Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 71 (2015) 10–49

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

A review of depression and suicide risk assessment using speech analysis

Nicholas Cummins [a,b,*], Stefan Scherer [c], Jarek Krajewski [d,e], Sebastian Schnieder [d,e], Julien Epps [a,b,*], Thomas F. Quatieri [f]

[a] School of Elec. Eng. and Telecomm., The University of New South Wales, Sydney, Australia
[b] ATP Research Laboratory, National ICT Australia (NICTA), Australia
[c] University of Southern California, Institute for Creative Technologies, Playa Vista, CA 90094, USA
[d] Experimental Industrial Psychology, University of Wuppertal, Wuppertal, Germany
[e] Industrial Psychology, Rhenish University of Applied Sciences Cologne, Germany
[f] MIT Lincoln Laboratory 244 Wood Street, Lexington, MA 02421, USA

Received 12 September 2014; received in revised form 27 March 2015; accepted 30 March 2015
Available online 6 April 2015

Abstract

This paper is the first review into the automatic analysis of speech for use as an objective predictor of depression and suicidality. Both conditions are major public health concerns; depression has long been recognised as a prominent cause of disability and burden worldwide, whilst suicide is a misunderstood and complex course of death that strongly impacts the quality of life and mental health of the families and communities left behind. Despite this prevalence the diagnosis of depression and assessment of suicide risk, due to their complex clinical characterisations, are difficult tasks, nominally achieved by the categorical assessment of a set of specific symptoms. However many of the key symptoms of either condition, such as altered mood and motivation, are not physical in nature; therefore assigning a categorical score to them introduces a range of subjective biases to the diagnostic procedure. Due to these difficulties, research into finding a set of biological, physiological and behavioural markers to aid clinical assessment is gaining in popularity. This review starts by building the case for speech to be considered a key objective marker for both conditions; reviewing current diagnostic and assessment methods for depression and suicidality including key non-speech biological, physiological and behavioural markers and highlighting the expected cognitive and physiological changes associated with both conditions which affect speech production. We then review the key characteristics; size, associated clinical scores and collection paradigm, of active depressed and suicidal speech databases. The main focus of this paper is on how common paralinguistic speech characteristics are affected by depression and suicidality and the application of this information in classification and prediction systems. The paper concludes with an in-depth discussion on the key challenges – improving the generalisability through greater research collaboration and increased standardisation of data collection, and the mitigating unwanted sources of variability – that will shape the future research directions of this rapidly growing field of speech processing research.

Keywords: Depression; Suicide; Automatic assessment; Behavioural markers; Paralinguistics; Classification

Contents

* Corresponding authors at: School of Elec. Eng. and Telecomm., The University of New South Wales, Sydney, Australia. Tel.: + 61 2 9385 6579 (J. Epps).
E-mail addresses: n.p.cummins@unsw.edu.au (N. Cummins), j.epps@unsw.edu.au (J. Epps).

http://dx.doi.org/10.1016/j.specom.2015.03.004
0167-6393/

1. Introduction

Clinical depression is a psychiatric mood disorder, caused by an individual's difficulty in coping with stressful life events, and presents persistent feelings of sadness, negativity and difficulty coping with everyday responsibilities. In 2002 the *World Health Organisation* (WHO) listed unipolar depression as the fourth most significant cause of disability worldwide, and predicted it will be the second leading cause by 2030 (Mathers and Loncar, 2006). Olesen et al. (2012) estimated that in 2010 the cost per patient of depression in Europe was €24,000 (in terms of relative value assessed across 30 European countries) and the total cost of depression in the European Union was €92 billion,

with €54 billion of this cost due to lost work productivity. Similarly in the United States, Stewart et al. (2003) estimated that in 2002, workers with depression cost the U.S. economy $44 billion in lost productivity (due either to absence from work or reduced performance), an increase of $31 billion when compared to the lost productivity costs for U.S. workers without depression.

Suicide is the result of a deliberate self-inflicted act undertaken with the intent to end one's life. Recently WHO estimated that over 800,000 people die from suicide every year and there are at least 20 times more attempted suicides (World Health Organisation, 2014). Often a private act, suicide has a profound negative impact on lives of those left behind; it is estimated that a single suicide intimately affects at least 6 other people (McIntosh, 2009). Despite the high socio-economic costs inflicted onto affected individuals, families and communities it remains a misunderstood and under-researched cause of death.

Depression often places an individual at higher risk of engaging in suicidal behaviours (Hawton et al., 2013; Lépine and Briley, 2011). It has been estimated that up to 50% of individuals who commit suicide meet criteria for a clinical diagnosis of a depressive illness (Joiner et al., 2005; McGirr et al., 2007). Given the associations between depression and suicide, effective diagnosis and treatment of depression has been identified as a key strategy in suicide prevention (Mann et al., 2005). Specific screening for suicide risk should also be undertaken for all individuals undergoing assessment or treatment for depression (Hawton and van Heeringen, 2009).

As there is no single clinical characterisation of a suicidal or depressed individual; this makes the diagnosis of both suicide risk and depression subjective in nature and time consuming. Gold-standard diagnostic and assessment tools for depression and suicidality remain rooted, almost exclusively, on the opinion of individual clinicians risking a range of subjective biases. Commonly used diagnostic tools include interview style assessment such as the Hamilton Rating Scale for Depression (Hamilton, 1960) and Suicide Probability Scales (Cull and Gill, 1982). These assessments rate the severity of symptoms and behaviours observed in depression or suicidality to give a patient a score which relates to their level of depression or suicide risk.

Diagnosis using this style of testing is complicated; it relies heavily on the ability, desire and honesty of a patient to communicate their symptoms, moods or cognitions when, by definition, their outlook and motivation are impaired. Therefore diagnostic information is time consuming to gather and requires a large degree of clinical training, practice, and certification to produce acceptable results (Mundt et al., 2007). Currently there is no objective measure, with clinical utility, for either depression or suicidality; this compromises optimal patient care, compounding an already high burden on health, social, and economic services.

To enhance current diagnostic methods an objective screening mechanism, based on biological, physiological and behavioural signals is needed. A wide range of biological markers such as low serotonin levels (Nock et al., 2008; Sharp and Cowen, 2011), neurotransmitter dysfunction (Luscher et al., 2011; Poulter et al., 2008) and genetic abnormalities (Dwivedi et al., 2003; Gatt et al., 2009) have been associated with depression and suicidal behaviour, however to date no specific biomarker has been found. Whilst biomarkers remain elusive, significant recent advances have been made in using affective computing and social signal processing as a diagnostic tool, for depression specifically (Cohn et al., 2009; Cummins et al., 2013c; Joshi et al., 2013; Scherer et al., 2013c; Williamson et al., 2013). These systems rely in particular on facial and body tracking algorithms to capture characteristic behavioural changes relating to depression.

In recent years, the problem of automatically detecting mental illness using speech, more specifically non-verbal paralinguistic cues has gained popularity. Speech is an attractive candidate for use in an automated system; it can be measured cheaply, remotely, non-invasively and non-intrusively. Clinicians often (subjectively) use the verbal behaviour of a patient during diagnosis; decreased verbal activity productivity, a diminished prosody and monotonous and "lifeless" sounding speech is indicative of depression (Hall et al., 1995; Sobin and Sackeim, 1997). Similarly it has been reported that as an individual becomes suicidal their speech quality changes to a hollow and toneless sound (Silverman, 1992).

Similar speech processing frameworks are likely to be effective when applied to assessment of either depression or suicide risk. Such a tool could be of great use in primary health care settings. Between 50% and 70% of individuals experiencing depression will consult their primary health care provider (Sikorski et al., 2012). However it has been estimated that General Practitioners have only a 50% success rate when diagnosing depressed individuals (Mitchell et al., 2009). Additional methods for early diagnosis could have a significant effect on suicide prevention, given that in up to 66% of suicides, patients have contacted a primary health care provider within a month prior to their death (Mann et al., 2005).

Whilst there has been significant research into correlations between prosodic, articulatory and acoustic features of speech and clinical ratings of both depression (Cummins et al., 2011; Flint et al., 1993; Low et al., 2011; Moore et al., 2008; Mundt et al., 2007; Nilsonne, 1988; Scherer et al., 2013b; Trevino et al., 2011; Williamson et al., 2013) and suicidality (Scherer et al., 2013a; Silverman and Silverman, 2006; Silverman, 1992) as well as work on the automatic analysis of speech as a predictor for both conditions (France et al., 2000; Ozdas et al., 2004a; Yingthawornsuk et al., 2007), there has never been an extensive review of this literature.

In order to investigate how speech might be used to index or classify depression or suicidality, it is necessary

to first understand how current diagnostic methods are used and what aspects of these may be relevant to speech analysis. As speech potentially represents just one diagnostic aid modality it is important to highlight current research into associated biological, physiological and behavioural markers as to gain understandings as to how speech could be used to augment systems analysis methods based on these systems. It is also instructive to review the characteristics of depressed and suicidal speech databases, to understand what kinds of data collection protocols and objectives already exist and are well suited for research in this area.

Primarily this literature review has been carried out to discuss the suitability of speech based features as a marker for both conditions and to review the investigations that have been carried out into the automatic analysis of speech as a predictor of suicidality and depression. This article concludes with some of the major challenges and potential future research directions associated with this potentially lifesaving field of speech processing research.

2. Current diagnostic methods

2.1. Defining clinical depression

The *Diagnostic and Statistical Manual of Mental Disorders* (DSM) published by the American Psychiatric Association is the most widely used resource in the diagnosis of mental disorders. It was first published in 1952 and is currently in its 5th edition (American-Psychiatric-Association, 2013). The DSM was designed to provide a common language and standard criteria for the classification of mental disorders, classifying disorders by their observed symptoms and the clinical course of the disorder.

The exact causes of depression are not universally agreed upon, it is generally considered to be a dysfunction - reduced activity and connectivity - of cortical-limbic systems (Deckersbach et al., 2006; Evans et al., 2006; Mayberg et al., 2005; Niemiec and Lithgow, 2005), resulting from interactions between a genetic predisposition and environmental factors including stress and emotional trauma (Nestler et al., 2002). Whilst most people feel some form of depression in their life, it is considered an illness, according to the DSM definition, when an individual has either a depressed mood or markedly diminished interest or pleasure in combination with four or more symptoms, shown in Table 1, for longer than a two-week period.

The DSM has often been criticised for the homogenous way it defines the boundaries between mental illnesses, leaving diagnosis open to subjective biases where a proper patient assessment does not have to be done to achieve a diagnosis (Brown et al., 2001; Kamphuis and Noordhof, 2009; Lux and Kendler, 2010; Oquendo et al., 2008; Stein et al., 2010; Watson, 2005). Noting that at least four of the DSM symptoms listed in Table 1 comprise two distinct manifestations, there are at least 1497 unique profiles of depression (Østergaard et al., 2011). Using the DSM it is possible for two depressed individuals, sharing no overlapping symptoms, to receive the same diagnosis (Balsters et al., 2012).

Table 1
Symptoms associated with depression (American-Psychiatric-Association, 2013).

Depressed Mood *and/or* Markedly diminished interest or pleasure
In combination with four of
Psychomotor retardation OR agitation
Diminished ability to think/concentrate OR Increased indecisiveness
Fatigue OR Loss of energy
Insomnia OR hypersomnia
Significant weight loss OR weight gain
Feelings of worthlessness OR Excessive/inappropriate guilt
Recurrent thoughts of death OR Recurrent suicidal ideation

2.2. Diagnosing depression

The diagnosis of depression, especially in primary care settings, is difficult. The large variation in depression profiles introduces a large degree of complexity when trying to fit the clinical profile of a depressed individual into an objective categorical level i.e. low or high level depression (Mitchell et al., 2009). Diagnosis is often further complicated by; the modest rate of clinical depression seen in primary settings increasing the chances of misidentifications, the time consuming nature of diagnosis, physical symptoms masking their underlying cause, and not all depressed patients outwardly express emotional symptoms such as sadness or hopelessness (Mitchell et al., 2009; Schumann et al., 2012).

2.2.1. Assessment tools for depression

Commonly used assessment tools include interview style assessments such as the *Hamilton Rating Scale for Depression* (HAMD, Hamilton (1960)) or self-assessments such as the *Beck Depression Index* (BDI) originally published in 1961 and revised in 1996 (Beck et al., 1996). Both assessment methodologies rate the severity of 21 symptoms observed in depression, to give a patient a score which relates to their level of depression. The major differences between the two scores are that HAMD is a clinician-rated questionnaire that can be completed in 20–30 min, while BDI is a self-reported questionnaire which can be completed in as little as 5–10 min. Both scales use different items; the HAMD favours neuro-vegetative symptoms (symptoms that affect an individual's day to day functioning such as weight, sleep, psychomotor retardation and fatigue) whilst the BDI favours negative self-evaluation symptoms, and different weighting schemes when producing their total score. Both assessments have been shown to have predictive validity and consistency when differentiating depressed from non-depressed patients (Cusin et al., 2010; Maust et al., 2012).

