G Model
BSPC-875; No. of Pages 7
ARTICLE IN PRESS
Biomedical Signal Processing and Control xxx (2016) xxx–xxx
Contents lists available at ScienceDirect
Biomedical Signal Processing and Control
journal homepage: www.elsevier.com/locate/bspc

On the harmonic-to-noise ratio as an acoustic cue of vocal timbre of Parkinson speakers

Abdellah Kacha [a,b,*], Christophe Mertens [b], Francis Grenez [b], Sabine Skodda [c], Jean Schoentgen [b,d]

[a] *Laboratoire de Physique de Rayonnement et Applications, Université de Jijel, Jijel, Algeria*
[b] *BEAMS, Université Libre de Bruxelles, Brussels, Belgium*
[c] *Department of Neurology, Knappschaftskrankenhaus, Ruhr-University of Bochum, Germany*
[d] *National Fund for Scientific Research, Belgium*

ARTICLE INFO

Article history:
Received 26 May 2016
Received in revised form 31 August 2016
Accepted 5 September 2016
Available online xxx

Keywords:
Harmonic-to-noise ratio
Cycle length jitter
Empirical mode decomposition
Parkinson disease

ABSTRACT

The relevance of the harmonic-to-noise ratio (HNR) and glottal cycle length jitter as cues of the vocal timbre of Parkinson speakers is investigated. HNR and vocal cycle length jitter are known to be suitable cues for the evaluation of the vocal timbre of dysphonic speakers. However, the question whether they are relevant descriptors of the voice quality of Parkinson speakers is still unanswered. Empirical mode decomposition (EMD) has been used to estimate the HNR by decomposing the log-magnitude spectrum of the speech signal into its harmonic, contour and noise components. Cycle length jitter has been estimated via the break-up by empirical mode decomposition of the cycle length time series into the intonation contour as well as the perturbations owing to tremor and jitter. HNR and cycle length jitter values of vowels [a] sustained by 205 Parkinson and 74 control speakers are in the same interval respectively and the differences are not statistically significant. Also, the standard deviations of the per-frame HNR values of an utterance do not differ statistically significantly between control and Parkinson speakers.

1. Introduction

Parkinson disease (PD) is a chronic neurological disorder caused by a loss of nerve cells in the part of the brain called the substantia nigra [1]. The core motor symptoms of PD include slow physical movements (hypokinesia), tremor, muscle stiffness (rigidity) and postural instability. Secondary symptoms include – among others – cognitive impairments, a variety of autonomic and neuropsychiatric features, sleep problems and speech and voice impairment. In a study based on a large sample, it has been reported that between 70% and 90% of patients with PD have problems related to speech and voice [2] that are referred to as hypokinetic dysarthria.

Conventional methods for assessing voice disorder in PD are based on the auditory evaluation of voice and speech by a clinician who assigns a score in the framework of an evaluation protocol such as the unified Parkinson disease rating scale (UPDRS) [3]. A drawback of a perceptual evaluation is intra and inter-judge variability. Recently, there has been a considerable interest in the acoustic analysis and classification of speech with a view to demonstrating that the values of selected acoustic cues differ for control and PD speakers [4–7]. A reliable method for PD speech assessment would fulfill the cliniciansí need for objective and reproducible means to categorize the severity and progression of the symptoms and to monitor therapeutic interventions.

Several acoustic analysis methods for the detection and clinical evaluation of voice disorders of patients with PD have been proposed in the literature. They attempt to quantify the severity of the disorder by focusing on phonation, articulation or prosody. A summary of acoustic cues assigned to aspects of motor control of speech has been presented in [8] in the context of the analysis of dysarthric speech related to PD and other neurologic diseases.

The Harmonic-to-noise ratio (HNR) commonly used in the general framework of the analysis of disordered voices to assess the degree of hoarseness [9,10] has been suggested in a number of studies as a cue for the assessment of voice quality of Parkinson speakers [11–13]. The HNR defined as the ratio between the power of the harmonic structure of the spectrum and that of the nonharmonic structure reports the deviation of the speech waveform from perfect periodicity and indicates to what extent a voice is perturbed.

A temporal acoustic marker for voice quality assessment of hoarse speakers is cycle length jitter. It has been used as a descrip-

* Corresponding author at: Department of Electronics, Faculty of Science and Technology, Université de Jijel, Algeria.
E-mail address: akacha@ulb.ac.be (A. Kacha).

http://dx.doi.org/10.1016/j.bspc.2016.09.004
1746-8094/

Please cite this article in press as: A. Kacha, et al., On the harmonic-to-noise ratio as an acoustic cue of vocal timbre of Parkinson speakers, Biomed. Signal Process. Control (2016), http://dx.doi.org/10.1016/j.bspc.2016.09.004

tor of the voice of Parkinson speakers in several studies [6,7,12]. It is defined as the variation from cycle-to-cycle of the glottal cycle duration [14]. Cycle length jitter is a consequence of frequency modulation noise that affects considerably the spectral properties of voiced speech sounds.

When the speech signal is periodic, its magnitude spectrum would have harmonics at the multiple integers of the fundamental frequency. When the signal deviates from perfect periodicity due to frequency modulation noise, sidebands referred to as inter-harmonics appear between harmonics. Since the HNR is a measure of the harmonicity of the speech signal and quantifies its degree of perturbation by measuring the ratio of the harmonic component to noise, the vocal frequency jitter has a major influence on the HNR when the analysis frame comprises more than few vocal cycles. An increase/decrease of vocal frequency jitter results in a decrease/increase of the HNR. Thus, cycle length and HNR are related.

