Neurocomputing 167 (2015) 94–111

Contents lists available at ScienceDirect

Neurocomputing

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

Robust and complex approach of pathological speech signal analysis

Jiri Mekyska [a,b], Eva Janousova [c], Pedro Gomez-Vilda [d], Zdenek Smekal [a,b], Irena Rektorova [e,f,*], Ilona Eliasova [e,f], Milena Kostalova [g,f], Martina Mrackova [e,f], Jesus B. Alonso-Hernandez [h], Marcos Faundez-Zanuy [i], Karmele López-de-Ipiña [j]

[a] Department of Telecommunications, Brno University of Technology, Technicka 10, 61600 Brno, Czech Republic
[b] SIX Research Centre, Technicka 10, 61600 Brno, Czech Republic
[c] Institute of Biostatistics and Analyses, Masaryk University, Kamenice 3, 62500 Brno, Czech Republic
[d] Facultad de Informatica, Universidad Politecnica de Madrid, Campus de Montegancedo, s/n, 28660 Boadilla del Monte, Madrid, Spain
[e] First Department of Neurology, St. Anne's University Hospital, Pekarska 53, 65691 Brno, Czech Republic
[f] Applied Neuroscience Research Group, Central European Institute of Technology, Masaryk University, Komenskeho nam. 2, 60200 Brno, Czech Republic
[g] Department of Neurology, Faculty Hospital and Masaryk University, Jihlavska 20, 63900 Brno, Czech Republic
[h] Institute for Technological Development and Innovation in Communications (IDeTIC), University of Las Palmas de Gran Canaria, 35001 Las Palmas de Gran Canaria, Spain
[i] Escola Universitaria Politecnica de Mataro, Tecnocampus, Avda. Ernest Lluch 32, 08302 Mataro, Barcelona, Spain
[j] Department of Systems Engineering and Automation, University of the Basque Country UPV/EHU, Av de Tolosa 54, 20018 Donostia, Spain

ARTICLE INFO

Article history:
Received 1 October 2014
Received in revised form
12 January 2015
Accepted 9 February 2015
Communicated by Carlos Manuel Travieso González
Available online 21 May 2015

Keywords:
Pathological speech
Disordered voice
Dysarthria
Speech processing
Bicepstrum
Non-linear dynamic features

ABSTRACT

This paper presents a study of the approaches in the state-of-the-art in the field of pathological speech signal analysis with a special focus on parametrization techniques. It provides a description of 92 speech features where some of them are already widely used in this field of science and some of them have not been tried yet (they come from different areas of speech signal processing like speech recognition or coding). As an original contribution, this work introduces 36 completely new pathological voice measures based on modulation spectra, inferior colliculus coefficients, bicepstrum, sample and approximate entropy and empirical mode decomposition. The significance of these features was tested on 3 (English, Spanish and Czech) pathological voice databases with respect to classification accuracy, sensitivity and specificity. To our best knowledge the introduced approach based on complex feature extraction and robust testing outperformed all works that have been published already in this field. The results (accuracy, sensitivity and specificity equal to $100.0 \pm 0.0\%$) are discussable in the case of Massachusetts Eye and Ear Infirmary (MEEI) database because of its limitation related to a length of sustained vowels, however in the case of Príncipe de Asturias (PdA) Hospital in Alcalá de Henares of Madrid database we made improvements in classification accuracy ($82.1 \pm 3.3\%$) and specificity ($83.8 \pm 5.1\%$) when considering a single-classifier approach. Hopefully, large improvements may be achieved in the case of Czech Parkinsonian Speech Database (PARCZ), which are discussed in this work as well. All the features introduced in this work were identified by Mann–Whitney U test as significant ($p < 0.05$) when processing at least one of the mentioned databases. The largest discriminative power from these proposed features has a cepstral peak prominence extracted from the first intrinsic mode function ($p = 6.9443 \times 10^{-32}$) which means, that among all newly designed features those that quantify especially hoarseness or breathiness are good candidates for pathological speech identification. The paper also mentions some ideas for the future work in the field of pathological speech signal analysis that can be valuable especially under the clinical point of view.