The HAMD has long been regarded as the gold standard assessment tool for depression for both diagnosis and

research purposes, although this status continually comes into question (Bagby et al., 2004; Gibbons et al., 1993; Maust et al., 2012). The HAMD assessment rates the severity of symptoms observed in depression, such as low mood, insomnia, agitation, anxiety and weight loss, to give a patient a score which relates to their level of depression. The clinician must choose the possible responses to each question by interviewing the patient and observing the patient's symptoms. Each of the 21 questions has 3–5 possible responses which range in severity; scored between 0–2, 0–3 or 0–4 depending on the importance of the symptom they represent (see Appendix A for HAMD questionnaire). The scores are summed and the total is arranged into 5 categories; *Normal* (0–7), *Mild* (8–13), *Moderate* (14–18), *Severe* (19–22) and *Very Severe* ($\geqslant$23).

Whilst the aggregate HAMD score has been proven to have predictive validity for determining if an individual is suffering depression, it is often criticised for its inability to rank patient in terms of their severity status (Gibbons et al., 1993) due in part the favouring neuro-vegetative symptoms (Maust et al., 2012) and weak interrater and retest coefficients at the individual item/question level (Bagby et al., 2004). Further, there is evidence that the HAMD total score is less sensitive measure of overall depressive severity than subscales derived from a more limited subset of scale items (Bech et al., 1981; Faries et al., 2000). Bech et al. (1981) show the *Melancholia Subscale* – consisting of only six items out of the HAMD; depressed mood, guilt, work and interests, retardation, psychic anxiety, and general somatic symptoms is as sufficient as the HAMD total when measuring severity of depression. Faries et al. (2000) report similar results for the *Maier and Phillips Severity Subscale* and *Gibbons Global Depression Severity Subscale*.

The BDI is the most widely use self-reported measure of depression (Cusin et al., 2010). It contains 21 items covering key cognitive, affective, and somatic symptoms observed in depression with a heavy focus on negative self-evaluations such as self dislikes and self-criticisms. Each question is scored as 0, 1, 2, or 3 depending on how the patient perceived the severity of that symptom over the previous week. The score range is 63 and is categorised 4 ways; *Minimal* (0–9), *Mild* (10–18), *Moderate* (19–29), *Severe* ($\geqslant$30). Although not specifically designed for use in primary care settings, results in Nuevo et al. (2009) show that it has reliability and validity for use in the general population. Whilst self-evaluated assessments offer greater convenience over clinician led assessments, in that they require less time and training to administer, their reliability can be reduced due to patient over-familiarity and patient reading ability (Cusin et al., 2010).

A range of depression ratings, Table 2, including the HAMD and BDI are discussed further in Maust et al. (2012), and other rating scales with clinical validity they discuss include: 10-item *Montgomery–Åsberg Depression Rating Scale* (MADRS), the 16-item *Quick Inventory of Depressive Symptomology* (QIDS) and the 9-item*Patient Health Questionnaire* (PHQ-9).

2.3. Defining suicidality

Suicidality is defined as a behaviour not an illness or a disease, and it is believed to be the result of individual susceptibility due to a wide range of triggering events and intertwined risk factors (Table 3). Suicidal behaviours range on a complex spectrum from suicidal thoughts, to non-fatal attempts, through to fatal acts, with all varying in terms of intention, impulsivity, seriousness and lethality (Goldney, 2008; Mann, 2003). It is believed that a similar cognitive process underlies most suicide attempts. Theories that attempt to explain this processes include the *Interpersonal Theory* (Van Orden et al., 2010), the *Cry of Pain Model* (Florentine and Crane, 2010) and the *Stress-Diathesis Model* (Mann, 2003). Common themes amongst these models include stressful life events, mental health issues, feelings of hopelessness and impulsivity and aggression.

As the exact underlying causes of both depression and suicidality are unknown, drawing parallels between the neurobiological factors which increase an individual's susceptibility to either depression or suicidality is a complicated task. The relationship between the two conditions is discussed in Section 2.5 and potential underlying neurobiological causes in Section 3.1.

Suicidal risk factors are defined as sufficient but not necessary conditions needed for an individual to engage in suicidal behaviours, they represent a long-term risk (months/years) for engaging in suicidal behaviours. Due to ethical and logistical challenges associated with studying suicidality, more data are needed to identify empirical correlations between risk factors and suicidality (Prinstein,

Table 2
Commonly used depression rating scales with predictive validity (Maust et al., 2012).

Scale	Acronym	Reference	Clinician-Led	Self-Report	Number of items	Minutes to complete
Hamilton Rating Scale for Depression	HAMD	Hamilton (1960)	√		17 or 21	20–30
Beck Depression Inventory	BDI	Beck et al. (1996)		√	21	5–10
Montgomery–Åsberg Depression Rating Scale	MARSD	Montgomery and Asberg (1979)	√		10	20–30
Quick Inventory of Depressive Symptomology	QIDS	Rush et al. (2003)		√	16	5–10
Patient Health Questionnaire	PHQ-9	Kroenke et al. (2001)		√	9	<5

Table 3
Triggering events and risk factors for suicidal behaviour.

Trigger event	Risk factors
Death of significant other	Mental illness
Physical illness	Substance abuse/intoxication
Conflict/victim of violence	Genetic factors/family history of suicide
Separation/loss of partner	Lowered serotonin levels
Legal problems/imprisonment	Previous attempts/exposure to suicide
Work related problems/ retirement	Combat exposure
Financial problems	Childhood maltreatment

2008). Known risk factors include: an individual's age and gender (younger 15–24 year olds and older 65+ males are at higher risk of suicide (Brendel et al., 2010; Cantor, 2008)), an individual's psychiatric diagnosis, (Brendel et al., 2010), a history of suicide behaviour (Brendel et al., 2010), a family history of suicide (Brendel et al., 2010) and personality type (Van Orden et al., 2010).

An individual who is having a suicidal crisis is at imminent risk (minutes, hours, days) of attempting suicide. Periods of exceptionally high risk are relatively short lived; on average attempts can be contemplated for as little as 5 min up to 1 h (Florentine and Crane, 2010). A pre-suicidal individual may display a range of intense affective states such as desperation, extreme hopelessness, feelings of abandonment, self-hatred, rage, anxiety, loneliness and guilt (Hendin et al., 2007). It is not understood why people attempt suicide. According to the Stress-Diathesis Model there exists a threshold level which varies for every individual. This threshold is determined by a wide range of factors including the individual's genetic predisposition, biochemical factors and personality traits as well as their physiological and emotional state at the time of a suicidal crisis (Mann, 2003).

2.4. Suicidal risk assessment

The aim of assessment is to quantify the suicidal ideation level of an at-risk individual. Assessment is complicated by the range of suicidal behaviours which includes initial ideation, non-fatal attempts and completed suicides. Bisconer and Gross (2007) list a range of socio-demographic and clinical factors associated with a complete assessment of suicidality including: assessing a patient's mental status, history of psychiatric illness, history of substance abuse, suicide ideation and history, parasuicidal behaviours and social support systems; see Appendix B for the complete list.

Assessment is extremely difficult, and the lack of specificity factors associated with suicide behaviour and the infrequency of attempts even among at-risk patients often results in a high number of false positive predictions (Goldney, 2008). Some suicidal patients may present an outwardly calm façade: deciding to attempt suicide can end a patient's anguish so they will openly lie about their intent, not wanting intervention resulting in a high number of false negative predictions (Hendin et al., 2001). Further complicating the task are factors such as irrational reasoning for engaging in suicidal behaviours (Maris, 2002), distrust and poor communication between patients and clinicians (Hendin et al., 2001), the honesty of the patients when completing the questionnaire (Cochrane-Brink et al., 2000), varying suicide ideation level (Beck et al., 1999) and periods of highest risk being short lived contemplation for attempts ranging between 5 min up to 1 h (Florentine and Crane, 2010).

2.4.1. Assessment tools for suicidal risk

Predictive scales for suicide risk have been developed, however on the basis of aggregate data, with a lack of specificity for subsequent suicidal behaviour in the individual person (Goldney, 2008). Two clinical predictive scales with predictive validity that attempt to account for the wide range of socio-demographic and clinical factors associated with suicidality include the *Suicide Probability Scales* (SPS, Cull and Gill (1982)) and the *Reasons for Living Inventory* (RFL, Linehan et al. (1983)). The SPS is a 36 item questionnaire, answered on a 4 point scale, designed to be a measure of suicidal risk. The items are designed to capture particular feelings and behaviours related to suicidal behaviours such as level of suicide ideation, hopelessness, hostility and anger impulsivity. The RFL is a 48 item questionnaire, answered using the Likert Scale, which is arranged into six subscales: survival and coping beliefs, responsibility to family, child concerns, fear of suicide, fear of social disproval, and moral objections. A lower overall score, the average of the 6 subscale scores, indicates a more imminent risk of the patient attempting suicide.

Although not specifically designed for assessing suicidality, the *Beck Hopelessness Scale* (BHS, Beck and Steer (1988)) has been found to have proven predictive validity. The BHS is a 20-item self-report inventory designed to measure feelings about the future, loss of motivation, and future expectation. The BHS has been found to be one of the better predictors of suicide up to a year in advance (Cochrane-Brink et al., 2000) and has been shown to be a better predictor than scales designed specifically for assessing suicidal behaviour (Bisconer and Gross, 2007).

2.5. Depression and suicidality

Suicide is a more common cause of death for individuals with psychiatric disorders than it is in the general population; it has been estimated that up to 90% of suicide attempters have a diagnosable psychiatric condition (Brendel et al., 2010; Mann, 2003). Psychiatric disorders associated with increased suicide risk include mental illnesses such as depression, bi-polar disorder and schizophrenia, and substance abuse. But despite the high prevalence in suicidality, psychiatric disorders cannot be regarded as a reliable predictor of suicide as only a small number of

people diagnosed with a psychiatric disorder will engage in suicidal behaviour. Brown et al. (2000) found over a 20 year period, in a prospective study to determine suicidal risk factors, there were 49 (1%) suicides and 170 (3%) natural deaths in a group of 6891 psychiatric outpatients.

Among psychiatric disorders associated with suicidality, depression is the most strongly associated (Hawton et al., 2013; Lépine and Briley, 2011). Patients with unipolar depression are up to twenty times more likely to commit suicide than an individual who has never had an affective or mood disorder (Lépine and Briley, 2011). Risk factors for depressed individuals engaging in suicidal behaviours include: depression severity, increased feelings of hopelessness, negative ruminations and desires of escapism, gender (males have a greater tendency to commit suicide), substance abuse, a family history of suicide and psychiatric disorders and a history of suicidal ideation and suicide attempts (Beck and Alford, 2008; Bolton et al., 2010; Hawton et al., 2013; McGirr et al., 2007). Depression also greatly increases an individual's chances of engaging in non-lethal suicidal behaviours (Kessler et al., 2005; Verona et al., 2004).

Despite the strong links, suicides are still infrequent amongst depressed individuals. Mattisson et al. (2007) found over a 50 year period, in a study of personality traits and mental disorders, a suicide rate of 5% in of a cohort of 344 depressed patients. In a similar study, Coryell and Young (2005) report, over a 21 year period, a suicide rate of 4.2% for a cohort of 785 depressed patients. Neither depression nor suicidality are necessary and sufficient conditions for each other; not everybody with depression commits suicide and not everybody who commits suicide is depressed. It is unclear as to why this is as the risk factors identified for increased suicide risk in depression are similar to those for suicide in general (Bolton et al., 2010; Hawton et al., 2013). Therefore the continual assessment of suicide risk for patients undergoing depression treatments is an integral part of suicide prevention in this at-risk group (Hawton and van Heeringen, 2009; Hawton et al., 2013; Mann et al., 2005).

3. Objective markers for depression and suicidality

Diagnosis on the basis of a set of observable behavioural signals or measurable biomarkers has not been fully embraced in mainstream psychiatry; however research in these fields is gaining in popularity. The focus of Section 3 is on non-speech based markers for depression and suicidality. Given the wide clinical profile of both conditions, it is likely that a multifaceted approach will be needed to find a true objective marker (Section 8); therefore it is important to briefly review the current research into associated biological, physiological and behavioural markers.

Objective markers of either condition have a wide range of potential uses in psychology. They could be used as an objective diagnostic aid in primary care settings and specialist clinics, provide a low effort tool for the automatic gathering and analysing of information over large populations and to provide immediate feedback and therapeutic advice for susceptible people. Proper sensing of non-verbal behavioural based cues could be used in a smartphone platform for remote diagnosis, assist in emergency triage situations as well as also provide support for an interactive virtual tool for use as diagnostic training aids (Scherer et al., 2013c). New areas of research could open up through the design and building of new scientific tools, such as building of diagnostic aids or tools to help locate potential markers. Ultimately research into objective biological, physiological and behavioural markers has the potential to help reduce the large socio-economic costs associated with these conditions and improve the psychiatric diagnosis accuracy and treatment efficacy for many mental illnesses (Balsters et al., 2012; Costanza et al., 2014).

3.1. Biological and physiological markers

The identification of biological and physiological markers associated with depression and suicidal behaviour is a key component in the search of objective markers for both conditions. However, due to the heterogeneity, range of comorbid conditions and the lack of simplistic one-to-one relationships between genes and behaviours (Hasler et al., 2004), this search has had limited success. The best known biological marker for both conditions is a lowered level of the molecular marker serotonin (Nock et al., 2008; Sharp and Cowen, 2011). However, whilst lower serotonin levels are reported in individuals who are depressed or suicidal, it is a non-specific marker and at best appears to represent a vulnerability to either mental state. It can also occur in healthy individuals with a family history of mental illnesses (Åsberg, 1997) and people who engage in impulsive and/or aggressive behaviours (Mann, 2003; Nock et al., 2008; Roy et al., 2008). Low functioning of the neurotransmitter *gamma-amino butyric acid* (GABA) has been linked with a vulnerability to both depression (Croarkin et al., 2011) and suicidality (Poulter et al., 2008). GABA transmission is also associated with increased stress reactivity (Hasler and Northoff, 2011) and inhibitory feedback loops (Croarkin et al., 2011).

