



# Prediction of major depression in adolescents using an optimized multi-channel weighted speech classification system



Kuan Ee Brian Ooi<sup>a</sup>, Margaret Lech<sup>a,\*</sup>, Nicholas Brian Allen<sup>b</sup>

<sup>a</sup> School of Electrical and Computer Engineering, Royal Melbourne Institute of Technology, 3001 Victoria, Australia

<sup>b</sup> Department of Psychological Sciences, University of Melbourne, Victoria 3010, Australia

## ARTICLE INFO

### Article history:

Received 17 April 2014

Received in revised form 8 July 2014

Accepted 7 August 2014

Available online 30 August 2014

### Keywords:

Multi-channel classification  
Prediction of clinical depression  
Classifier optimization  
Weighted features  
Decision fusion  
Speech classification

## ABSTRACT

This study addresses an urgent need for objective measures allowing an efficient, early prediction of risk for depression in adolescents. An early intervention preventing the onset of clinical depression could significantly reduce the social and economic burden of the disease. Previous studies have shown that acoustic speech parameters are strong indicators of full blown depression symptoms in adults and adolescents. The current study investigates the effectiveness of acoustic speech analysis and classification in prediction of depression in adolescents before the full blown symptoms become apparent. The proposed optimized multi-channel weighted speech classification (OMCWSC) method introduces a two-stage multi-channel classification procedure. In the first stage, each channel of the OMCWSC performs an independent classification based on a single type of features (glottal, prosodic and Teager energy operator parameters derived from glottal waveform (TEOG)). In the second stage, the weighted classification outcomes from all channels are combined to deliver the final decision classifying an individual as “At Risk” (AR) or “Not at Risk” (NAR) of developing depression symptoms within the next 2.5 years. The weight values of the OMCWSC system are optimized to maximize the overall system classification accuracy. Experiments based on speech recordings collected from 15 AR and 15 NAR adolescents showed that while each of the three individual channels provided prediction accuracy above the chance level (69% glottal, 63% prosodic, 67% TEOG), the adaptive weighting system of the proposed OMCWSC procedure led to further improvement of the prediction results achieving up to 74% accuracy and a well-balanced sensitivity to specificity ratio of 77%/70%.

Crown Copyright © 2014 Published by Elsevier Ltd. All rights reserved.

## 1. Introduction

The diverse nature of depression leads to numerous research challenges, including the understanding of its causes and effects, as well as the development of efficient treatments and prevention methods.

Although depressive disorders are more often a problem in adults, statistics have shown that children aged as early as 4 years can show the first signs of depression and the numbers increase towards adolescence [49].

If untreated, depression significantly reduces the social and vocational opportunities of affected individuals and in some cases can lead to suicide attempts. It is therefore vitally important that very early signs of depression can be detected to facilitate the

application of preventative measures to either eliminate or reduce the risk of developing full blown symptoms of depression.

This study describes an effective prediction system that is able to determine in objective terms, if an adolescent is at possible risk of clinical depression in the near future. The objective character of the proposed system comes from the fact that the prediction is based on the numerical analysis of acoustic speech parameters. Just as for example, blood analysis can determine if a patient suffers from a particular disease, the analysis of speech sounds can provide vital diagnostic information facilitating the assessment of the patient's mental health state. In other words, the speech signal is being treated here as a carrier of diagnostic information.

Aside from containing important diagnostic information, speech signals are very easy to capture in a discrete and noninvasive way. Moreover, the speech based mental health diagnosis (or prediction) is a low cost procedure and can be conducted in a fully automatic way by a person that does not necessarily have an extensive experience in mental health care. These attributes make the proposed system particularly suitable as a mass screening device for detecting

\* Corresponding author. Tel.: +61 3 99251028.

E-mail address: [margaret.lech@rmit.edu.au](mailto:margaret.lech@rmit.edu.au) (M. Lech).

“at risk” populations of adolescents. The proposed methodology can be also extended to testing of war veterans, elderly people, young mothers and other groups particularly vulnerable for depression.

The link between acoustic parameters in speech signals and depression has been extensively researched in relation to understanding the psychophysical mechanisms of depression, and searching for acoustic cues that allow discrimination between depressed and non-depressed individuals at the stage when full blown symptoms are already present [15,34,35,37,38,46].

The present study further extends these investigations and shows that the acoustic speech analysis has the potential to detect (or predict) very early signs of depression before full-blown symptoms have developed. This goal has been achieved through the analysis of speech signals of adolescents who were according to the DSM-IV criteria [5], healthy (non-depressed) at the time of data collection. The analysis showed clear differences between the speech characteristics of individuals who two years later were diagnosed with the DSM-IV criteria as depressed and those who were diagnosed as still non-depressed. These differences provided the basis for predictive classification of speech up to two years before the more advanced full-blown symptoms are developed. This study also investigates which of the acoustic speech parameters (or characteristic features) provide the most accurate prediction results within a single-channel classification framework (i.e. using a single type of acoustic features). The single-channel prediction results are further improved by the introduction of a new computational methodology based on the optimized multi-channel classification system with an optimized weighted decision procedure. The proposed methodology is tested using speech data collected from a community sample including 15 “At Risk” and 15 “Not at Risk” of depression adolescents.