1. Introduction

Voice Pathology description and characterization has always demanded attention from the physiological and medical fields [1], as well as from the voice function point of view [2]. The term voice is described by [3] both in a broad and a narrow sense. In the broad sense voice may be taken as synonymous of speech, therefore terms as Voice over IP (VoIP) can be found in the literature and media with the meaning of speech data on internet. In the narrow sense voice refers to the vibration of the vocal folds. Speech sounds resulting from the interaction of this vibration with the Oro-Naso-Pharyngeal Tract (ONFT) are referred to, as voiced. Speech sounds produced by turbulent flow within the ONFT are

* Corresponding author at: First Department of Neurology, St. Anne's University Hospital, Pekarska 53, 65691 Brno, Czech Republic. Tel.: +420 549 497 825.
E-mail address: irena.rektorova@fnusa.cz (I. Rektorova).

http://dx.doi.org/10.1016/j.neucom.2015.02.085
0925-2312/

termed voiceless. Phonation is the recommended term to refer to vocal fold vibration. A speaker showing an anomalous vocal fold vibration pattern is referred as dysphonic, and aphonic if there is no vocal fold vibration at all. If no anomalies are present the speaker is referred as normophonic.

Dysphonic voice is a perceptual and subjective term associated to Voice Pathology. Dysphonia is the perceptual quality of voice signaling that something wrong is happening in the phonation organs (mainly the larynx and its associated structures). Voice Pathologies and Dysphonic Voice are thus intrinsically related. The classification of Voice Pathologies or Disorders is described in [3] as tissue infection (e.g. laryngitis, bronchitis, croup, …), systemic changes (e.g. dehydration, pharmacological and drug effects, hormonal changes, …), mechanical stress (e.g. vocal nodules, polyps, ulcers, granulomae, laryngocele, hemorrhage, …), surface irritation (e.g. laryngitis, leukoplakia, gastroesophageal reflux, …), tissue changes (e.g. laryngeal carcinoma, keratosis, papillomas, cysts, …), neurological and muscular changes (e.g. bilateral and unilateral vocal fold paralysis, Parkinson's Disease (PD), Amyotrophic Lateral Sclerosis (ALS), myotonic dystrophy, Huntington's Chorea, myasthenia gravis, …), and abnormal muscle patterns (e.g. conversion aphonia or dysphonia, spasmodic dysphonia, mutational dysphonia, ventricular phonation, …). In [4] a description of vocal pathologies can be found. A last important group of neurological diseases which leave a correlate in voice and speech is that of cognitive origin, Alzheimer's Disease (AD) being the most relevant one for their impact in well-being and in aging specialized-attention demand. Some references on the influence of AD in speech and voice can be found in [5–8]. Going a step further, emotional alterations (either temporary or persistent) leave also correlates in the speech and phonation signature, and may be subjects of further study by acoustic analysis [9,10].

The relationship between acoustic correlates and voice pathology has been clinically established in the last decades, subjectively and quantitatively [11,12]. Acoustic Voice Quality Analysis (AVQA) is a wide term for a set of different methodologies designed to quantify acoustic correlates giving a definition of the quality of phonation or speech production. Therefore AVQA would be the procedural way to objectively quantify anomalies manifested in Dysphonic Voice. Modern signal processing technologies provide estimates of voice and speech correlates in time and frequency [13], allowing to better visualize and quantify phonation patterns. Spectral techniques facilitate the study of pathologic phonation, establishing relations between harmonic–harmonic and harmonic–formant ratios, which were found as important correlates to organic voice pathology [14]. Similarly, harmonic–noise ratios were found significant in characterizing certain types of dysphonic pathologies [15,16]. Time domain estimates, as jitter, shimmer and open-closed phase quotients are also used in describing dysphonic voice [17,18]. These descriptions gave rise to AVQA as a specific field [19].