3.1.1. Markers for depression

Neuroimaging studies suggest the interactions between the cortical and limbic systems play a major role in the onset of depression (Evans et al., 2006; Kemp et al., 2008; Niemiec and Lithgow, 2005; Sheline, 2003). Interactions between the *brain-derived neurotropic factor* (BDNF) gene and stress exposure have been linked to increased risk for comorbid depression and anxiety disorders (Gatt et al., 2009). There is evidence for the role of the basal ganglia, the area in the brain responsible for motor co-ordination connected to both the limbic system and prefrontal cortex, in the development of depression (Ring, 2002). Depression has also been associated with

small hippocampal volumes (MacQueen and Frodl, 2011). Frodl et al. (2007) found, in a sample of 60 depressed patients, smaller hippocampal volume when compared to the same number of, age and gender matched, controls. The authors also report, for combined cohort of both depressed patients and matched controls, that smaller hippocampal volumes are linked with the BDNF gene, concluding that BDNF abnormalities may represent a marker for increased risk of depression.

Other biomarkers associated with depression include molecular markers; such as insulin and serum (Domenici et al., 2010; Schmidt et al., 2011), protein molecules such as cytokines (Schmidt et al., 2011), and steroid hormones such as salivary cortisol levels (Owens et al., 2014). Physiological biomarkers include galvanic skin responses (Schneider et al., 2012), cardiovascular dysregulation (Carney et al., 2005), saccadic eye movements (Steiger and Kimura, 2010) and changes in REM sleep parameters (Hasler et al., 2004).

3.1.2. Markers for suicidality

Clinical studies of families, twins and adoptees show there is a genetic susceptibility to engaging in suicidal behaviour (Mann, 2003; Roy et al., 2008) but it is thought that this susceptibility is only likely to manifest itself in an individual at times of severe stress or suffering from a major psychiatric disorder (Roy et al., 2008). It is hypothesised that changes to the limbic system can alter a person's risk of engaging in suicidal behaviours as it plays a major role in influencing their depression and stress response (Sequeira et al., 2007). Decreases in the size of the basal ganglia have also been reported in individuals who have had recently attempted suicide (Vang et al., 2010).

Differing levels of serotonin and dopamine have been reported in the basal ganglia between suicidal individuals and controls (Ryding et al., 2006). Research has shown individuals who die by suicide have lower levels of serotonin (Mann, 2003; Nock et al., 2008; Roy et al., 2008), but as already mentioned, low serotonin levels are non-specific to suicidal behaviour. Recently it has reported that lower levels of the serotonin metabolite – 5-HIAA – in the cerebrospinal fluid is associated with short-term suicidal risk (Costanza et al., 2014; Jokinen et al., 2008). BDNF abnormalities have also been associated with suicidal behaviours (Costanza et al., 2014; Dwivedi et al., 2003). Dwivedi et al. (2003) found, in a post mortem study, statistically significant reduced levels of BDNF in both the prefrontal cortex and hippocampus in 27 suicide subjects, regardless of psychiatric diagnosis, compared with that of a matched control group. Le-Niculescu et al. (2013) found blood molecule abnormalities associated with the gene SAT1 in the blood of nine different men who had killed themselves. They also found high SAT1 levels in 49 male patients hospitalised for suicidal behaviours. For a more in-depth discussion of biomarkers of suicide, the reader is referred to Costanza et al. (2014).

3.2. Behavioural markers

Although a large part of human communication is non-verbal and large parts of this channel of communication are beyond conscious control, current diagnostic methods for depression and suicidality do not utilise this extra information (Balsters et al., 2012).

3.2.1. Markers for depression

Including many speech based markers discussed in depth in Section 6, there are a variety of behavioural signals that can be used to distinguish between depressed and healthy individuals (Table 4). Recently preliminary

Table 4
Non Speech based behavioural signals associated with clinical unipolar depression, where ↓ indicates a reduction in the behaviour whilst ↑ indicates a increase the behaviour.

Behavioural Signal	Effect	Reference
Social Interaction	↓	Bos et al. (2002) and Hall et al. (1995)
Clinical Interaction	↓	Parker et al. (1990)
Gross Motor Activity	↓	Balsters et al. (2012), Parker et al. (1990), and Sobin and Sackeim (1997)
Slumped Posture	↑	Parker et al. (1990) and Segrin (2000)
Gesturing	↓	Balsters et al. (2012) and Segrin (2000)
Self-Touching	↑	Scherer et al. (2013c), Segrin (2000) and Sobin and Sackeim (1997)
Head-Movements (Variability)	↓	Girard et al. (2013) and Scherer et al. (2013d)
Facial Activity		
Mobility	↓	Parker et al. (1990) and Sobin and Sackeim (1997)
Expressivity	↓	Ellgring and Scherer (1996), Gaebel and Wölwer (2004), Girard et al. (2013), Maddage et al. (2009), Schelde (1998), and Segrin (2000)
Smiling	↓	Balsters et al. (2012), Schelde (1998), Scherer et al. (2013c), Segrin (2000), and Sobin and Sackeim (1997)
Eye Movements		
Eyebrow movements	↓	Balsters et al. (2012), Schelde (1998), and Segrin (2000)
Horizontal pursuit	↓	Abel et al. (1991) and Lipton et al. (1980)
Saccades	↓	Abel et al. (1991) and Crawford et al. (1995)
Visual fixation	↑	Sweeney et al. (1998)

Table 5
Warning signs of a suicidal crisis, reproduced from Rudd et al. (2006).

Warning sign	Risk
Direct verbal suicide threat (Direct threats)	Imminent
Direct suicidal actions (Seeking means to attempt suicide)	Imminent
Talking/writing about suicide (Indirect threats)	Imminent
Increased hopelessness	Heightened
Rage, anger, seeking revenge	Heightened
Self-destructive behaviour	Heightened
Feeling trapped-like there's no way out	Heightened
Increased substance abuse	Heightened
Withdrawing from friends, family, or society	Heightened
Anxiety, agitation, unable to sleep, or sleeping all the time	Heightened
Dramatic changes in mood or behaviour	Heightened
No reason for living; no sense of purpose in life	Heightened

work has been carried out into the automatic analysis of depression using non speech based behavioural signals such as facial activity (Cohn et al., 2009; Joshi et al., 2013; Scherer et al., 2013c) hand movements and leg fidgeting (Scherer et al., 2013c), and eye movement (Alghowinem et al., 2013c).

3.2.2. Markers for suicidality

Due to the low rate of completed suicide in the general population the identification of behavioural signals that have both sensitivity and specificity when predicting suicidality is a difficult task (Mandrusiak et al., 2006). In 2006 the American Association of Suicidology identified 3 warning signs – behavioural markers – that signal an imminent suicidal crisis and a further 9 warning signs associated with a heightened risk of engaging in suicidal behaviours, Table 5.

4. Speech as an objective marker

Clinically, prosodic abnormalities associated with an individual's mental state are well documented. In 1921 Emil Kraepalin, regarded as the founder of modern psychiatry, described depressive voices as "*patients speak in a low voice, slowly, hesitatingly, monotonously, sometimes stuttering, whispering, try several times before they bring out a word, become mute in the middle of a sentence*" (Kraepelin, 1921). A pre-suicidal voice is described as sounding hollow and toneless, whilst monotony, dullness and reduced verbal output have long been associated with a depressed voice.

As a result of a complexity of speech production (Section 4.1) speech is a sensitive output system; slight physiological and cognitive changes potentially can produce noticeable acoustic changes (Scherer, 1986). We hypothesise that depression and suicidality produce both cognitive (Section 4.2) and physiological changes (Section 4.3) that influence the process of speech production, affecting the acoustic quality of the speech produced in a way that is measurable and possible to be objectively assessed.

4.1. Speech production

The process of speech production involves simultaneous cognitive planning and complex motoric muscular actions (Fig. 1). Cognitively, speech production includes the formation of the message a speaker wishes to communicate followed by the setup of the phonetic and prosodic information associated with this intended message. This information is then stored briefly in working memory, the short-term memory necessary to perform information manipulation and complex cognitive tasks including analysing visual information and language comprehension (Baddeley, 2003). Elements in working memory are transformed to phonetic and prosodic representations and the speaker then executes a series of neuromuscular commands to initiate the motoric actions required to produce speech. These motoric actions are viewed as a source-filter operation. The source is air forced from the lungs through the vocal folds and the space between them known as the glottis. The vocal tract acts as a filter, amplifying and attenuating different frequencies, imposing spectral shaping on the glottal waveform. Positioning of the articulators shapes the vocal tract to produce the different phonemes. A speaker continuously monitors and controls their own speech via two feedback loops; the proprioceptive loop which monitors the movement and shape of the muscles actions and the auditory loop where the speaker uses their own speech as feedback (Postma, 2000).

Speech production uses more motor fibres than any other human mechanical activity (Kent, 2000). The major muscle groups include the respiratory, laryngeal, and articulatory muscles. Respiratory muscles include the diaphragm and intercostal muscle groups whilst the larynx is made up of nine cartilages controlled by 5 muscle groups. The articulators include the mandible, lip, tongue, velum, jaw and pharyngeal constrictors all which play a part in determining the shape of the vocal tract. It is possible to produce 20–30 phonetic segments per second, which makes speech production the fastest discrete human motor performance (Kent, 2000).

4.2. Cognitive effects on speech production

Cognitive impairments associated with depression have been shown to affect an afflicted individual's working memory (Murphy et al., 1999). A key component in working memory is the phonological loop, which helps control the articulatory system and store speech-based information, typically for a few seconds. Interestingly both serotonin and BDNF have been linked to working memory performance (Brooks et al., 2014; Enge et al., 2011). Work by Christopher and MacDonald shows that depression affects the phonological loop causing phonation and articulation errors (Christopher and MacDonald, 2005). Working memory impairments have also been reported with suicidal behaviour (Raust et al., 2007).

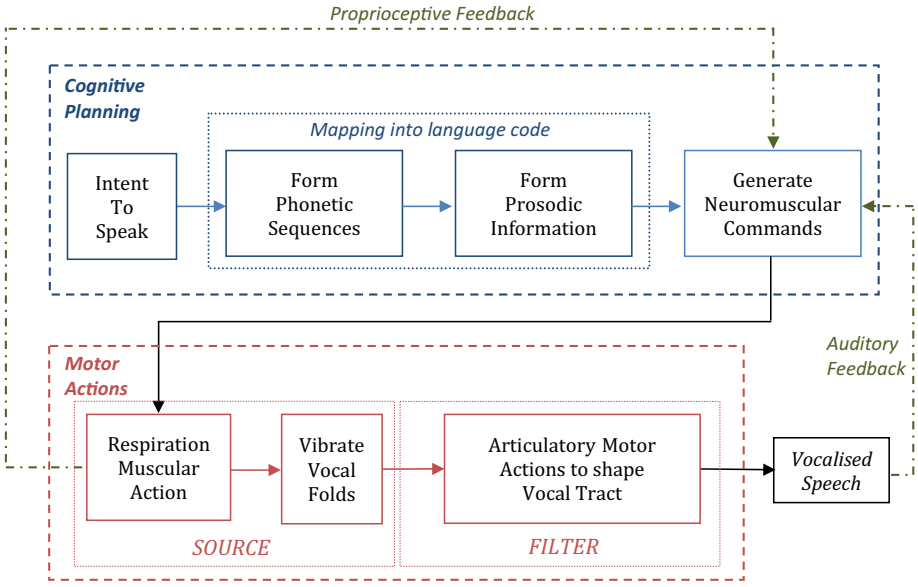

Fig. 1. Schematic diagram of speech production, adapted from Krajewski et al. (2012) and O'Shaughnessy (1999).

A reduction in cognitive ability and subsequent working memory impairments affects speech planning (Levelt et al., 1999), impairs the neuromuscular motor coordination processes (Krajewski et al., 2012) and alters the proprioceptive feedback loop affecting the feedback of articulator positions (Krajewski et al., 2012). Work in the literature confirms this; significant correlations between depression severity with pause related measures, showing that depressed individuals have difficulty choosing words (Alpert et al., 2001; Mundt et al., 2012). Williamson et al. (2013) use a feature space specifically designed to capture information relating to a reduction in articulatory coordination to achieve good prediction accuracy against the BDI scale. These results, and similar, are discussed in greater detail in Section 6.

Changes to the somatic nervous system (SNS) and automatic nervous system (ANS) cause disturbances in muscle tension (Scherer, 1986) and respiratory rate (Kreibig, 2010). The GABA neurotransmitter has also been linked to changes in muscle tonality (Croarkin et al., 2011). Changes in muscle tension and control will affect the prosody and quality of the speech produced; disturbances in laryngeal muscle tension will influence vocal fold behaviour whilst changes to respiratory muscles will affect subglottal pressure. Both prosodic and source features have been shown to be affected by a speaker's level of depression (Moore et al., 2008; Mundt et al., 2007; Quatieri and Malyska, 2012; Scherer et al., 2013c; Trevino et al., 2011) and suicidality (Ozdas et al., 2004a; Scherer et al., 2013a) and will be discussed further in Section 6.