The extent to which HNR can be useful for discriminating between control and Parkinson speakers needs further investigation. Indeed, some studies have reported contradictory results concerning the discrimination capability of the HNR in the framework of the assessment of PD [8,15].

In [4], a large number of acoustic cues grouped into sets that share common attributes have been proposed for the discrimination between PD and control speech. Most of these cues are known from the analysis of disordered voices in general. Among the tested features, mel frequency cepstral coefficients and cues that quantify noise have been retained, including the HNR, glottal-to-noise excitation and vocal fold excitation ratio. The findings are in agreement with those reported in [7] which found that the difference in HNR between control and Pakkinson speakers is statistically significant.

More recently, the HNR has been used in [6] with other acoustic cues to characterize voice disorders and dysarthria in Parkinson speakers. The acoustic cues have been estimated using existing software. The results did not show statistically significant differences between the values of the noise-related cues for Parkinson and control speakers. These findings are consistent with those reported in previous studies [13,15] in which the signal-to-noise ratio together with other acoustic perturbation features failed to discriminate between control and Parkinson speakers.

Even when taking into account that the corpora as well as the methods used in [4] and [6] for estimating the HNR differ, their conclusions remain contradictory. In [6], it has been found that the difference in HNR values between control and Parkinson speakers is not statistically significant. On the contrary, in [4], it has been found that the HNR is among the most powerful ten acoustic cues of a total of 132 acoustic cues in terms of the ability of discriminating between control and Parkinson speakers. The question whether the HNR is a relevant feature for characterizing PD speech therefore remains unanswered.

In the present study, the relevance of the HNR for the description of PD speech is investigated based on a large corpus of vowels sustained by control and Parkinson speakers. Empirical mode decomposition (EMD) based method is used to estimate the HNR. The EMD-based method for HNR estimation does not rely on the isolation of individual speech cycles or pseudo-harmonics/rhamonics in the speech spectrum/cepstrum as most techniques for HNR estimation. The isolation of individual speech cycles or pseudo-harmonics/rhamonics cannot be carried out reliably in speech produced by severely hoarse speakers and results in inaccurate HNR estimation. Previous studies in the framework of disordered voice analysis has shown that the EMD-based method is effective for estimating the HNR and outperforms known methods [16].

Also, the vocal cycle length jitter has been evaluated for comparative purposes. The cycle length jitter has been estimated via the break-up by empirical mode decomposition of the cycle length time series into the intonation contour and the perturbations owing to jitter as well as tremor. The use of an EMD-based method for the calculation of the HNR and cycle length jitter offers the opportunity to obtain a temporal and a spectral cue by means of similar methods.

The remainder of the article is organized as follows. The empirical mode decomposition based methods for HNR and cycle length jitter estimation are reviewed in Section 2. The corpus is described in Section 3. Results and discussion are presented in Section 4. Finally, conclusions are given in Section 5.

2. Methods

Glottal cycle length jitter and the harmonic-to-noise ratio are obtained after pre-processing, respectively, the glottal cycle length time series or the log-magnitude spectrum of an analysis frame by means of empirical mode decomposition. The latter breaks up a sequence of data into narrow-band oscillating intrinsic mode functions and a residue that when summed enable reconstructing the original data. Since cycle length jitter or the sidebands of the partials inform about rapid modulations of the vocal frequency, they are expected to be reported by intrinsic mode functions that are characterized by high frequencies or quefrencies, respectively. Assigning intrinsic mode functions to different categories according to their average frequency or quefrency therefore enables separating jitter from tremor and intonation in the temporal domain and modulation noise from the harmonics and spectral contour in the spectral domain.

2.1. Estimation of the harmonic-to-noise ratio

In speech signal analysis, a voiced speech frame $x(t)$ is modeled as a periodic source component $e(t)$ convolved (*) with the impulse response of the vocal tract $v(t)$ [17],

$$x(t) = e(t) * v(t). \quad (1)$$

Windowing the signal frame $x(t)$ and taking the magnitude of the Fourier transform gives

$$|X_w(f)| = |E_w(f) \times V(f)|, \quad (2)$$

where $X_w(f)$, $E_w(f)$ are short-time spectra of the windowed speech and excitation signal frames respectively, and $V(f)$ is the frequency response of the vocal tract.

Taking the logarithms changes the multiplication into an addition:

$$\log|X_w(f)| = \log|E_w(f)| + \log|V(f)|. \quad (3)$$

From (3), it is seen that the log-magnitude spectrum is actually the sum of two spectral components: $\log|E_w(f)|$, the log-magnitude spectrum of the windowed excitation signal and $\log|V(f)|$, the spectral contour due to the filtering characteristic of the vocal tract. Aspiration and modulation noise at the glottis disturb the pseudo-periodic excitation signal so that the excitation spectrum itself may be considered as a regularly spaced sequence of harmonics the magnitude of which decreases with frequency, superimposed on irregularly distributed noise.

The estimation of the three components of the log-magnitude spectrum has been carried out by means of empirical mode decomposition. When applied to the log-magnitude spectrum of a speech frame, the EMD algorithm outputs several oscillating intrinsic mode functions (IMFs) and a residue, the sum of which is exactly equal to the original [18]. The IMFs are clustered into three categories that are assigned to the log-magnitude spectrum of the noise, the sequence of harmonics and the contour. The latter includes the

Please cite this article in press as: A. Kacha, et al., On the harmonic-to-noise ratio as an acoustic cue of vocal timbre of Parkinson speakers, Biomed. Signal Process. Control (2016), http://dx.doi.org/10.1016/j.bspc.2016.09.004