The remaining sections are organized as follows. Section 2 provides a brief overview of the existing depression prediction methodologies. Section 3 describes the database used to test the prediction algorithms. Section 4 explains the single-channel and the proposed multi-channel prediction techniques. Section 5 describes the test results and their interpretation, and Section 6 provides the conclusions.

## 2. Previous work

Over recent years, the practice of delivering mental health services has changed dramatically. Medical and psychological assessment is being automated step-by-step and consultation is increasingly conducted via the Internet.

Given the prevalence of mental health disorders and the burden on economies worldwide, mental health informatics has become a dedicated branch of engineering and applied computer science. Automated, easy to use mental health diagnostic and prediction techniques are sought after for online and mobile phone applications. Mobile phones can record and transmit speech, text and movement data. They can store and communicate personal information, including medical diagnostic information. For the first time in history, objective data about the behaviour and mental health of hundreds of millions of individuals can be recorded, stored and analyzed. With the help of dedicated computer software, it also provides a new unique opportunity to predict and prevent mental health problems within large populations [31].

Speech and language are particularly interesting from the viewpoint of psychological assessment. For instance, depression may change the characteristics of the voice of individuals and these changes can be detected by a special form of speech analysis. Computational screening methods that utilize speech and language can detect subtle changes and alert clinicians as well as individuals and caregivers.

In many cases, the life-long reoccurrences of depression during adulthood are consequences of untreated or undetected depression symptoms appearing for the first time during childhood or adolescence [42]. The onset of depression is most likely between 13 and 17 years of age, with major depression (MD) being the most common form. An early intervention preventing the onset of clinical depression would provide a very important method for reducing the burden of the disease. Efficient depression prediction techniques are therefore needed to determine the risk for depression in young individuals.

Current treatment and prevention methods in adolescents focus on the study of the behavioural and psychological states of an individual [42,48,57]. Some of the often-used prediction approaches include the identification of characteristics such as family history, early adversity, gender, age or socioeconomic status [12,17,22,42]. Numerous studies have undertaken the task of recognizing early behavioural and psychological signs and symptoms of the disorder that are not classified as full-blown symptoms [39–41]. A longitudinal study of 230 late adolescents identified increased cortisol awakening response (CAR) as a significant prospective risk factor for the development of depression in young adults [1].

In [30], the concept of emotional inertia characterizing a specific type of psychological maladjustment showing high resistance to emotional changes was investigated. People with high emotional inertia are resistant to external influences and preserve their current emotional state for a long time. It was found that depressed individuals in particular tend to have higher levels of emotional inertia. In [30] emotional inertia was shown to predict the emergence of clinical depression in adolescents 2.5 years later.

The use of acoustic speech analysis to recognize symptoms of depression in adolescent [34] and adult [3,10,15,37,38,46,51,52], has shown promising results on the case where individuals have already experience full-blown disorder episodes. Depending on gender and the type of acoustic features used in the recognition process, the binary (depressed versus non-depressed) classification accuracy can be as high as 80–90%. The proposed depression diagnosis had a single-channel character, meaning that only a single type of features or a single combination of features was used to determine if the speaker is depressed or not. Within this single-channel framework, features such as the glottal speech parameters [34,37,38], as well as parameters derived from the Teager energy operator (TEO) [34,35] were found to be strongly correlated with depression symptoms.

Acoustic speech analysis has also been used to predict the risk for depression in adolescents up to 2 years before full-blown depressive symptoms become evident [32,43,44]. Ooi et al. [43] introduced a multi-channel classification approach for the prediction of depression in adolescents. The results led to the highest to date speech based prediction accuracy of 73% and sensitivity to specificity ratio of 79%/67% (these parameters are defined later in Section 5.1). In comparison, the reported speech based detection of full-blown symptoms of depression can achieve an accuracy of up to 87% (with sensitivity to specificity 81%/92%) for males and 79% (with sensitivity to specificity 81%/77%) for females [34]. These results indicate that the early prediction of depression appears to be more challenging, most likely due to less defined differences between acoustic parameters characterizing individuals that are at risk and individuals that are not at risk of developing depression in the near future. The multi-channel depression prediction approach in [43] used a weighted combination of four types of acoustic features (prosodic, glottal, TEO and spectral) to determine the final classification decision. The current study extends upon the approach described in [43] by introducing a new improved way of determining the weight values for the individual channels. The speech classification system in [43] used a weighted combination of the single-channel outputs to derive

Download English Version:

<https://daneshyari.com/en/article/558021>

Download Persian Version:

<https://daneshyari.com/article/558021>

[Daneshyari.com](https://daneshyari.com)