4.3. Changes in affect state and speech production

Changes in a speaker's affective state, common in both depression (Davidson et al., 2002; Goeleven et al., 2006) and suicidal behaviour (Hendin et al., 2007; Rudd et al., 2006) potentially alter the mechanisms involved in speech production, the phonation and articulation muscular systems, via changes to the somatic and autonomic nervous systems (Ozdas et al., 2004a; Scherer, 1986).

Changes to the somatic nervous system (SNS) and automatic nervous system (ANS) cause disturbances in muscle tension (Scherer, 1986) and respiratory rate (Kreibig, 2010). The GABA neurotransmitter has also been linked to changes in muscle tonality (Croarkin et al., 2011). Changes in muscle tension and control should affect the prosody and quality of the speech produced; disturbances in muscle tension will influence vocal fold behaviour whilst changes to respiratory muscles should affect subglottal pressure. Both prosodic and source features have been shown to be affected by a speaker's level of depression (Moore et al., 2008; Mundt et al., 2007; Quatieri and Malyska, 2012; Scherer et al., 2013c; Trevino et al., 2011) and suicidality (Ozdas et al., 2004a; Scherer et al., 2013a) and will be discussed further in Section 6.

Changes in muscle tension will also change vocal tract dynamics and constrain articulatory movement. Articulatory movements will be restricted due to both the common mechanical link they share with the already tightened laryngeal muscles, the hypolaryngeal complex, as well as the SNS affecting jaw and facial muscle tension (Roy et al., 2009). These changes will produce speech errors, a decrease in speech rate and an increase in hesitations (Cannizzaro et al., 2004; Ellgring and Scherer, 1996). Vocal tract properties are affected by both an increase in muscle tension and changes in salivation and mucus secretion controlled by the ANS response (Scherer, 1986). Changes in filter dynamics are well reported in the literature for both depression (Cummins et al., 2013a,b; Flint et al., 1993; Helfer et al., 2013; Quatieri and Malyska, 2012; Tolkmitt et al., 1982; Williamson et al., 2013) and suicidality (France et al., 2000; Keskinpala et al., 2007;

Landau et al., 2007; Ozdas et al., 2004a; Yingthawornsuk and Shiavi, 2008; Yingthawornsuk et al., 2007, 2006) and will be discussed further in Section 6.

5. Databases

An overview of 16 depressed and suicidal databases is provided in Table 6. This list comprises the corpora which are still currently used in research.[1] For each corpus additional information is provided: number of subjects, collection paradigm, clinical scores, rough content outline and a complete list of references. All Corpora are English language unless stated.

6. Prosodic and acoustic features

This section reviews how commonly used prosodic and acoustic speech characteristics are affected by depression and suicidality. Where possible, attempts have been made to include physical quantities of relevant features in regards to class (Low or Control versus High level of depression or suicide risk) mean and standard deviation as a table at the end of each section. Speech parameterisation or feature extraction is the transformation of the raw speech signal into a more abstract representation of the signal with fewer redundancies (Kinnunen and Li, 2009). Features commonly used in the literature, as well as more recently introduced features, are extracted from a speech sample on a short-term time scale using overlapping frames of 10–40 ms in length, but can also include extractions on much longer time scales. Typically temporal information and long-term information is captured either statically through the use of utterance level statistics/functionals or dynamically through frame-based delta (Δ) and delta–delta ($\Delta\Delta$) coefficients, reflecting differences between the adjacent vectors' feature coefficients (Schuller et al., 2013, 2011). The foundation for feature extraction methodologies are not covered in this section; for this the reader is referred to either Kinnunen and Li (2009) or Quatieri (2001).

Before starting the feature review it is worthwhile considering the properties of an *ideal* speech feature(s) for detecting either depression or suicidality. The authors consider the ideal feature(s) should be able to: (i) capture a commonly occurring and easy to measure vocal effect; (ii) exhibit large between-class variability, small within class variability; (iii) be individualised, i.e. be able to predict within-subject changes of mental state; (iv) capture effects specific to an increased mental state, i.e. not be a marker of a related symptom such as increased anxiety or fatigue (unless specifically designed as one, see Section 8.2.2); (v) be able to predict onset of either condition; and (vi) stable over time. Further, if this feature is to be used in an automated classification or prediction system (Section 7), it should also be robust to different recording environments and phonetic content (Section 8.2.2) and have low dimensionality to help avoid the *curse of dimensionality*.[2] Challenges in finding an ideal speech-based marker of either condition are discussed in Section 8 of this review.

6.1. Prosodic features

Prosodic features represent the long-time (phoneme level) variations in perceived rhythm, stress, and intonation of speech.[3] Popular examples include the speaking rate, the pitch (auditory perception of tone) and loudness, and energy dynamics. In practice the fundamental frequency (F_0, rate of vocal fold vibration) and energy are the most widely used prosodic features as they relate to the perceptual characteristics of pitch[4] and loudness.

6.1.1. Depressed speech

Early paralinguistic investigations into depressed speech found that patients consistently demonstrated prosodic speech abnormalities such as reduced pitch, reduced pitch range, slower speaking rate and articulation errors. The first investigation undertaken to find specific correlates between patterns of speech and depression focused on these results. Darby and Hollien (1977) found listeners could perceive a change in the pitch, loudness, speaking rate and articulation of depressed patients before and after treatment. Hollien (1980) suggests depressed patients use different speaking patterns and highlights 5 potential characteristics: reduced speaking intensity, reduced pitch range, slower speech, reduced intonation and a lack of linguistic stress. Nilsonne and Sundberg (1985) ran a series of listening tests using signals composed of F_0 contours to see if evaluators, none of whom had a psychology background, could identify if a specific contour belonging to a speaker who was depressed or had successfully undergone treatment reporting an accuracy of 80%. The F_0 contour was chosen for these listening experiments as it contains information from a wide range of prosodic information including F_0 variability as well as speech rate and pause time.

Given the dull, monotonous and "lifeless" descriptors of speech affected by depression (Section 4), it is not surprising that correlations between both a reduced F_0 range and a reduced F_0 average with increasing levels of depression severity are reported throughout the literature (Breznitz, 1992; Darby et al., 1984; Hönig et al., 2014;

[1] We define this to be any databases which have had results published on them in last 10 years.

[2] A machine learning phenomenon where features with large dimensionality degrade the performance of a classifier as their dimensionality is considerably larger than the number of training samples used to train the classifier. Overfitting is a potential reason for the discrepancy between the development and test sets baselines in the 2013 Audio/Visual Emotion Challenge baseline paper (Valstar et al., 2013), Section 7.1.3.

[3] Whilst prosodic structures are influenced by lexical and semantic information in speech this paper does not focus on using lexical or semantic information as a maker of depression.

[4] F_0 correlates well with pitch, a change in the rate of vocal fold vibration leads to a change in perceived pitch (O'Shaughnessy, 1999).

Table 6
Summary of depressed and suicidal speech databases which have had results published on them in the last 10 years. Abbreviations: *DPRD* – Depressed, *SCDL* – Suicidal, *NTRL* – Neutral, not depressed or suicidal, *M* – Number of males, *F* – Number of Females *DSM* – Diagnostic and Statistical Manual of Mental Disorders, *HAMD* – Hamilton Rating Scale for Depression, *BDI* – Beck Depression Inventory, *QIDS* – Quick Inventory of Depressive Symptomology, *PHQ-9* – Patient Health Questionnaire, *C-SSRS* – Columbia Suicide Severity Rating Scale, *SIQ-Jr version* – Suicidal Ideation Questionnaire – Junior. *Note:* where DSM is present as a clinical score all depressed patients in corpus meet criteria for Major Depressive Disorder.

1st Published (Name)	Subjects	Clinical scores	Vocal exercises	Read speech	Free response or interview	Free speech	Additional notes	Other references
France et al. (2000) Vanderbilt II Study	115: 59 DPRD (21M, 38F) 22 SCDL (all M) 34 NTRL (24M, 10F)	DSM-IV BDI (DPRS = BDI > 20)			√	√	*Recorded therapy sessions or suicide notes* Mean file length: *2 min 30 s* Age range: 25–65 Medications Present: *Imipramine-hydrochloride*	*Similar corpus used in:* Ozdas et al. (2004a,b, 2000) and Hashim et al. (2012)
Moore et al. (2004)	33: 15 DPRD (6M, 9F) 18 NTRL (9M, 9F)	DSM-IV		√			Utterances per speaker: 65 Mean file length: *3 min* Age range: 19-57	Moore et al. (2008) *Similar corpus used in:* Moore et al. (2003)
Yingthawornsuk et al. (2006)	32(all M): 10 SCDL 13 DPRD 9 Remitted Patients	BDI (DPRD = BDI > 20)		√	√		Mean file length: *Free Response – 8 min* *Read Speech – 2 min* Age range: 25–65	*Similar corpus used in:* Keskinpala et al. (2007), Landau et al. (2007), Yingthawornsuk et al. (2007), and Hashim et al. (2012)
Mundt et al. (2007)	35: DPRD (15M, 20F)	HAMD Mean: 14.9 ± 6.3 Range: 3–27 QIDS Mean: 12.4 ± 6.1 Range: 0–26	√	√	√		Mean age: 41.8 Medications: *Range present*	Sturim et al. (2011), Trevino et al. (2011), Quatieri and Malyska (2012), Cummins et al. (2013a,b), and Helfer et al. (2013)
Cohn et al. (2009)	57: DPRD (24M, 34F)	DSM-IV HAMD (HAMD ⩾ 5)			√		Min. vocalisation per speaker: *100 s* Mean age: 39.7 Age range: 19–65 Medications: *SSRI's present*	*Extended version:* Yang et al. (2012)
Low et al. (2009)	139: 68 DPRD (49F, 19M) 71 NTRL (71M, 44F)	N/A			√	√	Recordings per subject: 3 Mean file length: *20 min* Age range: 12–19	Memon et al. (2009), Low et al. (2011, 2010), and Ooi et al. (2013, 2012)
Alghowinem et al. (2012)	80: 40 DPRD 40 NTRL	DSM-IV		√	√		Mean file length: *40 min*	Alghowinem et al. (2013a,b), and Cummins et al. (2013b) *Subset published in:* Cummins et al. (2011)
Mundt et al. (2012)	165: All DPRD (61M, 104F)	DSM-IV HAMD QIDS	√	√	√		Age range: 21–75 Mean age: 37.8 Medications: *Sertraline*	None
Scherer et al. (2013a)	60: 30 SCDL 30 NTRL	C-SSRS SIQ-Jr version			√		Age range: 13–17	None
Scherer et al. (2013c) *Distress Assessment Interview Corpus*	110: 29%: DPRD 32%: PSTD 62%: Anxiety	PHQ-9 (DPRD = PHQ-9 > 10)			√		Data per participant: *30–60 min* Age range: 18–65	Scherer et al. (2013b,d)

Valstar et al. (2013) *Audio-Visual Depressive Language Corpus (AViD Corpus)*	292: *AVEC 2013*: 150 files each containing a range of mix of vocal exercises, free and read speech tasks *AVEC 2014*: 150 files each containing a read speech passage (Die Sonne und der Wind) and an answer to a free response question. *Note* AVEC 2014 is a shortened (file length) version of AVEC 2013. 5 files were replaced in 2014 due to unsuitable data	DBI *Mean AVEC 2013* *Training Set: 15.1 ± 12.3* *Development Set: 14.8 ± 11.8* *Mean AVEC 2014* *Training Set: 15.0 ± 12.3* *Development Set: 15.6 ± 12.0*	√	√	√	*For AViD-Corpus:* German Language Mean file length: *25 min* Age range: 18–63 Mean age: *31.5 ± 12.3*	*AVEC 2013 Papers:* Cummins et al. (2014a,b, 2013c), Kaya and Salah (2014), Kaya et al. (2014b), and Williamson et al. (2013) *AVEC 2014 Papers:* Valstar et al. (2014), Gupta et al. (2014), Kaya et al. (2014a), Mitra et al. (2014), Pérez et al. (2014), Senoussaoui et al. (2014), Sidorov and Minker (2014), and Williamson et al. (2014). *Similar corpus used in:* (Hönig et al., 2014) – 1122 recordings taken from 219 speakers in AVDL Corpus

Kuny and Stassen, 1993; Mundt et al., 2007; Nilsonne, 1988; Nilsonne et al., 1987; Stassen et al., 1998; Tolkmitt et al., 1982), as seen in Table 7. However a number of papers report no significant correlation between F_0 variables and depression, (Alpert et al., 2001; Cannizzaro et al., 2004; Mundt et al., 2012; Quatieri and Malyska, 2012; Teasdale et al., 1980; Yang et al., 2012) also seen in Table 7. It is possible these conflicting results are due to the heterogeneous nature of depression symptoms, the fact that F_0 is both a marker of the physical state of the vocal folds and a marker of a speaker's affective state, gender dependence of F_0 and a lack of standardisation in F_0 extraction techniques.

