Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/bbe

Original Research Article

Fully automated speaker identification and intelligibility assessment in dysarthria disease using auditory knowledge

Kamil Lahcene Kadi [a,*], *Sid Ahmed Selouani* [b], *Bachir Boudraa* [a], *Malika Boudraa* [a]

[a] *Faculty of Electronics and Computer Science, University of Sciences and Technology Houari Boumediene, Algiers, Algeria*

[b] *Department of Information Management, University of Moncton, Campus of Shippagan, Shippagan, NB E8S 1P6, Canada*

ARTICLE INFO

Article history:
Received 10 January 2015
Received in revised form
13 September 2015
Accepted 9 November 2015
Available online 27 November 2015

Keywords:
Dysarthria
Speech processing
Auditory cues
GMM
SVM
Hybrid GMM/SVM

ABSTRACT

Millions of children and adults suffer from acquired or congenital neuro-motor communication disorders that can affect their speech intelligibility. The automatically characterization of speech impairment can contribute to improve the patient's life quality, and assist experts in assessment and treatment design. In this paper, we present new approaches to improve the analysis and classification of disordered speech. First, we propose an automatic speaker recognition approach especially adapted to identify dysarthric speakers. Secondly, we suggest a method for the automatic assessment of the dysarthria severity level. For this purpose, a model simulating the external, middle and inner parts of the ear is presented. This ear model provides relevant auditory-based cues that are combined with the usual Mel-Frequency Cepstral Coefficients (MFCC) to represent atypical speech utterances. The experiments are carried out by using data of both Nemours and Torgo databases of dysarthric speech. Gaussian Mixture Models (GMMs), Support Vector Machines (SVMs) and hybrid GMM/SVM systems are tested and compared in the context of dysarthric speaker identification and assessment. The experimental results achieve a correct speaker identification rate of 97.2% which can be considered promising for this novel approach; also the existing assessment systems are outperformed with a 93.2% correct classification rate of dysarthria severity levels.

* *Corresponding author at*: Faculty of Electronics and Computer Science, University of Sciences and Technology Houari Boumediene, 32 El Alia, 16111 Bab Ezzouar, Algiers, Algeria.

E-mail addresses: kkadi@usthb.dz (K.L. Kadi), selouani@umcs.ca (S.A. Selouani), bboudraa@usthb.dz (B. Boudraa), mboudraa@usthb.dz (M. Boudraa).

http://dx.doi.org/10.1016/j.bbe.2015.11.004

1. Introduction

Communication is a multidimensional dynamic process that is necessary to express thoughts, emotions and needs, allowing interaction between people and their environment. Cognition, hearing, speech production and motor coordination are involved in the communication process. If one or more of these aspects is impaired, the communication is disordered [1]. A communication disorder has a large impact on the life quality; it prevents individuals from expressing needs, wants and opinions, it also reduces the capacity to express personality, exercise autonomy and often has an impact on relationships and self-esteem [2]. Therefore, it is necessary to enhance the communication quality of individuals suffering from a verbal communication disability by offering them more possibilities to interact with their environment.

Our research focuses on one of the most common speech communication disorders associated with a neurological impairment called dysarthria. In fact, millions of adults and children through the world are affected by dysarthria which induces a reduction of their speech intelligibility [2,3]. Dysarthria is a motor speech disorder resulting from disturbed muscular control of the speech mechanism, caused by damage to the central or peripheral nervous system [4]. A few causes of dysarthria include Parkinson's disease, head injury, stroke, tumors, muscular dystrophy and cerebral palsy.

Numerous tools and methods have been developed to help dysarthric speakers. Indeed, prominent works were achieved in the field of speech recognition, speech intelligibility enhancement and automatic evaluation. In the last few years, based on the significant and newer Torgo database, Rudzicz developed a dysarthric speech recognition system and improved the intelligibility of dysarthric speech [5,6]. Although, it is worthy to note that the research effort did not focus on further functionalities such as speaker recognition which is increasingly used in various security or identity management systems. It is also important to note that the automated diagnosis and assessment that can assist clinicians to support patients suffering from speech impairments has not received enough attention and research efforts in this field are very sporadic.

This paper presents two original techniques of dysarthric speech processing, the first one is based on a biometric approach for automatic recognition of dysarthric speakers, and the second technique is designed to automatically assess the dysarthric speech with respect to the severity level. Both techniques use relevant features based on Distinctive Auditory-based Cues and Mel-Frequency Cepstral Coefficients (MFCC). The classification performance of the Gaussian Mixture Models (GMMs) and Support Vector Machines (SVMs) are compared in various experimental conditions. The experiments have been carried out by using two important world-class databases of dysarthric speech, namely, the Nemours database of dysarthric speech [7] and the Torgo database of acoustic and articulatory speech [8]. Both databases were used to build and evaluate the developed tools, which will ensure the availability of a large amount of data and rational diversity in recordings.

The original contributions reported in this paper can be summarized by the following three aspects:

(i) An original dysarthric speaker recognition system is proposed in a context where the individuals affected by verbal communication disorders are excluded from speech-enabled biometric solutions.
(ii) Computational models of auditory perceptual knowledge are proposed to improve the effectiveness of dysarthria assessment.
(iii) A new global assessment score is proposed for the Torgo database based on the second edition of Frenchay Dysarthria Assessment (FDA-2).

The remaining of the paper is structured as follows. Section 2 describes the acoustic analysis. Sections 3 and 4 present the automatic dysarthric speaker recognition system, and the technique of dysarthric speech assessment, respectively. In Section 5, the experiments and their outcomes are presented and discussed. Section 6 contains our concluding comments.

2. Acoustic analysis

The extraction of reliable parameters to represent the speech utterance waveform is an important issue in pattern recognition. This extraction process aims at extracting the relevant information contained in the speech signal while excluding the irrelevant part. Several features can be used as input parameters in speaker recognition and disorder characterization systems. Among these features, we can cite Linear Predictive Coding (LPC) coefficients, MFCC, short-time spectral envelope, short-time energy, zero crossing rate. It is important to mention that numerous studies showed that the use of human hearing properties can provide a potentially useful front-end speech representation [9].

Our acoustical analysis method consists of consolidating different sources of information about the speech signal that could be missing if just one type of features (e.g. MFCC) is used to represent the speech utterance waveform. The first information source consists of the conventional MFCCs, while the second source consists of acoustic cues derived from hearing phenomena studies.

2.1. *Cepstral acoustic features*

Short-term MFCC is an appropriate parameterization approach for dysarthric speech. The MFCCs have been often used in the signal processing field applied for speech disorder, like the recognition of dysarthric speech or speech disorders classification [10]. The Mel-scale introduced by Davis and Mermelstein is a mapping from a linear to a nonlinear frequency scale based on human auditory perception [11]. The Mel-scale approximation is:

$$Mel(f) = 2595\log_{10}\left(1 + \frac{f}{700}\right) \quad (1)$$

where f represents the linear frequency scale.

To compute the MFCCs, a discrete cosine transform is applied to the outputs of M critical band-pass filters. These