-voiced intervals (1628 sentences). A two-parameters classification related to rhythm metrics was used to determine the most relevant measures investigated through bi-dimensional representations. Results show the relevance of rhythm metrics to distinguish healthy speech from dysarthrias and to discriminate the levels of dysarthria severity. The majority of parameters was more than 54% successful in classifying speech into its appropriate group (90% for the dysarthric patient classification in the feature space (%V, ΔV)). The results were not significant for voiced and unvoiced intervals relatively to the vocalic and intervocalic intervals (the highest recognition rates were: 62.98 and 90.30% for dysarthric patient and healthy control classification respectively in the feature space (ΔDNV, %DV)).

1. Introduction

Dysarthria covers various speech impairments resulting from neurological problems and it probably represents a significant proportion of all acquired neurological communication disorders (as cited in Ziegler and von Cramon (1986)). These disorders are linked to the disturbance of brain and nerve stimuli of the muscles involved in the production of speech. Ultimately they induce disturbances in the strength, speed, range, tone, steadiness, timing, or accuracy of movements necessary for prosodically normal, efficient and intelligible speech (Liss et al., 2009; Yunusova et al., 2008). All types of dysarthria affect the articulation of consonants leading to slurring speech. Vowels may as well be distorted in very severe dysarthria. Rhythm troubles may be the most common characteristic of various types of dysarthria. Many studies state that most dysarthric patients

* Corresponding author.
E-mail address: anout_1999@yahoo.fr (H. Dahmani).
Peer review under responsibility of King Saud University.

ELSEVIER Production and hosting by Elsevier

1319-1578
http://dx.doi.org/10.1016/j.jksuci.2012.05.005

have slow speaking rates with long vowel and consonant segments as compared to standard control samples (Liss et al., 2009; Yunusova et al., 2008).

The present paper focuses on the assessment of rhythmic disturbance in dysarthria caused by cerebral palsy and head trauma. Cerebral palsy refers to a variety of developmental neuromuscular pathologies, occurring in three main forms: spastic, athetoid, and ataxic, associated with bilateral lesions of upper motor neuron pathways that innervate relevant cranial and spinal nerves. Dysarthria severity can be indexed in several ways, but quantitative measures usually consider features such as intelligibility and speaking rate. Disturbance of rhythm in the speech flow process is one of the important factors in dysarthric abnormalities (Liss et al., 2009). Even if the rhythm is identified as the main feature that characterizes dysarthria, assessment methods are mainly based on perceptual evaluation measures. Despite their numerous advantages that include the ease of use, low cost and clinicians' familiarity with related procedures, perceptual-based methods suffer a number of inadequacies and aspects that affect their reliability. These methods also lack evaluation protocols that may help standardization of judgments between clinicians and/or evaluation tools. Therefore, the aim of this work is to quantify rhythm abnormalities in the dysarthric speech by using the rhythm metrics developed recently in the language identification domain (Arvaniti and Rhythm, 2009).

This paper is organized as follows. Section 2 gives some definitions related to the rhythm metrics. Section 3 presents our method including the speech material, subjects and procedures used throughout experiments. In Section 4, we discuss the relevance of the rhythm metrics to assess the severity of dysarthrias. We describe the Gaussian Bayes classification system and its results in section 5. Section 6 concludes this paper.

2. Rhythm metrics

Rhythm metrics are based on acoustic measures of the duration of vocalic and consonantal intervals in continuous speech, they take into account variability in these durations, and they can be calculated in both raw and rate-normalized forms. A list of rhythm metrics used in our experiments is given at the end of this section. Grab and Low calculate durational variability in successive acoustic-phonetic intervals using Pairwise Variability Indices (PVI) (Grabe and Low, 2002). The raw Pairwise Variability Index (rPVI) is given in Eq. (1):

$$\text{rPVI} = \sum_{k=1}^{N-1} |d_k - d_{k+1}|/(N-1) \tag{1}$$

where dk is the length of the kth vocalic or intervocalic segment and N the number of segments.

A normalized version of the PVI index (noted nPVI) is defined by:

$$\text{nPVI} = \sum_{k=1}^{N-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| /(N-1) \tag{2}$$

Ramus et al. (1999) based their quantitative approach of speech rhythm on purely phonetic characteristics of the speech signal. They measured vowel durations and the duration of intervals between vowels. They computed three acoustic correlates of rhythm from the measurements:

(a) %V: the proportion of time of vocalic intervals in the sentence;
(b) ΔV: the standard deviation of vocalic intervals;
(c) ΔC: the standard deviation of inter-vowel intervals.

Ramus et al. (1999) found that a combination of %V and ΔC provided the best acoustic correlate of rhythm classes. Our goal is to use these metrics in order to distinguish between the healthy and dysarthria speakers and to assess the speech intelligibility since the alterations of rhythm may also impact speech intelligibility. For each dysarthric sentence of each speaker, we have measured the durations of the vocalic, consonantal, voiced and unvoiced segments. In addition to the Vocalic-rPVI, Vocalic-nPVI, Intervocalic-rPVI, Intervocalic-nPVI, %V, ΔC, and ΔV, we computed the %DV, the duration of voiced intervals expressed in percent, ΔDV and ΔDNV, the standard deviation of voiced and non-voiced intervals, respectively.

3. Method

3.1. Speech material

Nemours is one of the few databases of recorded dysarthric speech. It contains records of American patients suffering different types of dysarthrias (Polikoff and Bunnell, 1999; James et al., 1996). The evaluation methodology followed in Nemours is inspired by the work of Kent et al. (1989). The test consists of a list of words from which four words are selected. The patient is supposed to listen to these words and repeat them aloud. The full set of stimuli consists of 74 monosyllabic names and 37 bi-syllabic verbs embedded in short nonsense sentences. Each Speaker pronounced 74 different sentences. Sentences have the following form: THE noun 1 IS verb-ING THE noun 2. The recording session was conducted by a speech pathologist considered as the healthy control (HC). The speech waveforms were sampled at 16 kHz and 16 bit sample resolution after low pass filtering at a nominal 7500 Hz cutoff frequency with a 90 dB/Octave filter.

3.2. Subjects

The speakers are eleven young adult males with dysarthria caused by cerebral palsy (CP) or head trauma (HT) and one non-dysarthric adult male (the experimenter). Seven speakers have CP, among whom three have CP with spastic quadriplegia and two have athetoid CP, and both have a mixture of spastic and athetoid CP with quadriplegia. The four remaining subjects are victims of head trauma. A two-letter code was assigned to each patient: BB, BK, BV, FB, JF, KS, LL, MH, RK, RL and SC. Thanks to the Frenchay dysarthria assessment scores (see Table 1 and (James et al., 1996; Enderby and Pamela, 1983)), the patients can be divided into three subgroups: one mild, including subjects FB, BB, MH and LL; the second subgroup includes the subjects RK, RL, and JF and the third is severe and includes subjects KS, SC, BV, and BK. The perceptual data and the speech assessment did not take into consideration the too severe case (patient KS) and the too mild case (patient FB).

4. Experiments and results

The mean and the standard deviation of vocalic and consonantal interval durations are given in the Fig. 1. These measures