1. Introduction

Dysarthria is a neurological impairment of controlling the motor speech articulators that compromises the speech signal. Mel-Frequency Cepstral Coefficients (MFCCs) have been proven to be an appropriate representation of dysarthric speech, but the question of which MFCC-based feature set represents dysarthric acoustic features most effectively has not been answered. Moreover, an speaker independent (SI) ASR model for users with dysarthria is required in order to recognize more accurately the speech of a wide range of users with speech disabilities; such ASR system can then be accessed by speech-disabled people using public services. The objectives of this study are:

1. To identify the most effective MFCC-based feature set for representing dysarthric acoustic signals in order to provide an ANN-based dysarthric ASR model.
2. To study the application of ANNs in a fixed-length, isolated-word SI dysarthric ASR system.

2. Materials and Participants

We used speech materials provided by the UA-Speech Database for dysarthria, which was produced by the University of Illinois. Table 1 provides more information about the subjects with dysarthria used in this research.

3. Dysarthric ANN-based ASR Model

22 frames of MFCC features for each digit utterance was selected in which the frame size was 16.25 ms with a sliding hamming window of 81 ms. The number of frames was selected to match the maximum length of the utterances (1782 ms) provided by the database.

4. Evaluation

Experiment 1: Identifying the Best-Performing Set of MFCC Parameters

We trained 28 ANN-based speech recognisers (speaker-dependent) for seven dysarthric subjects. Four SD speech recognisers were provided for each subject, each of which was trained with a different set of MFCC parameters. These sets are:

a) 12 MFCCs, (i.e. mel cepstrum with 12 coefficients)
b) 12 MFCCs first derivatives
c) 12 MFCCs second derivatives
d) Set a = Set b + Set c

Experiment 2: Speaker-Independent Dysarthric ANN-based ASR

The speech materials of 13 dysarthric subjects (Table 1) were considered for training. The evaluation subjects were M07 (28% speech intelligibility), M05 (speech intelligibility 58%), and F05 (speech intelligibility 95%).

5. Results

Acknowledgment

This research is supported by UM High Impact Research Grant UMC/HIR/MOHE/FC/SIT/05 from the Ministry of Higher Education Malaysia.
Artificial neural networks as speech recognisers for dysarthric speech: Identifying the best-performing set of MFCC parameters and studying a speaker-independent approach

Seyed Reza Shahamiri *, Siti Salwah Binti Salim

Department of Software Engineering, Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia

Abstract

Dysarthria is a neurological impairment of controlling the motor speech articulators that compromises the speech signal. Automatic Speech Recognition (ASR) can be very helpful for speakers with dysarthria because the disabled persons are often physically incapacitated. Mel-Frequency Cepstral Coefficients (MFCCs) have been proven to be an appropriate representation of dysarthric speech, but the question of which MFCC-based feature set represents dysarthric acoustic features most effectively has not been answered. Moreover, most of the current dysarthric speech recognisers are either speaker-dependent (SD) or speaker-adaptive (SA), and they perform poorly in terms of generalisability as a speaker-independent (SI) model. First, by comparing the results of 28 dysarthric SD speech recognisers, this study identifies the best-performing set of MFCC parameters, which can represent dysarthric acoustic features to be used in Artificial Neural Network (ANN)-based ASR. Next, this paper studies the application of ANNs as a fixed-length isolated-word SI ASR for individuals who suffer from dysarthria. The results show that the speech recognisers trained by the conventional 12 coefficients MFCC features without the use of delta and acceleration features provided the best accuracy, and the proposed SI ASR recognised the speech of the unforeseen dysarthric evaluation subjects with word recognition rate of 68.38%.

1. Introduction

Dysarthria is a neurological impairment that damages the control of the motor speech articulators, which the malfunction is caused by the lack of control over the speech-related muscles, the lack of coordination among them, or their paralysis. It is often associated with irregular phonation and amplitude [1,2]. As a result of the impairment, the speech signal is compromised and its intelligibility is reduced [3,4]. According to [5], low intelligibility is one of the most detrimental social characteristics of dysarthria that affects different aspects of the lives of people with such disability.

Automatic Speech Recognition (ASR) systems identify the uttered word(s) represented as an acoustic signal and rely on a given lexicon to recognise the spoken word(s). They have several applications in health care, the military, telephony, and other domains [6]. They can be very helpful for speakers with dysarthria, because the disabled persons are often physically incapacitated and unable to use keyboards [7,8].

Most state-of-the-art commercial ASR systems are designed for speakers without speech disabilities, (i.e. non-speech disordered) and exclude those with speech disabilities [9]. These ASR systems record lower performance for individuals who suffer from dysarthria (specifically severe dysarthria [10,11]) than for people without speech disabilities as dysarthric speech is different from normal speech [12–14]. Therefore, there has recently been a trend towards creating specialised ASR systems for individuals with dysarthria instead of using ASR systems designed primarily for speakers without speech disabilities [3,10,15,16]. Thus, it is necessary to propose an ASR model specifically built for users with dysarthria that delivers adequate accuracy; specialised systems have generally achieved comparatively better performance for people with speech disorders [2,4,10,16].

According to [9], it is easier for people with dysarthria to utter isolated words rather than a continuous sequence of words. Similarly, it is more effective when the size of the ASR vocabulary is small and includes only simple words with one or two syllables in order to boost recognition rates with reduction or minimisation of dysarthric ASR errors [9]. Therefore, isolated-word and small-vocabulary ASR models are in greater demand for dysarthric speech recognition [10,15].