As the larynx is an intricate neuromuscular system, small disturbances in muscle tension due to the effects of *psychomotor retardation* (PMR, the slowing of thought and reduction of physical movements) and/or changes in the speaker's affective state (Section 4.3) can be used to explain the observed reduction in F_0 variability. Increases in monotony could be a result of PMR reducing laryngeal control and dynamics (Horwitz et al., 2013; Quatieri and Malyska, 2012). Evidence for a lack of laryngeal control is provided by reports in the literature of increases in aspiration with depressed speech (Section 6.2.1). Another potential cause could be increases in vocal tract tension tightening the vocal folds resulting in less variable, more monotonous sounding speech (Cannizzaro et al., 2004; Ellgring and Scherer, 1996; Nilsonne et al., 1987; Sobin and Sackeim, 1997; Tolkmitt et al., 1982).

Whilst decreases in F_0 variation can potentially be explained by increase in muscle tension; this theory does not explain reported reduced average F_0; as vocal fold muscle tension increases F_0 should also increase. F_0 is more than just a marker for vocal fold behaviour; it is a major paralinguistic marker, carrying information relating to the expressiveness of speech and as a result is affected by many different speaker states and traits (Nilsonne and Sundberg, 1985). Individuals with depression can differ, in regards to personality traits, in many ways from those without depression (Kotov et al., 2010). Work in the literature shows F_0 is affected by changes to a person's underlying mood (Ellgring and Scherer, 1996), level of agitation and anxiety (Alpert et al., 2001; Tolkmitt et al., 1982) and personality traits (Yang et al., 2012) associated with their depressive status. Therefore, F_0 effects in depression elicited from comparisons with non-depressed individuals (Breznitz, 1992; Darby et al., 1984; Kuny and Stassen, 1993; Nilsonne, 1988) potentially lack specificity for depression. Further, studies that have investigated F_0 change in depression severity over time have reported very small effect sizes that depended on large numbers of participants (Mundt et al., 2012) – small effect sizes are unlikely to have much utility for classification tasks.

As with F_0, there are conflicting results on the effect of depression on energy parameters (Table 7). Darby et al. (1984) report that depressed patients before treatment had reduced variation in loudness due to lack of speaking

Table 7
Examples, taken from the literature, of prosodic measures for low (control) or high levels of speaker depression. *Abbreviations:* N.S. – Not Significant, n – number of samples in class.

Feature	Reference	Low (Control)	High	Significance (Test)
F_0 range[a] (Hz)	Nilsonne (1987)	21 ± 2 ($n = 16$)	15 ± 2 ($n = 16$)	$p \leqslant 0.001$ (t-test)
F_0 range[b] (Hz)	Breznitz (1992)	38.3 ± 11.3 ($n = 11$)	15.8 ± 18.2 ($n = 11$)	$p \leqslant 0.004$ (t-test)
F_0 mean (Hz)	Alpert et al. (2001)	150.6 ± 31.4 ($n = 19$)	142.0 ± 27.2 ($n = 22$)	N.S. (t-test)
F_0 mean (Hz)	Mundt et al. (2012)	153.3 ± 35.7[c] ($n = 51$)	151.8 ± 36.5[d] ($n = 54$)	N.S. (t-test)
F_0 variation (Hz)	Yang et al. (2012)	0.23 ± 0.1[e] ($n = 16$)	0.20 ± 0.1 ($n = 10$)	N.S. (t-test)
Energy per second[f] (mV^2)	Kuny and Stassen (1993)	11.0 ± 4.8 ($n = 30$)	9.9 ± 3.7 ($n = 30$)	$p \leqslant 0.01$ (Wilcoxon)
Loudness[g] (dB)	Alpert et al. (2001)	14.2 ± 7.33 ($n = 19$)	18.1 ± 6.37 ($n = 22$)	N.S. (t-test)
Mean pause duration (s)	Alpert et al. (2001)	0.68 ± 0.136 ($n = 19$)	0.70 ± 0.162 ($n = 22$)	$p \leqslant 0.05$ (t-test)
Total pause time (s)	Mundt et al. (2012)	36.4 ± 19.4[c] ($n = 51$)	51.9 ± 31.5[d] ($n = 54$)	$p \leqslant 0.001$ (t-test)
Pause variability (s)	Mundt et al. (2012)	0.51 ± 0.15[c] ($n = 51$)	0.69 ± 0.25[d] ($n = 54$)	$p \leqslant 0.001$ (t-test)

[a] Both genders used when calculating F_0 range.
[b] Females used when calculating F_0 range.
[c] Patients who responded to treatment during study.
[d] Patients who did not respond to treatment during study.
[e] Clinician used as low group.
[f] Measure of amount of energy associated with a syllable (Kuny and Stassen, 1993).
[g] F_0 Amplitude.

effort which significantly improved after treatment. However, Kuny and Stassen (1993) found only mild correlation in improvements in mean loudness and variation of loudness with patient recovery, whilst Alpert et al. (2001) found that depressed patients spoke louder than controls (22 Depressed subjects, 19 Controls), however not at a significant level. Stassen et al. (1991) report mixed results: depressed patients either demonstrated a lack of energy dynamics which improved after treatment or were overly loud before treatment commenced and decreased to a normal speaking level after treatment. Stassen et al. (1991) argued this result was due to the homogeneous nature of depression. Recently Quatieri and Malyska (2012) reported mildly significant negative correlations between energy variability and depression and significant positive correlation between energy velocity and depression. The authors argue the energy velocity results are an indication of the improvement of motor coordination at lower levels of depression.

The most promising prosodic features for recognising depression are related to speech rate. As already stated, many initial studies report depressed individuals speak at a slower rate than controls (Darby and Hollien, 1977; Godfrey and Knight, 1984; Greden and Carroll, 1980; Greden et al., 1981; Hardy et al., 1984; Hollien, 1980; Szabadi et al., 1976; Teasdale et al., 1980). Among more recent studies, Stassen et al. (1998) found, for 60% of patients in their study, that speech pause duration was significantly correlated with their HAMD score. Alpert et al. (2001) also report significant differences in speech pause duration between the free speech of a control and depression group. Investigations undertaken in Cannizzaro et al. (2004) found, albeit on a database of 7 speakers, that reduced speaking rate had a significant correlation with HAMD measures.

Two recent studies led by Mundt confirmed these findings on larger databases. In the first study, (Mundt et al., 2007), it was observed that patients who responded to treatment (relative reduction of 50% in their HAMD score) during a 6 week study into depression severity, paused less and spoke faster at the end of the 6 week period. In a follow-up study, Mundt et al. (2012) found 6 prosodic timing measures were significantly correlated with depression severity; total speech time, total pause time, percentage pause time, speech pause ratio and speaking rate. Alghowinem et al. (2012) analysed the differences between speech rate, articulation rate and pause rate between samples of depressed and control free response speech. In the study only average syllable duration – total speech duration divided by number of syllables – was found to be significantly lower for the depressed group. Hönig et al. (2014) also analysed average syllable duration and report a positive correlation with increasing levels of speaker depression. Both Alghowinem et al. (2012) and Hönig et al. (2014) results are consistent with an overall increase in speech rate with depression.

Recent results reported on Mundt's 35-speaker database, Trevino et al. (2011), indicate that changes in speech rate are potentially stronger when extracted at the phoneme level of speech production. Trevino et al. (2011) found that dissecting average measures of speech rate into phone-specific characteristics and, in particular, combining phone-duration measures (linear combination of the average phone durations that are highly correlated with depression) uncovers stronger relationships between speech rate and depression severity than global measures previously reported for a speech rate marker. The authors also found consistency in the correlations when grouping individual phonemes together by manner of articulation, i.e. vowels or fricatives, and strong correlation between

the phone-based measures with specific HAMD sub-symptoms, PMR and mood in particular (Section 8.2.2). The authors conclude that their results show the potential of phoneme-based indicators of speech rate as a biomarker of depression.

Whilst it is arguably one of the strongest speech features associated with depression severity, it is unclear in the literature whether a reduction in speech rate associated with depression is a potential measure of motor retardation or of cognitive impairment. There are two ways to slow speech rate (Cannizzaro et al., 2004); the first is a decrease in the rate of speech sound production which is reflective of a motor impairment. The second method is to insert longer or more frequent pauses into an utterance; this would suggest a cognitive impairment where an individual has difficulty choosing their words. Cannizzaro et al. (2004) argue that their findings, a decrease in speech rate but no significant increase in speech pause measures, show that motor retardation causes a decrease in motor speed and agility slowing speech. Alpert et al. (2001) found that whilst speech pause measures significantly increased in a depressed group of patients when compared to a control group, there was no decrease in speech intelligibility. They argue that this result shows speech slowing in depression is a marker of lowered cognitive functioning related to speaker motivation. It should be noted that the results of both papers are limited by the size and nature of their databases.

Whilst many clinicians subjectively use prosodic features during diagnosis it is possible that due to natural variations in individual speaking styles and the wide clinical profile of depression, a single dimensional prosodic feature does not contain sufficient discriminatory information for use as a clinical marker of depression. Nilsonne et al. (1987) speculate that the natural F_0 variation present in non-depressed speech encompasses the speech behaviour of depressive speech and state that F_0 variation would be a more useful measure for tracking within-patient changes.[5] Stassen et al. (1991) found no significant differences between basic, single dimensional, prosodic features and clinically defined subgroups within a group of 20 depressed patients. The authors speculate that both the complexity of speech and depression mean that a multivariate approach is needed to diagnose depression using speech parameters. Moore et al. (2008) speculate that F_0 is a high level view of vocal fold dynamics not subtle enough to capture vocal fold tension. Their results show that glottal parameters contain more useful discriminatory information than prosodic features (Section 6.2.1).

6.1.2. Suicidal speech

The first investigation into the paralinguistic effects of suicidality were undertaken by Dr. S. Silverman: after listening to recorded psychiatry sessions associated with suicidal patients and recorded suicide notes, he noted that as an individual becomes pre-suicidal, their speech undergoes discernible changes in its quality (Silverman, 1992). Silverman identified four possible acoustic effects; hollow toneless sounds, the loss in intensity over an utterance, unusual vocal patterns such as monotonous, mechanical and repetitious phrasing, and unexpected emphasis, unrelated to normal speech syntax. Silverman hypothesised that being in a pre-suicidal mental state causes changes to speech production mechanisms altering the acoustic properties of speech in measurable ways (Silverman and Silverman, 2006). Silverman investigated the speech of individuals suffering from depression who subsequently attempted suicide; which represent a subset of all suicide attempts. Therefore it cannot be presumed that the identified features will be transferable to a larger database representing a wider range of suicidal behaviours.

Outside of the initial investigations done by Silverman there has been very little work done in investigation of prosodic properties of suicidal speech. Rhythm based features derived from temporal patterns of either voiced, unvoiced and silent speech have been used to classify patients as either suicidal or depressed with accuracies greater that 70% (Hashim et al., 2012). To the best of the authors' knowledge no-one has found significant correlations between prosodic features and suicidality.

6.2. Source features

Source features capture information relating the source of voice production, the air flow from the lungs through glottis. They can either parameterize this flow directly via glottal features, or vocal fold movements via voice quality features. If being in a clinically depressed or suicidal mental state affects laryngeal control (Sections 4.2 and 4.3), source features should therefore capture information relating to both conditions. However, that has not been universally found across the literature. Nevertheless, the discrepancies between findings could be due, in part, to the difficulty in extracting accurate source information from a speech signal.

Many source features are estimated from time-length measurements of the extracted glottal flow signal (Airas, 2008; Doval et al., 2006); however these time instants are difficult to automatically extract due to non-uniform vocal fold behaviour and formant ripple and noise remaining after the inverse filtering required in their extraction, needed to remove the effects of a constantly changing vocal tract (Walker and Murphy, 2007). For more information on modelling glottal flow and extraction techniques, the reader is referred to Doval et al. (2006), Drugman et al. (2010) and Walker and Murphy (2007).

Source features are a major marker of *Voice Quality*, the auditory perception of changes in vocal fold vibration and vocal tract shape, outside of pitch, loudness, and phonetic category. Often voice quality features measure irregular

[5] This natural variation in speaking style means there are problems relating to normalization of prosodic features i.e. calculating a baseline to work out potential improvements from. This often not addressed in the literature.

phonation capturing information relating to phonation types or laryngeal qualities, such as breathiness, creakiness or harshness (Gobl and Ní Chasaide, 2003). Voice quality measures commonly seen in both depressed and suicidal speech literature include *jitter*, the small cycle-to-cycle variations in glottal pulse timing during voicing; *shimmer*, the small cycle-to-cycle variations in glottal pulse amplitude in voiced regions; and *harmonic-to-noise ratio* (HNR), a ratio of harmonics to inharmonic (spectral components which are not a whole number multiple of F_0) components. These features have been shown to correlate strongly with depression, PMR and suicidality (Ozdas et al., 2004a; Quatieri and Malyska, 2012) as they are related to vocal fold vibration, which is influenced by vocal fold tension and subglottal pressure (Sundberg et al., 2011). However, due to a lack of standardisation in extraction techniques – factors such as window duration, sampling frequency and F_0 extraction influence jitter and shimmer values - it is very difficult to compare results between studies (McAllister et al., 1996; Orlikoff and Kahane, 1991).