Under the point of view of AVQA the following objectives can be established in order of difficulty: dysphonic voice detection, dysphonic voice grading, and dysphonic voice classification according to etiology. Dysphonic voice detection would be the task of assigning normophonic (normative) or dysphonic (non-normative) labels to a given phonation produced by a specific speaker. The determination of the dysphonic grade is traditionally carried out by independent referees according to a subjective criterion on a given scale. One of the most popular is GRBAS (grade, roughness, breathiness, asthenia and strain) [11]. The relative dependence of the assigned grade to the referee's subjective opinion results in wide grading differences among referees. To overcome this problem requires the design of clinical assessment methodologies [20]. The task of dysphonic voice classification according to etiology is far more difficult, as a given acoustic correlate may be attributed to different pathologies. If this problem is stated in terms of associating acoustic correlates to specific pathologies, the potential risk is that it will remain unsolved for long, because it is an ill-posed problem (a many-to-one subjective mapping). A preliminary step to be covered first is to define the implications of different pathologies in the vocal function, especially at the level of the larynx. Under the functional point of view, the following main behaviors may be observed in the abnormal operation of the vocal folds: asymmetric vibration, contact defects and dystonia (hypo-, hyper-tension and tremor). Most organic larynx pathologies reproduce either one or another behavior, or all of them. Asymmetric vocal fold vibration is to be expected in pathologies as polyps, cysts, carcinomae, vocal fold paralysis, ulcers, cysts, and papillomae, and produces acoustic correlates as jitter, shimmer, poor harmonic–noise ratios, unbalance, and sub- and inter-harmonics. On its turn, contact defects are to be expected in pathologies as polyps, nodules, edemae, and cysts, where full closure of the glottal gap is not granted by vocal fold adduction and produce acoustic correlates as open and close phase perturbations, recovery phase attenuation, and harmonic display reduction. Other pathologies, especially those of neurological origin produce deviations in biomechanical parameters, as vocal fold tension (hypo- and hyper-tonic) and tremor, which can also be manifested as modulations in amplitude or frequency, and as changes in the vocal fold tension. As many organic pathologies induce a hyper-tonic behavior, the main problem when dealing with correlates produced by neurological pathology is to differentiate their origin from that of organic origin. Contact defect pathologies can also show asymmetric vocal fold vibration, and this may also be the case in aging voice (presbyphonia). To establish differentiation criterion, for using acoustic correlates when dealing with organic, neurologic or aging-induced perturbations, is a major issue in AVQA [21].

Another problem regarding AVQA is the lack of good reference baselines to establish the methodologies for voice pathology classification from acoustic analysis. Rigorous databases, acquired to represent each of the different pathologies under well-defined sample population size and recording conditions are scant. Many times the problem comes from the simultaneous presence of different larynx pathologies in the same patient, either related or unrelated (e.g. it is common to find a counter-lateral lesion as a consequence of a unilateral polyp). The problem then is how to decide if a specific acoustic correlate is produced by one cause or another. Most of the times acoustic databases are produced by laryngological services as part of examination protocols [22,23], but this is not always the case, as many laryngologists prefer to depend on visual exploration, neglecting the possibilities offered by AVQA for different reasons [24]. Speech therapists rely more on acoustic exploration, but many times it is mainly restricted to measurements of vocal effort, long term frequency analysis, respiratory efficiency, or distortion measures. Therefore voice records from speakers ranked by etiology and severity index, using compatible standards (digitalization, channel, microphones) are scant. Most of the available databases are either incomplete, inconsistent (made up of recordings taken under different conditions) or deficient (many non-frequent pathologies are not well represented in sample size). Besides, there is a lack of good normative databases, as most of the records produced by medical services contain information from dysphonic voice, but normative speakers are absent or poorly represented (this is the case of the most widely used database [25]). This is a severe limitation for systematic AVQA. Another problem is cross-lingual representation. As far as sustained vowels are concerned, this would not be a problem, but it becomes a major obstacle when segmental parameters are involved, as in the use of passages (either read or spontaneous).