Biomedical Signal Processing and Control 46 (2018) 293–301

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Thomson Multitaper MFCC and PLP voice features for early detection of Parkinson disease

Savitha S. Upadhya [a,*], A.N. Cheeran [b], J.H. Nirmal [c]

[a] Electrical Engineering Department, Veermata Jijabai Technological Institute, Matunga, Mumbai, Maharashtra 400019, India
[b] Electrical Engineering Department, Veermata Jijabai Technological Institute, Matunga, Mumbai, Maharashtra, 400019, India
[c] Electronics Engineering Department, K J Somaiya College of Engineering, Vidyavihar, Mumbai, India

ARTICLE INFO

Article history:
Received 11 October 2017
Received in revised form 7 June 2018
Accepted 31 July 2018

Keywords:
MFCC
PLP
Single taper smooth window
Multitaper
Multitaper weights
Classification accuracy
Equal error rate

ABSTRACT

In this paper, MFCC and PLP voice features extracted using Single Taper Smooth (STS) window and Thomson Multitaper (TMT) windowing technique together with a neural network classifier is used in the classification of Healthy people from early stage Parkinson diseased patients and a performance comparison of the two techniques is reported. Parkinson disease in their early stages, not only affects the muscular movements of the human body but also influences the articulatory process of the speech production mechanism. This signifies change in the shape of the vocal tract which manifests itself in the short time power spectrum. The MFCC and PLP features used in this investigation, which represent the vocal tract parameters are derived from the short time spectrum. It is therefore crucial to estimate this short time power spectrum accurately. Generally, the short time speech power spectrum is estimated using STS window. But this power spectrum computed manifests large variance in the spectral estimates. Hence a variance reduced power spectrum is attained by computing the weighted average of the short time speech spectra obtained using a set of TMT windows. This spectrum is then used to compute the PLP and MFCC features. In this paper, extraction of both these voice features using STS window as well as TMT technique with three different weights namely Uniform, Eigen value (EV) and Adaptive weights is implemented using the speech samples of healthy and Parkinson diseased individuals. The experiment was carried out for several Thomson tapers ranging from 1 to 12 and the optimal number of tapers needed for the application and dataset is reported. A comparative performance analysis of the techniques implemented using both MFCC and PLP as features is then carried out in terms of classification accuracy, Equal Error Rate, sensitivity, selectivity and F1 score for the optimal taper value. The results obtained show that in comparison with the STS window a maximum improvement in the classification accuracy was obtained to be 6.6% for nine tapers, adaptive weights using MFCC as features and 6.9% for five tapers, EV weights using PLP as features for experimental dataset 1 and 6.0% using MFCC and 6.4% using PLP for experimental dataset 2. A performance improvement in other measures for the optimal taper value is also observed and reported for experimental dataset 1.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

The reduction in the production of the chemical dopamine by the brain of the human body results in Parkinson disease (PD), which affects the neuro mediator systems responsible for the muscular movements of the body [1]. PD is normally detected by performing neuro tests and scanning of the brain. These tests are costly and causes inconvenience to the PD patients. Hence researchers have investigated the detection of PD using signals like EEG [2], rate of finger tapping, pace of walking [3] and speech [4]. As acquisition of speech signal is easy, simple and also noninvasive, speech signals are used in our investigation.

The production of speech involves three processes. They are the respiration, phonation and articulation process. All these subsystems of the speech production mechanism are affected by Parkinson disease [5]. As a result, the quality and intelligibility of speech is reduced, which tends to worsen with the progression of the disease. The phonation subsystem which involves the vocal

* Corresponding author.
E-mail addresses: savitha.upadhya@fcrit.ac.in, ssupadhya_p15@ee.vjti.ac.in (S.S. Upadhya), ancheeran@ee.vjti.ac.in (A.N. Cheeran), jhnirmal@somaiya.edu (J.H. Nirmal).

https://doi.org/10.1016/j.bspc.2018.07.019
1746-8094/© 2018 Elsevier Ltd. All rights reserved.

folds movements and the articulatory subsystem which involves the movement of the different articulators are also affected by the disease, even in their early stages. So, extraction and analysis of phonation and articulation features of speech may provide the necessary cue in detection of PD.

PD detection using various speech features have been studied by researchers in the past [4,6–14]. The speech features prominently include the phonation parameters namely the short time jitter and shimmer parameters, noise parameters [4,6–10]. Using these features to discriminate PD and healthy subjects, an accuracy of more than ninety percent was reported by Mohammad Shahbakhi et al. [4] and by Athanasios Tsanas et al [6]. Research was also carried on articulatory speech features like MFCC, PLP, Linear Predictive Cepstral Coefficients (LPCC) [12,13]. Orozco-Arroyave J. R. et al. [12] extracted the MFCC features using a dataset which included sixty recordings of vowel /a/ for PD detection. Using a dataset with 17 healthy and PD patients respectively and MFCC coefficients, Achraf Benba et al. [13] reported an accuracy of greater than ninety percent. The same authors obtained an accuracy of ninety percent with PLP coefficients and SVM classifier [14]. The database used by most of the researchers had voice samples of people suffering from PD for 0 to 28 years.

