Available online at www.sciencedirect.com
ScienceDirect
Computer Speech & Language xxx (2017) xxx-xxx
www.elsevier.com/locate/csl

Characterisation of voice quality of Parkinson's disease using differential phonological posterior features☆

Q1 Milos Cernak*[a], Juan Rafael Orozco-Arroyave[b,e], Frank Rudzicz[c], Heidi Christensen[d], Juan Camilo Vásquez[b,e], Elmar Nöth[e]

Q2 [a] Idiap Research Institute, Martigny, Switzerland
[b] Universidad de Antioquia Medellín, Colombia
[c] University of Toronto, Canada
[d] University of Sheffield, UK
[e] Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

Received 13 March 2017; received in revised form 12 May 2017; accepted 12 June 2017

Abstract

Change in voice quality (VQ) is one of the first precursors of Parkinson's disease (PD). Specifically, impacted phonation and articulation causes the patient to have a breathy, husky-semiwhisper and hoarse voice. A goal of this paper is to characterise a VQ spectrum – the composition of non-modal phonations – of voice in PD. The paper relates non-modal healthy phonations: breathy, creaky, tense, falsetto and harsh, with disordered phonation in PD. First, statistics are learned to differentiate the modal and non-modal phonations. Statistics are computed using phonological posteriors, the probabilities of phonological features inferred from the speech signal using a deep learning approach. Second, statistics of disordered speech are learned from PD speech data comprising 50 patients and 50 healthy controls. Third, Euclidean distance is used to calculate similarity of non-modal and disordered statistics, and the inverse of the distances is used to obtain the composition of non-modal phonation in PD. Thus, pathological voice quality is characterised using healthy non-modal voice quality "base/eigenspace". The obtained results are interpreted as the voice of an average patient with PD and can be characterised by the voice quality spectrum composed of 30% breathy voice, 23% creaky voice, 20% tense voice, 15% falsetto voice and 12% harsh voice. In addition, the proposed features were applied for prediction of the dysarthria level according to the Frenchay assessment score related to the larynx, and significant improvement is obtained for reading speech task. The proposed characterisation of VQ might also be applied to other kinds of pathological speech.

Keywords: Phonological features; Non-modal phonation; Parkinson's disease

1. Introduction

Speech of hypokinetic dysarthria in Parkinson's disease (PD) is characterised by hypokinesia (rigid, less motion describing decreased range and frequency of movement) of the vocal folds and articulators. Besides of impacted

☆ This paper has been recommended for acceptance by Roger Moore.

* Corresponding author.
E-mail address: milos.cernak@idiap.ch (M. Cernak).

http://dx.doi.org/10.1016/j.csl.2017.06.004
0885-2308/ 2017 Published by Elsevier Ltd.

prosody and articulation, phonation is impacted by incomplete vocal fold adduction. Clinicians, otolaryngologists and speech-language pathologists, consider hoarseness − a rough quality of voice − as a basic symptom of a voice disorder in PD. When hoarse, the voice may sound breathy, raspy, or strained, and if this abnormal/pathological voice quality is accompanied with relatively constant loudness and pitch deviations, it is diagnosed as Parkinsonian dysphonia (Aronson and Bless, 2011).

Healthy subjects may also produce speech sounds of different voice quality based on different modes of vibration of the vocal folds. Laver (1980) defines the term of voice quality in a broad sense as the characteristic auditory colouring of an individual speaker's voice, and not just in a narrow sense coming from the laryngeal activity. The neutral mode phonation, often used in *modal voice*, is one against which the other modes can be contrastively described, also called non-modal phonations. Ladefoged and Johnson (2014) describe four basic states of the glottis (which is defined as the space between the vocal folds). The position of the vocal folds is adjusted by the arytenoid cartilages placed toward the back. In (i) a voiced sound, the vocal folds are close together (adducted) and vibrating, whereas in (ii) a voiceless sound, they are pulled apart (abducted). If there is considerable airflow, the abducted vocal folds will be set vibrating − flapping in the airstream − producing what is called (iii) *breathy voice*, or murmur. Alternatively, breathy voice is produced with the vocal folds apart only between the arytenoid cartilages in the lower (posterior) part. If the arytenoid cartilages are tightly together, so that the vocal folds can vibrate only at the anterior end, (iv) *creaky voice* is produced. Creaky-voiced sounds may also be called laryngealised. Besides these basic non-modal phonation, Laver (1980) defines *tense, harsh* and *falsetto* phonations. Such voice qualities impact the production of the speech sounds, and we hypothesise that these changes might be captured by changes in phonological features.

The goal of this paper is to present a study on the production of speech sounds with healthy non-modal phonation, and project its non-modal statistics to analysise disordered production of speech sounds with pathological phonation. This approach might help to aleviate a problem of missing data in research of pathological speech. Voice quality of the speech sounds can be characterised by phonological features (Cernak et al., 2017b), and the current work proposes to use differential phonological posterior features (between modal and non-modal, and between healthy and disordered phonations) for characterisation of both healthy non-modal and parkinsionian phonations. Comparing to the work of Cernak et al. (2017b), the novel aspects of this paper is in using pathological speech and characterisation of pathological voice quality using healthy non-modal voice quality "base/eigenspace". An Euclidean distance between the non-modal and disordered phonation characterisations quantifies the composition of non-modal voice qualities in PD. This characterisation of non-modal phonation in PD is novel, and shows objective quantification of voice quality using phonological features not investigated in previous approaches.

For studying speech with non-modal phonation, the read-VQ database of Kane (2012) is used, the recording of which was inspired by prototype voice quality examples produced by Laver (1980). Laver's recordings are considered as recordings of non-modal phonation with excellent quality, however only one utterance per phonation type is available, and thus they are speaker-specific. The read-VQ database contains recordings from four speakers. The database covers five different non-modal phonations: falsetto, creaky, harshness, tense and breathiness. For studying speech with pathological phonation, the Colombian-Spanish database (Orozco-Arroyave et al., 2014) is used, which contains speech recordings of 50 patients with PD and 50 healthy controls (HC).

The structure of the paper is as follows: Section 2.1 gives an overview of the non-modal (healthy) and pathological (Parkinsonian) phonation types considered in this work. Section 3 introduces differential phonological posterior (DPP) features used in further characterisation of VQ. Section 4 describes experimental setup and evaluation databases, and Section 5 presents results and their validation. Finally, Section 6 concludes the paper.

2. Voice quality of Parkinson's disease

2.1. Non-modal (healthy) phonation

Different modes of vibration of the vocal folds contribute significantly to VQ. The modal (periodic) phonation can be contrastively described against the other modes, also called non-modal (aperiodic) phonations.

Recent work on non-modal phonation focuses on detection (Drugman et al., 2014), analysis (Malyska, 2008; Malyska et al., 2011) and synthesis (Bangayan et al., 1997) of speech with non-modal phonation. Modern computational paralinguistics tries to 1) get rid of non-modal phonation, or 2) model it, for example, for classification

Please cite this article as: M. Cernak et al., Characterisation of voice quality of Parkinson's disease using differential phonological posterior features, Computer Speech & Language (2017), http://dx.doi.org/10.1016/j.csl.2017.06.004