G Model
BSPC-973; No. of Pages 9
ARTICLE IN PRESS
Biomedical Signal Processing and Control xxx (2017) xxx–xxx
Contents lists available at ScienceDirect
Biomedical Signal Processing and Control
journal homepage: www.elsevier.com/locate/bspc

Features of vocal frequency contour and speech rhythm in bipolar disorder

A. Guidi [a,b,*], J. Schoentgen [c], G. Bertschy [d], C. Gentili [e], E.P. Scilingo [a,b], N. Vanello [a,b]

[a] Dipartimento di Ingegneria dell'Informazione, University of Pisa, Via G. Caruso 16, 56122, Pisa, Italy
[b] Research Center "E. Piaggio", University of Pisa, Largo L. Lazzarino 1, 56122, Pisa, Italy
[c] Department of Signals, Images and Acoustics, Faculty of Applied Sciences, Université Libre de Bruxelles, Av. F. Roosevelt 50, 1050, Bruxelles, Belgium
[d] Department of Psychiatry, Strasburg University Hospital; Translational Medicine Federation, University of Strasburg, INSERM u1114, 1 place de l'Hôpital, 67000 Strasburg, France
[e] Department of General Psychology, University of Padua, Via Venezia 8, 35131, Padua, Italy

A R T I C L E I N F O

Article history:
Received 1 June 2016
Received in revised form 10 October 2016
Accepted 28 January 2017
Available online xxx

Keywords:
Bipolar disorder
Mood state
Speech contour
Speech rhythm
Spectral analysis

A B S T R A C T

Mental diseases are increasingly common. Among these, bipolar disorders heavily affect patients' lives given the mood swings ranging from mania to depression. Voice has been shown to be an important cue to be investigated in relation with this kind of disease. In fact, several speech-related features have been used to characterize voice in depressed speakers. The goal is to develop a decision support system facilitating diagnosis and possibly predicting mood changes. Recently, efforts were devoted to studies concerning bipolar patients. A spectral analysis of F0-contours extracted from audio recordings of a text read by bipolar patients and healthy control subject is reported. The algorithm is automatic and the obtained features describe parsimoniously speech rhythm and intonation. Bipolar patients were recorded while experiencing different mood states, whereas the control subjects were recorded at different days. Feature trends are detected in bipolar patients across different mood states, while no significant differences are observed in healthy subjects.

1. Introduction

Mental illnesses have an increasing impact in contemporary society [1]. In particular, the lives of persons suffering from bipolar disorder may be impaired due to periodic, and sometimes extreme, mood swings. Patients may experience oscillations between depression, mania or hypomania, euthymia and a mixed condition [2]. More specifically, depression is a very low mood state characterized by sadness and hopelessness. Mania and hypomania, i.e. a less severe form of mania, are states of hyperarousal that leads to euphoria or irritability, excessive energy, and increased activity. Euthymia is a mood state in which symptoms are mostly absent, and mixed condition is a mood state in which both depressive and manic symptoms are associated. The development of decision support systems may be useful in helping physicians in formulating a diagnosis. With this aim some biomedical studies have been carried out to detect physiological correlates of mood changes [3,4]. Moreover, several studies have focused on possible relations between voice and mental disease, especially in persons suffering from depression. Indeed, almost the whole central nervous system is involved in voice production [5]. The speech motor system is able to finely control the laryngeal muscles, mucosae, tongue and lips, while the vagus nerve, that is also responsible for the innervation of the motor parasympathetic fibers, activates the pharynx, the soft palate and the laryngeal muscles [5]. In addition, because both the autonomic and somatic nervous systems control the respiratory system, they are expected to affect also the speakers' prosody. As a consequence, speech analysis represents an interesting, non-invasive and economic approach to the study of a speaker's mental state since speech production is modulated by different psychological and/or mental states [6].

Usually three different categories of speech related features can be taken into account in speech analysis: source, vocal tract and prosody-related features [7]. Source features aim at describing the glottal excitation. Vocal tract features report indirectly different shapes of the vocal tract cross-section during articulation. Finally, prosodic features describe the supra-segmental modulation of into-

* Corresponding author at: Dipartimento di Ingegneria dell'Informazione, University of Pisa, Via G. Caruso 16, 56122, Pisa, Italy.
E-mail addresses: andrea.guidi@for.unipi.it (A. Guidi), jschoent@ulb.ac.be (J. Schoentgen), gilles.bertschy@chru-strasbourg.fr (G. Bertschy), c.gentili@unipd.it (C. Gentili), e.scilingo@centropiaggio.unipi.it (E.P. Scilingo), nicola.vanello@iet.unipi.it (N. Vanello).

http://dx.doi.org/10.1016/j.bspc.2017.01.017
1746-8094/© 2017 Published by Elsevier Ltd.

nation and intensity of speech. Several studies have been conducted on the latter, because prosody has a role in conveying emotion. Moreover, prosodic feature estimation is fairly robust in noise. Therefore, a system based on a smartphone device, and aiming at analyzing prosodic features in running speech, has been proposed to monitor bipolar patients [8].