The computation of MFCC and PLP features involves filter banks which are designed according to the perceptual criteria of the human ear and also involves the estimation of short time spectrum. It is important that this spectrum obtained by computing the DFT of the windowed speech frames should be estimated accurately. The window traditionally used is the STS window like the Hamming window. Smooth windows are preferred over rectangular windows because they offer reduced spectral leakage caused by the sidelobes, but the problem of increased variance in the spectral estimation persists. A technique for reducing the spectral variance is to replace a smooth windowed periodogram estimate with a multiple windowed spectrum estimate. This reduced variance spectral estimate is attained by computing the weighted addition of individual speech spectra obtained using M orthogonal windows [15–20]. In literature, the various windows used to reduce the spectral variance estimate includes the TMT [21], Sinusoidal weighted cepstrum estimator(SWCE) taper [22] and Multi-peak multi-taper [23]. Once the spectrum is obtained, the other procedures of computing the MFCC and PLP features is the traditional way as introduced by Davis and Mermelstein et al. [24] for MFCC and Hermansky H [25] for PLP features.

Researchers have used the Multitaper technique in experiments on verification of speakers [16,17], recognition of emotion [26] and other applications of speech too [27]. The experiment on recognition of emotion carried out by Attabi Y et al. [26] based on the Multitaper PLP and MFCC features reported an improvement for both MFCC and PLP systems when compared with STS window. Equal error rate reduction was reported by Alam M J et al. [16] for PLP and MFCC features respectively in the speaker verification experiment conducted.

In this paper, the speech corpus used has Healthy people and people suffering from PD for 3 months to 4 years (early stage). The objective of this investigation is to compare the performance of MFCC and PLP features extracted using STS and TMT windowing techniques for the application of classifying healthy and PD people using the above speech corpus. The performance measures used in the comparative analysis are classification accuracy, sensitivity, specificity, Equal Error Rate (EER), False Positive Rate (FPR), False alarm rate (FAR), False Negative Rate (FNR) and F1 score. A performance comparison of multitapered spectrum weighted with Adaptive, EV and Uniform weights is obtained in terms of classification accuracy to study the effect of the different weighing schemes.

This paper is organized as follows. Section 2 gives the description of TMT technique and the associated weights. Section 3 is the implementation and results section which discusses the implementation details and the results obtained. Finally, the conclusion is presented in Section 4.

2. Thomson Multitaper (TMT) technique

The technique of Multitaper described by Thomson D J [21], improves the estimation of the spectrum by addressing the effect of spectral leakage and its variance. In this technique, M tapers are used and each of these tapers are slightly different and hence energy leakage is reduced across frequencies [21]. Also, the TMT are orthogonal to each other and hence M orthogonal samples of data is obtained. These M data samples provide M orthogonal spectral estimates called the sub-spectrum. These M orthogonal spectral estimates are then weighted using Uniform (simple average), Non-uniform or Adaptive weights and summed up to obtain the weighted average spectrum called Multitapered spectrum. The sub-spectrums obtained using the orthogonal tapers are statistically independent because of their orthogonality property [28]. The speech spectrum computation using Multitaper technique is shown in Fig. 1. In this Figure, the dotted portion signifies power spectrum computation using the conventional STS window.

The TMT windows [19] are also called as the Discrete Prolate Spheroidal Sequences or DPSS [30]. These tapers are derived from the time-frequency concentration problem. Let x[n] be a finite energy sequence of length L defined over an interval [L1, L1 + L2]. Let the sampling rate be Fs and B be the bandwidth given by $\left|B\right| < \frac{F_s}{2}$.

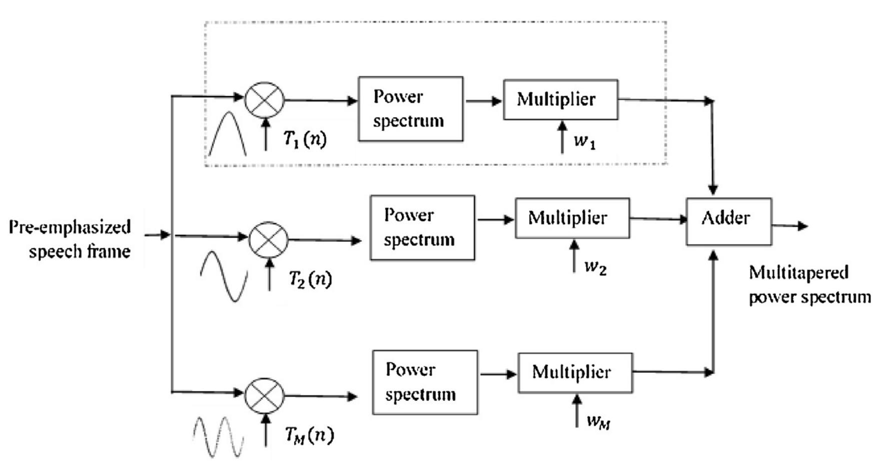

Fig. 1. Block diagram of Multitaper spectrum computation [29].

Download English Version:

https://daneshyari.com/en/article/6950682

Download Persian Version:

https://daneshyari.com/article/6950682

Daneshyari.com