A further confounding factor when using jitter, shimmer and HNR is the type of speech signal they are extracted from, i.e. held vowels or continuous speech (Laver et al., 1992). Due to their periodicity, the use of held vowels allows for simpler extraction of these features but leaves them open to errors due to differing sound pressure levels between and within individuals, potentially rendering unreliable for clinical analysis (Orlikoff and Kahane, 1991). Analysis of continuous speech is more difficult when compared to held vowels, due to the problems associated with finding suitable voiced sections in a given utterance (Laver et al., 1992). The automatic identification of suitable speech segments for glottal source analysis of continuous speech is an active area of speech processing research (Kane et al., 2014, 2013) and important area of future research relating to depressed and suicidal speech analysis.

6.2.1. Depressed speech

Only a small number of papers have studied the effect of depression on source measures, mainly voice quality, with very inconsistent results reported. Flint et al. (1993) found increased spirantization, a measure that reflects aspirated leakage at the vocal folds, in depressed individuals when compared to healthy controls (Table 8). Ozdas et al. (2004a) report significance via the F-test, between their depressed and control classes for jitter (Table 8) but not for spectral slope (Table 8); when using a two sample t-test they report significant differences between the depressed and control class for glottal spectral slope but not for jitter.

Quatieri and Malyska (2012) report that, on the Mundt 35 speaker dataset, aspiration defined as the inverse of HNR, jitter and shimmer were significantly correlated with depression severity (HAMD). Quatieri and Malyska (2012) use these results to hypothesise that the presence of motor retardation in depression reduces laryngeal muscle tension resulting in a more open, turbulent glottis. Results presented in Hönig et al. (2014) show strong negative correlations between shimmer, spectral harmonic and spectral tilt, all indicative of a more breathy phonation in depressed speech, and increasing levels of speaker depression supporting Quatieri and Malyska (2012) hypothesis. Low et al. (2011), using pairwise ANOVA comparison, found that the *Teager Energy Operator* (TEO) Autocorrelation feature was significantly different between depressed and control patients. Teager energy operators capture amplitude and frequency modulations of vocal tract resonances generated by nonlinear airflow in the vocal tract (Maragos et al., 1993) one application being detection of speech under stress (Zhou et al., 2001). Such nonlinear production mechanisms were first observed by Teager and Teager (1990).

Three recent papers authored by Scherer all show promising trends between voice quality features and depression severity (Scherer et al., 2013b,c,d). Using both the *Normalised Amplitude Quotient* (NAQ), a feature derived from the glottal flow derivative, and the *Quasi-Open-Quotient* (QOQ), a feature derived from amplitude measurements of the glottal flow (both extracted fully automatically using the IAIF algorithm (Alku, 1992) to parameterise the glottal flow) these papers all show statistical significance in discerning speakers with moderate to severe depression and speakers without depression (Table 8). These results indicate that depressed voices potentially have a more tense voice quality, matching results presented in Darby et al. (1984), Flint et al. (1993) and France et al. (2000) showing an increase in vocal fold

Table 8
Examples, taken from the literature, of source measures for Low (Control) or High levels of speaker depression. *Abbreviations:* N.S. – Not significant, n – number of samples in class.

Feature	Reference	Low (Control)	High	Significance (Test)
Spirantization (ratio[a])	Flint et al. (1993)	0.32 ± 0.43 ($n = 30$)	0.59 ± 0.56 ($n = 31$)	$p \leqslant 0.02$ (ANOVA)
Jitter (ratio[b])	Ozdas et al. (2004a)	0.0165 ± 0.002 ($n = 10$)	0.187 ± 0.005 ($n = 10$)	N.S. (t-test)
Spectral slope (kHz/dB)	Ozdas et al. (2004a)	-83.3 ± 5.46 ($n = 10$)	-62.4 ± 9.02 ($n = 10$)	$p \leqslant 0.05$ (t-test)
NAQ (ratio[c])	Scherer et al. (2013d)	0.100 ± 0.027 ($n = 25$)	0.067 ± 0.034 ($n = 14$)	$p \leqslant 0.002$ (t-test)
QOQ (ratio[d])	Scherer et al. (2013d)	0.367 ± 0.072 ($n = 25$)	0.292 ± 0.103 ($n = 14$)	$p \leqslant 0.002$ (t-test)

[a] (Present/Absent).
[b] Mean F_0 difference of adjacent glottal cycles/the mean F_0.
[c] Peak glottal flow amplitude/(negative peak amplitude of the glottal flow derivative $\times$ F_0 period).
[d] Quasi-open time of the glottis/quasi-closed time of the glottis. (Note *quasi* as perfect periodicity of vocal folds vibration is never achieved.)

tension associated with depressed individuals. This effect is discussed further in Sections 6.3.1 and 6.4.1.

There has been a small amount of work done analysing the glottal spectrum. Ozdas et al. (2004a) speculate that their results, depressed patients exhibit higher energy in the upper frequency bands of the glottal spectrum, are the result of irregularly shaped glottal pulses caused by increasing laryngeal tension. Excessive vocal fold tension and a lack of coordination of the laryngeal and articulatory muscles would interfere with the vibration of vocal folds causing their action to become more irregular and aperiodic (Ozdas et al., 2004a; Scherer, 1986). Quatieri and Malyska (2012) performed a sub-band decomposition of the glottal spectrum and found uniformly positive high-frequency correlations in the glottal spectrum with depression severity. This result and their voice quality measure correlations support Ozdas et al. (2004a) speculation.

Glottal features have mainly been used for depressive speech classification systems, which are discussed further in Section 7.1.1. In their initial system design, both Moore et al. (2008) and Low et al. (2011) used ANOVA analysis and found that glottal parameters, grouped together either as timing of glottal frequency measures, exhibited significant differences between groupings of depressed and control patients.

6.2.2. Suicidal speech

As well as analysing voice quality for depressive and control speech Ozdas et al. (2004a), using a two sample *t*-test, report significant differences between the glottal spectral slope of their suicidal and depressed classes, and for both jitter (Table 9) and glottal slope for depressed (Table 9) between their suicidal and control classes. The glottal spectral slope results indicate a shift in glottal flow spectrum from lower to higher bands with increased mental state; this result is discussed further in Sections 6.3.2 and 6.4.2.

Statistically significant differences between features, relating to a breathy voice quality, between suicidal adolescent voices and matched control have been reported. Scherer et al. (2013a) analysed the effect size, shift of the mean of one class needed to match the mean of the other, of range of source parameters in the speech of suicidal and non-suicidal adolescents and found NAQ, *Open Quotient* (OQ), and PeakSlope, a voice quality parameter relating to glottal closure instance, all showed a significant effect (Hedges' g value: $g \geqslant 0.4$) between the two classes (Table 9). These results indicate that suicidal adolescent voices have a potential breathy-tense quality to them.

6.3. Formant features

If as hypothesised in Section 4.3, changes in vocal tract properties are affected by both an increase in muscle tension and changes in salivation and mucus secretion relating to changes in a speaker's mental state, then these changes should be captured in detail by formant features, which contain information relating to the acoustic resonances of the vocal tract.

Decreased formant frequencies reported with depressed speech provide evidence for a decrease in articulatory effort with increasing levels of speaker depression (Flint et al., 1993; Mundt et al., 2007). It is possible these effects are due to PMR tightening the vocal tract (Flint et al., 1993; France et al., 2000), or a lack of motor coordination which improves as the effects of PMR decreases (Cummins et al., 2013a; Helfer et al., 2013; Quatieri and Malyska, 2012; Trevino et al., 2011; Williamson et al., 2013) or possibly a result of anti-depressant medication drying out the vocal tract and mouth, directly affecting formant properties and energy distribution (France et al., 2000). Issue relating to effects of medication will be discussed further in Section 8.

6.3.1. Depressed speech

Flint et al. (1993) report significant differences in formant measures, in particular the second formant location, for the phoneme /aɪ/, in depressive patients (all of whom scored 1 or above on the PMR HAMD item) when compared with a matched control group (Table 10), hypothesising that this reduced F_2 location was due to a slowing of the tongue in low-back to high-front moments. They found this abnormality was comparable to results they gained from individuals suffering from Parkinson's disease in which articulatory errors are a result of a depletion in dopamine. They go on to argue that PMR, a result of reduced dopamine, induces these similar articulatory errors

Table 9
Examples, taken from the literature, of source measures for Low (Control) or High levels of speaker suicidality. *Abbreviations:* N.S. – Not Significant, n – number of samples in class.

Feature	Reference	Low (Control)	High	Significance (Test)
Jitter (ratio[a])	Ozdas et al. (2004a)	0.0165 ± 0.002 ($n = 10$)	0.0217 ± 0.005 ($n = 10$)	$p \leqslant 0.05$ (*t*-test)
Spectral slope (kHz/dB)	Ozdas et al. (2004a)	-83.3 ± 5.46 ($n = 10$)	-75.56 ± 8.53 ($n = 10$)	$p \leqslant 0.05$ (*t*-test)
NAQ (ratio[b])	Scherer et al. (2013a)	0.09 ± 0.04 ($n = 8$)	0.12 ± 0.05 ($n = 8$)	$p \leqslant 0.002$ (*t*-test)
OQ (ratio[c])	Scherer et al. (2013a)	0.31 ± 0.13 ($n = 8$)	0.42 ± 0.2 ($n = 8$)	$p \leqslant 0.002$ (*t*-test)
PeakSlope (steepness[d])	Scherer et al. (2013a)	-0.23 ± 0.05 ($n = 8$)	-0.25 ± 0.04 ($n = 8$)	$p \leqslant 0.002$ (*t*-test)

[a] Mean F_0 difference of adjacent glottal cycles/the mean F_0.
[b] Peak glottal flow amplitude/(negative peak amplitude of the glottal flow derivative $\times$ F_0 period).
[c] Open phase of vocal folds/F_0 period.
[d] Regression line fitted to (Wavelet peak amplitudes/Filter Centre frequencies) (Kane and Gobl, 2011).

Table 10
Examples, taken from the literature, of formant measures for Low (Control) or High levels of speaker depression. **Abbreviations: *N.S.* – Not Significant, *n* – number of samples in class**

Feature	Reference	Low (Control)	High	Significance (Test)
F_2 location (Hz)	Flint et al. (1993)	1132.7 ± 264.2 ($n = 30$)	944.5 ± 380.8 ($n = 31$)	$p \leqslant 0.02$ (t-test)
F_1 location (Hz)	Mundt et al. (2012)	546.8 ± 67.1[a] ($n = 51$)	558.2 ± 51.8[b] ($n = 54$)	N.S. (t-test)

[a] Patients who responded to treatment during study.
[b] Patients who did not respond to treatment during study.

through slowness in articulatory muscles and increased muscle tone.

Changes in formant features associated with depression could also be a result of increased muscular tension. The dampening of the vocal tract resonances due to increased tension should have the effect of narrowing formant bandwidth (Scherer, 1986). Increased facial tension and reduced smiling associated with depression, (see Table 4), shortens the vocal tract via retraction of the tongue. This should have the effect of rising the first formant (F_1) and lowering F_2, especially in front vowels (Laukkanen et al., 2006; Scherer, 1986). A higher F_1 and lower F_2 is associated with a throaty voice quality (Laukkanen et al., 2006).

However, this trend of formant behaviour, higher F_1 and lower F_2, has not been consistently observed in depressed speech literature. France et al. found a trend of increased formant frequencies (F_1-F_3) and first formant bandwidth and decreased higher formant bandwidths (France et al., 2000). In the first of the papers led by Mundt, it is reported that F_1 variability was not significantly correlated with depression whilst F_2 variability was mildly correlated (Mundt et al., 2007). However, in the follow-up study, both F_1 and F_2, location (Table 10) and variability were not significantly correlated with depression (Mundt et al., 2012). A potential reason for the discrepancies in formant results is the complex relationship between source and filter dynamics, which means the relationship between articulatory behaviour and formants is not straightforward (Roy et al., 2009); improper glottal closure, Section 6.2.1, can introduce zeroes into all-pole vocal tract models, raising F1 and F2 (Klatt and Klatt, 1990).

Formant based features are very popular in depressive speech classification systems. In their system design, Low et al. (2011) found that the first three formant frequencies and bandwidths, grouped together, had significant differences between their depressed and control patients ($p < 0.05$). Helfer et al. (2013) constructed a two class low/high depression classifier using features derived from formant frequencies, specifically including their dynamics given by velocity and acceleration, and reported accuracies of 70% and 73% using Gaussian Mixture Model and Support Vector Machine respectively. Recently, Williamson et al. (2013) introduced a new promising approach, to predict an individual's level of depression, by extracting information relating to change in coordination across articulators as reflected in formant frequencies. This approach is discussed in further detail in Sections 7.1.3 and 8.2.2. Other papers which used formants in their classification system include (Alghowinem et al., 2013a,b, 2012; Cummins et al., 2011; France et al., 2000; Hönig et al., 2014; Moore et al., 2004), the reader is referred to Section 7.1 for more details.

6.3.2. Suicidal Speech

France et al. (2000) could be characterised by an increase in formant frequencies and a decrease in formant bandwidth. The major difference between the spectral properties of depressed and suicidal speech were significant shifts in power from lower to higher frequency in suicidal patients, although the authors noted the effects of anti-depressants[6] and artefact differences due to the uncontrolled recording conditions of database could have affected results.

