EURASIP Journal on Advances in Signal Processing

RESEARCH Open Access

MFCC in Audio Signal Processing For Voice Disorder: A Review

Manjit Singh Sidhu*[1], Nur Atiqah Abdul Latib[1], Kaur Kirandeep Kulwant[2,*] and Shaliza Jumahat[1]

*Correspondence:
manjit@uniten.edu.my

[1] Universiti Tenaga Nasional, College of Computing and Informatics 43000, Putrajaya, Selangor Malaysia
[2] Universiti Putra Malaysia, Faculty of Health Sciences and Medicine, 43000, Serdang, Selangor, Malaysia

Abstract

Voice Disorder or Dysphonia has caught the attention of audio signal process engineers and researchers. The efficiency of several feature extraction and classifier implementation techniques in identifying voice abnormalities has been investigated. Mel-Frequency Cepstral Coefficient (MFCC) has been extensively used as a feature extractor. This paper adopts a Comparative Review Method to assess the effectiveness of feature extraction and classifier methods in detecting voice disorders. By examining the pairing of the Mel-Frequency Cepstral Coefficient (MFCC) with various classifiers, including Support Vector Machine (SVM), Artificial Neural Network (ANN), Decision Tree (DT), and other online or commercial classifiers, the study aims to review the robustness of MFCC in this context. Additionally, the paper acknowledges the importance of selecting the appropriate database, considering the diverse aetiology of pathological diseases, and its potential impact on the efficiency of voice disorder detection.

Keywords: Voice disorder; feature extractor; MFCC; classifier; SVM; ANN; Decision Tree

1 Introduction

Voice is an important medium for a human to be able to communicate as it contains information that individuals desire to share. Wrong pronunciation and inconsistent tone, pitch and loudness would lead to misinterpretation during a conversation. A voice disorder or dysphonia is defined as a difference in the person's voice quality, pitch, volume or flexibility in comparison to those of similar age, sex and cultural groups [1, 2]. The existence or absence of structural or neurological illnesses of the larynx determines whether two primary categories of voice disorders, known as organic and non-organic (functional), exist. Non-organic voice disorder can be simply explained as being brought on by an outside influence, such as excessive speech or voice abuse. [3]. Among the adult population functional voice disorder is the most frequently diagnosed type of voice disorder [4]. Voice disorders can create a financial burden in terms of time off work, treatment costs and reduced productivity [5].

Nur Atiqah *et al.*
EURASIP Journal on Audio, Speech and Music Processing (2023) 2023:15
https://doi.org/10.1187/s12477-023-01023-z

EURASIP Journal on Audio, Speech and Music Processing

RESEARCH Open Access

Classification of Non-Organic Voice Disorder Using Mel-Frequency Cepstral Coefficient (MFCC) with Support Vector Machine (SVM)

Nur Atiqah Abdul Latib[1], Manjit Singh Sidhu*[1], Kaur Kirandeep Kulwant[2], * and Shaliza Jumahat[1]

*Correspondence:
manjit@uniten.edu.my

[1] Universiti Tenaga Nasional, College of Computing and Informatics 43000, Putrajaya, Selangor Malaysia
[2] Universiti Putra Malaysia, Faculty of Health Sciences and Medicine, 43000, Serdang, Selangor, Malaysia

Abstract

Voice disorder detection has garnered interest among machine learning researchers. While previous studies focused on differentiating between healthy and pathological voices, few explored multi-class detection. Support Vector Machine (SVM) with additional features has been extensively studied. This project proposes a hierarchical SVM for non-organic voice disorders. Challenges and overfitting risks were addressed through k-fold cross-validation and spectral parameters. This method achieved impressive results, with 98.98% accuracy for healthy vs. hypo functional dysphonia and a minimum of 70.73% accuracy for differentiating between functional dysphonia, psychogenic dysphonia, dysodie, and dysphonie.

Keywords: Voice Disorder Detector, Support Vector Machine, Hierarchical, Non-Organic

1 Introduction

In the Oxford Languages Dictionary [1], voice is defined as the sound produced in a person's larynx and uttered through the mouth as speech or song. Scientifically, voice is a production of sound from sub glottal pressure that controls various parameters such as airflow, glottal area, fundamental frequency, and sound pressure level [2].

Voice is essential for human communication to share our thoughts, feelings, and emotions through spoken words. However, any deviation in the normal voice quality, pitch, and loudness could lead to a voice disorder [3]. There are many types of voice disorders but they can be categorized into two categories which are Organic and Non-Organic. Organic voice disorders usually result from changes in respiratory, laryngeal, or vocal tract mechanisms/ structures while non-organic voice disorders resulted from ineffective use of the vocal mechanism when the physical structure is normal [4] due to external factors such as injury, misuse of the voice, and prolonged speech [5].