Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 121 (2017) 19–26

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies, CENTERIS / ProjMAN / HCist 2017, 8-10 November 2017, Barcelona, Spain

Vocal Acoustic Analysis – Classification of Dysphonic Voices with Artificial Neural Networks

João Paulo Teixeira[a,b,*], Paula Odete Fernandes[a,b,c], Nuno Alves[a]

[a]*Polytechnic Institute of Bragança, Campus de Sta. Apolónia, 5300-253 Bragança, Portugal*
[b]*Applied Management Research Unit (UNIAG)*
[c]*Núcleo de Estudos em Ciências Empresariais (NECE)*

Abstract

Voice acoustic analysis is becoming nowadays a useful tool for detection of laryngological pathologies. This techniques enables a non-invasive and low cost assessment of voice disorders allowing a more efficient fast and objective diagnosis, permitting the patients to get a suitable treatment. In this work, the best predictors/parameters for diagnose of dysphonia were experimented. A vector made up of 4 Jitter parameters, 4 Shimmer parameters and Harmonic to Noise Ratio (HNR), determined from 3 different vowels at 3 different tones, in a total of 81 features, was used. Variable selection and dimension reduction techniques such as hierarchical clustering, multilinear regression analysis and principal component analysis (PCA) was applied. For the classification models based on artificial neural network (ANN) was used. The methods/models found allowed us to obtain an Accuracy of 100% for female voices and 90% for male voices using only Jitter Shimmer and HNR parameters.

Peer-review under responsibility of the scientific committee of the CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies.

* Corresponding author. Tel.: +351 273 303129; fax: +351 273 303051
E-mail address: joaopt@ipb.pt

1877-0509 © 2017 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the scientific committee of the CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies.
10.1016/j.procs.2017.11.004

Keywords: Voice Analysis; Dysphony; ANN; PCA; Multilinear Regression Analysis; Hierarquical Clustering.

1. Introduction

Voice analysis techniques are often used for voice disorders assessment[1,2,3]. The usefulness of such techniques rely upon the non-invasive character when compared with, for example, laryngoscopy exams[1]. Although, voice disorders may be diagnosed by an auditory perceptual analysis, this may lead to different results depending on the practitioner experience[4].

It is common in daily life of primary care facilities the people complains about hoarseness in their voices. The dysphonia affects 30% of adults and 50% of older adults. This disease modifies voice quality and has significant impacts on life quality. This also represents a significant economic burden. In patients with a progressive pathology it is important to do a diagnosis as fast as possible for the sake of having access to better treatment and prognosis[5].

There are several parameters extracted from speech signal processing. Teixeira and Fernandes[4] analyzed the reliability of Jitter, Shimmer and HNR parameters for dysphonia detection. A statistical analyzes was performed over the three parameters for the vowels /a/, /i/ and /u/ at three different tones, high, low and normal. In this work Jitter and Shimmer are suggested as good parameters to be used in intelligent diagnosis system of dysphonia pathologies.

To test this analysis we need to apply an intelligent tool and some reduction dimension and variable selection techniques. Variable selection is intended to select the best subset of predictors. The variable selection problem arises from large datasets who may contain redundant information and variables that have little or no predictive power[6]. The correct choice of input variables leads to a small subset that may boost/improve the training process when intelligent tools are used.

Henríquez *et al.*[7] studied the usefulness of six nonlinear chaotic measures based on nonlinear dynamics theory in the discrimination between two levels of voice quality: healthy and pathological. The studied measures are first and second order Rényi entropies, the correlation entropy and the correlation dimension. The values of the first minimum of mutual information function and Shannon entropy were also studied. Two databases were used to assess the usefulness of the measures: a multi-quality and a commercial database (MEEI Voice Disorders). A classifier based on standard neural networks was implemented in order to evaluate the measures proposed. Global success rates of 82.5% (multi-quality database) and 99.7% (commercial database) were obtained.

In Forero *et al.*[8], several parameters of glottal signal were used to identify nodule, unilateral paralysis or healthy voices. The database, obtained from a speech therapist, was composed of records of voices from 12 speakers with nodule, 8 speakers with vocal fold paralysis and 11 speakers with normal voices. Eight voice records were taken of each speaker making a total of 248 records. Three different classifiers were used, an Artificial Neural Network, a Support Vector Machine (SVM) and Hidden Markov Model. The best accuracy, 97.2%, was reached using glottal signal parameters and MFCC's with a SVM as classifier.

Markaki *et al.*[9] explored the information provided by a joint acoustic and modulation frequency representation, referred to as modulation spectrum, for detection and discrimination of voice disorders. The initial representation is first transformed to a lower dimensional domain using higher order singular value decomposition (HOSVD). For voice pathology detection an accuracy of 94.1% was achieved using SVM as classifier.

In Panek *et al.*[10], a vector made up of 28 acoustic parameters is evaluated using Principal Component Analysis (PCA), kernel principal component analysis (kPCA) and an auto-associative neural network (NLPCA) in four kinds of pathology detection (hyperfunctional dysphonia, functional dysphonia, laryngitis, vocal cord paralysis) using the /a/, /i/ and /u/ vowels, spoken at a high, low and normal pitch. The results shows a best efficiency levels of around 100%.

Al-Nasheri *et al.*[11] investigated different frequency bands using correlation functions. The authors extracted maximum peak values and their corresponding lag values from each frame of a voiced signal by using correlation functions as features to detect and classify pathological samples. Three different databases were used, Arabic Voice Pathology Database (AVPD), Saarbruecken Voice Database (SVD) and Massachusetts Eye and Ear Infirmary (MEEI). A Support Vector Machine was used as classifier. For detection of pathology an accuracy of 99,8%, 90.9% and 91.1% was achieved for the three databases respectively. In classification of the pathology task an accuracy of 99.2%, 98,9% and 95.1%, respectively, was achieved for the three databases.