6.4. Spectral analysis

Spectral features characterise the speech spectrum; the frequency distribution of the speech signal at a specific time instance information in some high dimensional representation. Commonly used spectral features include *Power Spectral Density* (PSD) and *Mel Frequency Cepstral Features* (MFCCs). As with formants, spectral based features have consistently been observed to change with a speaker's mental state, although there is some disagreement as to the nature of the effect. Some depression papers have reported a relative shift in energy from lower to higher frequency bands (France et al., 2000; Ozdas et al., 2004a), while other depression papers report a reduction in sub-band energy variability (Cummins et al., 2013a; Quatieri and Malyska, 2012). But whilst spectral features are suited to capturing information relating to changes in muscle tension and control as they offer more detailed information on vocal tract behaviour, these features are representative of all the information contained in speech, linguistic and paralinguistic, which can potentially hinder the usefulness in classification systems (Section 8.2).

[6] The effect of anti-depressants was not directly tested. Dataset does not contain control and either depressed or pre-suicidal voice samples from the same individual.

6.4.1. Depressed speech

Tolkmitt et al. (1982) first reported a shift in spectral energy, from below 500 Hz to 500–1000 Hz, with increasing depression severity. This relative shift in energy from lower to higher frequency bands with increasing depression severity was also found in France et al. (2000) and Ozdas et al. (2004a). This effect was not found by Yingthawornsuk et al. (2006), where elevated energy in 0–500 Hz and decreased energy in the 500–1000 Hz and 1000–1500 Hz bands, were found comparing depressed speech to the remitted speech.

This shift in energy is potentially a result of increase in vocal tract and vocal fold tension changing the resonance properties of the vocal tract filter (Scherer, 1986). Speech produced under these conditions is often described as having a throaty, strained or tense quality (Laukkanen et al., 2006; Scherer, 1986). Magnetic Resonance Imaging has shown that throaty voice quality is associated with increased vocal tract tension (Laukkanen et al., 2006). France et al. (2000) and Ozdas et al. (2004a) both speculate that alterations in spectral properties, observed in depressive voices, are a result of increased muscle tone in the vocal tract tension caused by PMR.

MFCCs are one of the most popular spectral features used in speech parameterisation. The signal is filtered by a bank of non-linearly spaced band pass filters (mel-filters) whose frequency response is based on the cochlear of the human auditory system. As the filters span the magnitude spectrum they obtain an estimate of vocal tract characteristics: the low frequency filters capture information about the fundamental frequency and the first coefficient represents the average spectral tilt of the spectrum, for example. MFCCs in combination with *Gaussian Mixture Models* (GMM) is a popular method of speech parameterisation, and these have been shown to be suitable for classifying either low/high levels of depression (Cummins et al., 2013b; Sturim et al., 2011) or the presence/absence of depression (Alghowinem et al., 2012; Cummins et al., 2011), see Section 7.1.

It is standard practise in speaker recognition tasks to concatenate MFCC with time derivatives (delta) features, which reflect rapid temporal information captured in the MFCC extraction. In a recent investigation Cummins et al. (2013b) found significant negative correlation of MFCC's appended with time derivatives, in both the feature and acoustic space variance, with level of depression. Cummins et al. (2013b) argue that these results, showing a decrease in temporal variations with increasing depression severity, fit the dull and monotonous descriptions often associated with depressive speech.

Sub-band energy variability has also been shown to decrease with increasing levels of depression. Quatieri and Malyska (2012) report uniformly negative correlations, although not significant, of energy variance with depression severity. Cummins et al. (2013a) applied Log Mean Subtraction to sub-band energy coefficients, to isolate the spectral variability associated with the speech production mechanism, reporting negative correlations between this measure and depression severity. The negative correlations show that with increasing level of speaker depression there is a decrease in the energy variability associated with speech production mechanism. The results in both Quatieri and Malyska (2012) and Cummins et al. (2013a) are consistent with conclusion drawn in Cannizzaro et al. (2004) who related reduce articulation rate to increased muscle tone. However, it should be noted that the Cannizzaro et al. (2004) results were found on a smaller dataset with inconsistent experimental procedures so their conclusions are somewhat tentatively reached.

A recent group of papers have examined the effect of PMR on energy and spectral based features. As well as reporting increases in phoneme length with increasing depression severity (Section 6.1.1) Trevino et al. (2011) also found pause length increased with various depression subsymptoms including PMR, arguing that this increase is due to increased muscular tension. Quatieri and Malyska (2012) report significant positive correlations between energy velocity and PMR. Cummins et al. (2013a) also report positive correlations between PMR and sub-band spectral variability associated with speech production mechanisms. These results indicate that more effort might be required to produce and sustain PMR-affected speech due to a lack of motor coordination which improves as the effects of PMR decrease (Cummins et al., 2013a; Quatieri and Malyska, 2012; Trevino et al., 2011). Given the heterogeneous nature of depression it is potentially more feasible to create an overall objective marker of depression through a symptom-specific speech based approach (Horwitz et al., 2013; Trevino et al., 2011). The advantages of such an approach are discussed further in Section 8.2.2.

6.4.2. Suicidal speech

For suicidal speech, studies linking energy shift as captured by PSD sub-bands include (Keskinpala et al., 2007; Landau et al., 2007; Yingthawornsuk and Shiavi, 2008; Yingthawornsuk et al., 2007, 2006), although there is some disagreement as to whether or not this is a relative shift in energy from lower to higher frequencies (France et al., 2000; Ozdas et al., 2004a) or vice versa (Yingthawornsuk and Shiavi, 2008; Yingthawornsuk et al., 2007, 2006).

Unfortunately, the databases used in France et al. (2000), Keskinpala et al. (2007), Landau et al. (2007), Ozdas et al. (2004a), Yingthawornsuk and Shiavi (2008) and Yingthawornsuk et al. (2007, 2006) are small and their recording equipment, environment and experimental procedures were not consistent through their speech data, so their conclusions are somewhat tentatively reached. The other issue is that these studies investigated the speech of individuals suffering from depression who subsequently attempted suicide; which represents a subset of all suicide attempts. Therefore it cannot be presumed that the identified features will be transferable to a larger database representing a wider range of suicidal behaviours. However, in the context of the small amount of literature

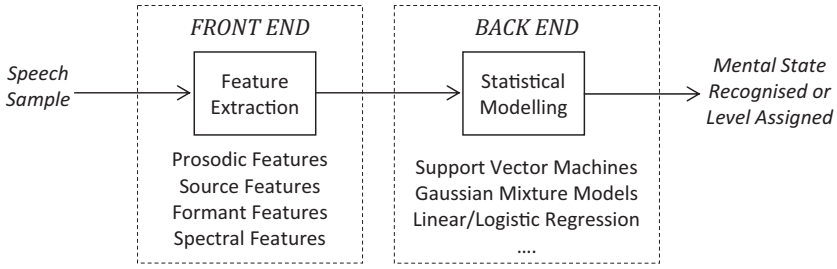

Fig. 2. Overview of a typical classifier or prediction systems.

on the subject and the difficulty inherent in obtaining suitable data, this research on real-world suicidal data is a vital resource which establishes preliminary results on which to base future studies.

7. Classification and score level prediction

Recent work has seen the first steps towards automatic classification of speech as an objective measure for both depression and suicidality. Speech-based classification papers can be divided into one of three different groups of problems - presence, severity or score level prediction – each differing in style of analysis undertaken. *Presence* is a detection problem, i.e. the categorical assignment of unknown (in terms of mental state), voice recording into one of two categories: the presence or absence of either condition. Corpora for this problem must contain two distinct unambiguously annotated classes: speech samples from individuals suffering either condition or samples from a control grouping made up of individuals who were not depressed or suicidal at the time of collection. Performance of classifiers used in this problem is nominally reported in terms of classification accuracy. *Severity* is the categorical assignment of an unknown voice sample into two or more distinct classes. The make-up of these classes is defined by scores of a particular mental state assessment scale, e.g. classification into the five HAMD groupings; *Normal* (0–7), *Mild* (8–13), *Moderate* (14–18), *Severe* (19–22) and *Very Severe* ($\geqslant$23). Again, performance of classifiers used in this problem is nominally reported in terms of classification accuracy. *Score level prediction* is the assignment of an unknown voice sample to continuous-valued mental state assessment scale score. Performance of a score level prediction system is nominally reported as a measure of the differences between values predicted and the values actually observed, such as the *Root Mean Square Error* (RMSE, measure of average error magnitude) and *Mean Absolute Error* (MAE, measure of average error magnitude).

It should be noted here that whilst for the presence-style problems the differences between the distinct classes used are well-defined, the same cannot be said for severity. Depression assessments, in particular, are often multi-component in nature, summing the severity of symptoms observed to give a patient a score which relates to their level of depression, as explained in Section 2.2. The final assessment scores, especially for the HAMD (Faries et al., 2000), should be regarded as ordinal rather than numerical, and the numbers were originally proposed for ease of clinical use rather than containing any kind of reliable numerical severity measure. The ordinal nature will potentially affect the accuracy of a severity problem and is discussed further in Section 8. This ordinal nature also means depression scales are potentially not well suited to score level prediction analysis, except perhaps at a coarse level thus effectively becoming a multi-level severity of depression problem.

Typically classifier or prediction systems consist of two main parts (Fig. 2). The front-end extracts parameters to capture acoustic properties in the speech signal, using the Prosodic and Acoustic Features discussed in Section 6. Classification or score prediction is achieved through a stochastic modelling and matching process carried out in the back-end of the classifier. The back-end has two main phases; in the training phase, a function/model is learnt which can predict a certain output (speaker mental state) for a given input (features extracted from speech files which labelled with ground truth mental state information). In the testing phase, the back-end uses this function/model to assign a mental state label that best represents an unknown (mental state of speaker not known) speech sample.

The two most popular modelling and classification techniques used in the literature include *Support Vector Machines* (SVM) and *Gaussian Mixture Models* (GMM). The popularity of both methods is due, in part, to their ability to robustly handle smaller/sparse datasets,[7] their relative lack of computational expense and the existence of established software implementations. Both techniques have proven adept for mental state classification (Cummins et al., 2013a,b, 2011; Helfer et al., 2013; Scherer et al., 2013a,b,d; Trevino et al., 2011) and prediction (Cummins et al., 2013c; Valstar et al., 2013; Williamson et al., 2013). For in-depth explanations on SVM, SVM Regression (SVR) and GMM usage, the reader is referred to Burges (1998), Chang and Lin (2011), Smola and Schölkopf (2004), Reynolds and Rose (1995), and Reynolds et al. (2000).

Note that all classification accuracies reported in Sections 7.1 and 7.2, *sensitivity* (sens.) and *specificity* (spec.) or F_1 measure are also reported where available.

[7] Depression and Suicidal datasets are smaller both in terms of number of speakers and duration, when compared to speaker recognition data sets such as those available through NIST (Section 5).

7.1. Automatic classification of depressed speech

This section reviews investigations into the automatic analysis of speech as a predictor of depression with a section dedicated to each of the three different groups of problems; presence, severity or score level prediction. The key characteristic of all corpus used to generate the results discussed in this section are covered in Section 5.

7.1.1. Presence of depression

A wide range of features have been trialled for automatic depressed speech classification. Moore et al. (2008), Low et al. (2011) and Ooi et al. (2013) investigated the suitability of forming a classification system from combinations of prosodic, voice quality, spectral and glottal features. Moore et al. (2008) utilised statistical measures (pairwise ANOVA comparison) to construct a classifier using quadratic discriminant analysis, reporting a maximum accuracies of 91% (Sens. 0.89, Spec. 0.93) for male speakers and 96% (Sens. 0.98, Spec. 0.94) for female speakers, using leave-one out between cross validation. This analysis revealed the suitability of glottal features for discriminating between depressed subjects and healthy controls; however the authors acknowledge that due to their small sample size these results might not generalise to larger corpora.

A similar approach was used in Low et al. (2011), who reported classification accuracies ranging from 50% to 75% for a 2-class gender-independent GMM classifier. The analysis undertaken in the above papers indicate that both *Teager Energy Operator* (TEO) energy and glottal features improved – over either individual or combined prosodic- or spectral-based classifiers – the accuracy of a presence classification system; however these results were obtained by fine tuning the parameters and their back-end modelling technique, not allowing for consistent comparison between features. The results in both (Low et al., 2011; Moore et al., 2008) support the earlier discussed hypothesis that the effects of depression on muscle tension and laryngeal control result in more turbulent glottal flow (Section 6.2.1).

Ooi et al. (2013) also used a multi-feature approach, but instead of the feature space fusion techniques used by Moore et al. (2008) and Low et al. (2011), a final classification decision was made using the weighted sum of the intermediate decisions generated by separate GMM classifiers training on a particular feature space. Using this approach the authors report a binary classification of 73% (Sens. 0.79, Spec. 0.67) when differentiating between the speech of adolescents who are at risk of developing depression versus individuals not at risk.[8] The weighted fusion technique outperformed all single feature groups, although glottal 69% (Sens. 0.76, Spec. 0.63) and prosodic 63% (Sens. 0.73, Spec. 0.53) feature groups provided reasonable prediction results.

Several papers have investigated the suitability of individual prosodic, voice quality, spectral and glottal features. Amongst a wide range of acoustic features and spectral features tested using a GMM back-end in Cummins et al. (2011), MFCCs (77%), and formants (74%), and a combination of both (79%) displayed the strongest discriminatory characteristics when classifying the presence of depression. Similar results are also reported in Alghowinem et al. (2013a,b, 2012). Alghowinem et al. (2013b) also compared a range of different back-ends in their experiments and concluded that GMMs and SVMs were the better back-ends when classifying the presence of depression. Helfer et al. (2013) also report the stronger performance of SVMs over GMMs when classifying the severity of depression.