Till recently, the study of speech changes in depression was the most often investigated topic [9–13]. Especially, speaking rate has been found to correlate negatively with the Hamilton Depression Rating Scale score [9]. Five possible descriptors, i.e. decreased speaking intensity, decreased vocal frequency range, slower speech, flat intonation and a lack of linguistic stress, have been highlighted in depressed patients by Hollien [11]. F0 contours were perceptually investigated and demonstrated to contain information about a wide range of prosodic information such as F0 variability, speech rate and pause time [13]. On the contrary, the relationship between mania and speech features has been less explored. More precisely, pressured speech, i.e. the tendency to speak quickly and loudly [14], has been observed to be one of the most common symptoms in children and adolescent affected by mania [15]. Recently, more attention has been paid to the analysis of speech changes in bipolar disorder. To our knowledge, two European Projects investigated speech of patients affected by Bipolar Disorder: namely PSYCHE [3] and MONARCA [4]. As a result, several articles on this topic have been published [16–23]. Additional studies focused on that topic are [24,25]. Specifically, intra-subject studies [16,18] reported significant differences in vocal frequency (F0) variability and average between different mood states. Moreover, the speech intonation contour has been found to be a reliable indicator of mood changes from an euthymic to either a depressed or a manic state in [17,20]. A system, based on smartphone-sensing, aiming at the recognition of depressive and manic states and the detection of state changes in patients suffering from bipolar disorder was studied in [21,22,24]. In these studies, statistics about the phone calls, the verbal interaction of the patients with the other talker, and speech-related features, extracted with the open-source “openSmile” toolbox [26] providing several low-level descriptors, were investigated. A good performance in terms of pattern recognition accuracy was obtained [22]. In [25], a smartphone-based system was proposed to monitor bipolar patients in terms of social rhythms by investigating statistics about conversations, speaking rate and F0 changes. Audio, accelerometer and self-assessment related data were used to classify mood states in bipolar patients in day-to-day phone conversation, achieving an accuracy better than 80% [23].

Despite the relevance of these studies, the analysis of speech changes in mood disorders remains a challenging task. Particularly, the direction of the features changes was not always coherent across subjects [17]. The need of selecting patient specific features and building personalized models was stressed in [22] and [24], respectively. Moreover, in some studies, no significant correlation between F0 variables and depression was found [9,27]. The improvement of existing models performance could be obtained both by improving subject status characterization, e.g. by evaluating anxiety level [17,28,29], and by investigating other features. Novel speech features might in fact capture relevant information about the specific phenomenon under investigation [30]. Furthermore, the exploration of novel speech features [30] that parsimoniously describe the phenomenon of interest in a specific application could improve the classification performance, by avoiding overfitting.

The aim of the present study is to explore a new feature set to characterize speech production in patients affected by bipolar disorder. More precisely, a spectral analysis of the F0 contour is proposed to investigate differences in mood states in patients suffering from bipolar disorder. Specifically, a parsimonious description of the F0-contour is presented. The proposed features are related both to modulation of F0 and to speech rhythm, i.e. speech rate, distance between pauses and pause lengths. Moreover, to better characterize the proposed features, we explore whether and to what extent they are related to rhythm features or whether they represent a complementary source of information. Therefore the rhythmic properties of the audio recordings and a correlational analysis between rhythm features and F0-contour features have been studied. The analysis is carried out in patients and healthy controls. Patients have been recorded at different days, while reading a neutral text. In addition, a study of healthy control speakers is presented. Preliminary results are reported and discussed.

Table 1
Patient mood labels.

Subj.	Label day1	Label day2	Label day3
A	Hypomania	Euthymia	
B	Hypomania	Euthymia	Depression
C	Hypomania	Euthymia	
D	Depression	Euthymia	
E	Depression	Hypomania	
F	Depression	Euthymia	
G	Hypomania	Euthymia	
H	Depression	Euthymia	
I	Depression	Euthymia	
L	Depression	Euthymia	
M	Mixed	Depression	Euthymia

2. Methods

2.1. Experimental protocol and data

In this study, eleven patients (five females and six males, 40.00 ± 9.02 years) suffering from bipolar disease were enrolled in the Psyche European project [3]. All were able to lead independent and active lives, and they were free from substance use disorder. Seven psychiatric patients were French native speakers, while four of them were Italian native speakers. The experimental protocol, approved by the clinical ethical committee, consisted in the reading of a neutral text during each recording session. Sessions were held at two or three different days. Seven patients out of eleven (patients A–G) (Table 1) were recorded twice each day. A physician labeled the patient’s mood status before each recording using clinician-administered rating scales. Four different mood states were identified in this study, namely depressed, euthymic, hypomanic and mixed. A high quality directional microphone was used to record signals at a sampling frequency of 48 kHz and with a resolution of 32 bits (AKG Perception P220 Condenser Microphone, M-Audio Fast-Track).

Eighteen healthy control subjects (nine males and nine females, 30.00 ± 5.00 years) were enrolled. Healthy control subjects did not report any actual or past psychiatric disorder, and had no history of neurological or major somatic conditions. They were recorded twice at two different days to test for inter-day variability. Typically, the second session took place 7 days after the first one. All healthy subjects were Italian native speakers.

The CMU Arctic Database [31] was used to evaluate the performance of a Voice Activity Detection (VAD) algorithm that is used to classify audio frames. The database provides audio and electroglottographic (EGG) recordings. The EGG signal is a signal that is related to the impedance changes during vocal folds contact. The corpus is formed of about 1100 short sentences, comprising more than 8000 vowels. Audio and EGG recordings were sampled at a rate of 32 kHz with a resolution equal to 32 bit.