7.1.2. Severity of depression

Papers which have explored building classifiers that identify classes based on severity of depression tend to focus on a specific feature or feature set when designing their system. Cohn et al. (2009) and Trevino et al. (2011) both exploited prosodic changes associated with depression when building their classification systems. Cohn et al. (2009) used F_0 and speech/pause ratio as entries to a gender independent SVM classifier; reporting an accuracy of 79% (Sens. 0.88, Spec. 0.64) when classifying patients who either responded or not responded to depression treatment. Using their results showing that timing measures relating to phoneme length had stronger correlations with depression severity than global measures of speech rate (Section 6.1.1), Trevino et al. (2011) constructed a GMM back-end in a 5 class level-of-depression classifier using average phoneme lengths and average silence values as the input features and reported an RMSE of 1.6. By using a subset of the speech rate features, which were either manually or automatically selected to find an optimal solution, small improvements in performance were reported, resulting in a RMSE of approximately 1.25 for both selection methods.

Two papers led by Scherer have focused on building SVM classification systems based on source features (Scherer et al., 2013b,d). Both papers use the 9-item *Patient Health Questionnaire* (PHQ-9) scores associated with their dataset to divide their data into high and low classes: participants who scored above 10, corresponding to moderate depression and above, and participants who scored below 8, corresponding to mild to no depressive systems respectively (Manea et al., 2012). In the first paper, (Scherer et al., 2013b), use a combination of *Normalised Amplitude Quotient* (NAQ), *Quasi-Open-Quotient* (QOQ), PeakSlope and *Open Quotient Neural Network*[9] (OQ_{NN})

[8] Dataset obtained from a longitudinal study into adolescent depression. *Depressed* partition comprises adolescent voice sample taken 2 years before diagnosed with depression, *Control* partition comprises voice sample of adolescent in the study who did not develop depression.

[9] Ratio between the open phase of the glottal pulse and the length of the fundamental period estimated using standard Mel frequency cepstral coefficients and a trained neural network.

with a SVM employing a radial basis kernel, and report a classification accuracy of 75% (F_1 score for depression 0.77 and no-depression 0.73). The second paper (Scherer et al., 2013c) reported a combination of NAQ, QOQ and OQ_{NN}, yielding a classification accuracy of 51.28% (Mean F_1 Score 0.51) based on a SVM this time employing a 3rd order polynomial kernel. The authors attribute this poor result to an unsuitable SVM kernel, and used linear discriminant analysis to report an accuracy of 76.92% (Mean F_1 Score 0.75), supporting this claim.

Cummins et al. (2013a) built both 2-class and 5-class SVM (with RBF kernel) classifiers to investigate the potential discriminatory advantages of including long term (100–300 ms) spectral information when classifying low/high levels of depression. Maximum classification accuracies of 66.9% and 36.4% were obtained for the 2- and 5-class problems respectively, from features derived from the *Modulation Spectrum* (MS). The MS comprises the frequency components of time-evolving sub-band magnitudes from a spectrogram representation of speech, characterising both slow and fast rates of spectral change (Kinnunen et al., 2008). MFCC in combination with Shifted Delta Coefficients also displayed strong classification accuracy, 61.9% 2-class and 44.9% 5-class, leading the authors to conclude that medium-to-long term spectro-temporal information has strong discriminatory properties when classifying an individual's level of clinical depression from their speech.

Helfer et al. (2013) used features based on the first three formant trajectories and associated dynamic (velocity and acceleration) information in combination with *Principal Component Analysis* (PCA), to classify the 35-speaker Mundt dataset into high and low classes, comprising samples from speakers whose HAMD scores were above 17 or below 7 respectively. Using a GMM back-end, the authors report a maximum *area under the ROC curve*[10] (AUC) of 0.70 (Sens. 0.86, Spec. 0.64), whilst for a linear SVM a maximum AUC of 0.76 (Sens. 0.77, Spec. 0.77) was reported. The authors found that the maximum AUC was achieved for both classification systems when both formant trajectories and dynamic information found from either free response speech or held vowels were included in system training.

Cummins et al. (2013b) explored how depression is modelled and classified when using MFCCs in combination with the *GMM-universal background model* (UBM) and *Maximum A Posteriori* (MAP) adaptation to model acoustic space (Reynolds et al., 2000). The authors compared the classification accuracies found when performing full MAP adaption versus mean-only, variance-only and weight-only adaptation. The results of these experiments, in which strong performance of variance-only and weight-only adaptation results were observed, show the importance of including spectral variability when using spectral based features to classify either the presence or severity of depression.

7.1.3. Depression score prediction

The *Audio/Visual Emotion Challenges* in 2013 (AVEC 2013) and 2014 (AVEC 2014) involved participants predicting (using multimodal signal processing techniques) an individual's self-reported level of depression (Beck Depression Index, BDI, score) from a given multimedia file (Valstar et al., 2014, 2013). The Challenge organisers provided a depression dataset of 150 files[11] – divided into training, development and test partitions of 50 files each – for each year (Section 5). Each file is a recording from both a webcam and a microphone of the subject performing a different range of Human–Computer Interaction tasks (Valstar et al., 2014, 2013). The organisers of the challenges set audio published RMSE baselines of 10.75 and 11.52, for the AVEC 2013 and AVEC 2014 development sets respectively, and 14.12 and 12.57 for the AVEC 2013 and AVEC 2014 test sets respectively. All baselines were set with a linear SVR using the *brute-force* approach (Schuller et al., 2008) to construct a feature space with a dimensionality of 2268 (Valstar et al., 2014, 2013).

A typical approach when dealing with large feature spaces, such as the AVEC challenge's feature sets, is to perform dimensionality reduction. Two papers authored by Kaya et al. have explored different *data-driven*[12] dimensionality reduction approaches (Kaya et al., 2014a,b). Kaya et al. (2014a) used PCA in combination with the *Moore–Penrose Generalized Inverse* to reduce the feature space dimensionality of the challenge audio feature set with an *extreme learning machine* regressor to achieve a AVEC 2014 test set RMSE of 9.98. In Kaya et al. (2014b) *Canonical Correlation Analysis* feature selection was combined with a *Bagging-REPTree* (decision tree) regressor in order to reduce the large dimensionality of the challenge feature set. The authors report minimum RMSEs of 10.22 for the AVEC 2013 development set and 9.78 for the AVEC 2013 test set.

The AVEC 2013 and AVEC 2014 papers authored by Williamson et al. used an approach – as discussed in Section 6.3.1 – that in part aimed to exploit changes in coordination across articulators as reflected in formant frequencies (Williamson et al., 2014, 2013). Specifically, the authors investigated changes in correlation that occur at different time scales across formant frequencies and also across channels of the delta-mel-cepstrum. In this approach, channel-delay correlation and covariance matrices were computed from multiple time series channels.

[10] Percentage that the classifier correctly indentifies a pair of random patients in a binary test.

[11] For AVEC 2014 5 files were replaced from AVEC 2013 due to unsuitable data.

[12] Feature reduced on properties of training data as opposed to *knowledge-driven* dimensionality reduction where prior knowledge on vocal phenomena being modelled is used to select relevant features.

Each matrix contains correlation or covariance coefficients between the channels at multiple relative time delays. The approach is motivated by the observation that auto- and cross-correlations of measured signals can reveal hidden parameters in the stochastic-dynamical systems that generate the signals. Changes over time in the eigenvalue spectra of these channel-delay matrices register the temporal changes in coupling strengths among the channels.

For their 2013 challenge entry, Williamson et al. (2013) formed predictors using a feature set based on these eigenvalue spectra – coordination features – and combined them with a feature set involving average phonetic durations, i.e. phonetic-based speaking rates (Trevino et al., 2011). With these coordination- and phoneme-rate-based features, using a specially designed GMM-based regression system (a Gaussian "staircase" regression) and PCA to reduce feature space dimensionality, the authors report minimum development-stage RMSEs of 8.40 for the formant domain, 9.16 for the delta-mel domain, and 9.25 for phoneme-rate features, and an RMSE of 6.46 for the fused predictors. For the test stage of the challenge, the authors report minimum RMSE of 8.50 for the fused predictors.

This approach was fined tuned and extended for AVEC 2014 (Williamson et al., 2014). The phonetic-based features were expanded to include phoneme-dependent pitch dynamics (pitch slope over each phoneme), as well as the original phoneme-dependent rate feature. The frequency at which a phoneme occurs was also exploited. New coordination-based features were also added, including the correlation between formant frequencies and a cepstral-peak-prominence measure (Heman-Ackah et al., 2003), reflecting coordination between articulators and the vocal source. In addition, a correlation methodology for facial expression, analogous to that for articulatory coordination, was also applied. Specifically, facial features during speaking were derived from correlations across facial action units (Ekman et al., 1980), reflecting coordination of muscle units. The frequency at which a facial action unit occurs was also used. Fusing predictions from their vocal and facial coordination features, and phonetic-based durations and pitch slopes, together with complementary Gaussian "staircase" (used in their AVEC 2013 submission) and an *extreme learning machine* regressor, Williamson et al. (2014) achieved a best 2014 test set RMSE of 8.12. The advantages of combining both audio and visual information to improve depression score prediction is discussed further in Section 8.1.3.

In the 2013 challenge, Cummins et al. (2013c) employed MFCC/GMM supervectors (Kinnunen and Li, 2009) in combination with either *Kullback–Leibler* (KL-means) divergence kernel (Campbell et al., 2006), *Bhattacharyya Distance* based GMM-UBM mean interval kernel (You et al., 2010), or *UBM weight posterior probability* kernel (Li et al., 2013), which achieved RMSEs of 9.60, 9.56 and 12.01, respectively, on the AVEC 2013 development set. The KL-means, as most consistent performing audio feature in system development, was used as one of the authors' challenge entries, obtaining a test set RMSE of 10.17. Cummins et al. (2013c) also trialled techniques to remove unwanted acoustic variability in the supervector approach; this is discussed further in Section 8.2.2.

One popular approach to improving system robustness in speaker identification and verification is the *i-vector* paradigm (Dehak et al., 2011). An i-vector is the result of a linear transformation from a high dimensional supervector space to a low dimensional subspace. This mapping has the property that most of the variability (useful information) present in the supervector space is retained. An issue when using i-vectors for depression prediction is the small amount of training data available to train the mapping. Cummins et al. (2014a) presented an oversampling technique to train there i-vector predictor and report a 2013 development set RMSE of 10.13 and 2013 test set RMSE of 11.58. As part of the AVEC 2014 challenge Senoussaoui et al. (2014) used the TIMIT database (Zue et al., 1990) to supplement their i-vector training and report a best 2014 development set RMSE of 9.37 and 2014 test set RMSE of 11.03.

Mitra et al. (2014), in their AVEC-2014 entry, fused a range of different i-vector systems including standard (MFCC-based) i-vectors, prosodic based i-vectors and modulation based i-vectors in combination with a SVR backend. They report a best 2014 development set RMSE of 7.71, however this approach did not generalise well to the test set where it gave a RMSE of 11.10. The authors speculate their lack of generalisation was due to their strategy for training fusion weights (Mitra et al., 2014). The systems presented by Cummins et al. (2014a), Senoussaoui et al. (2014) and Mitra et al. (2014) all struggled to generalise well to their respective test set; challenges in minimising nuisance effects for depressed and suicidal speech is discussed in depth in Section 8.2.2.

Results presented by Cummins et al. (2013c) indicate that this phonetic variability is a strong confounding factor for Depression Score Prediction, similar results have been reported in emotion recognition (Sethu et al., 2008). Therefore comparing system performance between many of the AVEC papers is not a straightforward as doing a direct comparison of the reported RMSE. This is due to the differing phonetic content between files. Cummins et al. (2014a, 2013c), Kaya et al. (2014b) and Valstar et al. (2013) all used the entire provided AVEC 2013 files, which varied greatly in length and contained a mix vocal exercises, free and read speech tasks, noting also that not all files include all tasks. Williamson et al. (2013) reduced phonetic variability present in AVEC 2013 by isolating a read passage taken from the novel *Homo Faber* by Max Frisch.

In AVEC 2014, the challenge organizers provided only two passages from each session: the read passage *Die Sonne und der Wind* and a *free response* speech passage. The AVEC 2014 baselines were set using the *free response*

passages. Williamson et al. (2014) used the read passage to obtain their articulatory coordination features, and used both passages to obtain their phoneme-based features. They also used both passages for obtaining facial action unit frequency features, and the free response passage for their facial coordination features. Whilst Kaya et al. (2014a), Mitra et al. (2014) and Senoussaoui et al. (2014) combined information from the free response and read passage when making their prediction. The role of speech elicitation when classifying or predicting depression is discussed further in Section 8.1.2.

Recently, Hönig et al. (2014) compared the performance of *brute-forcing* to both a *knowledge-driven* data feature selection, and *data-driven* feature selection methodologies when performing score level prediction. Using ridge regression and a 4-fold speaker-independent cross fold validation set-up, the authors report average cross fold correlations – between the predicted and reference values – of 0.44 for brute-forcing, 3805 features, 0.39 for the knowledge-driven selection, 34 features which had previously been found suitable for modelling speaker fatigue, and 0.36 for the data-driven feature selection, 34 features chosen using a greedy forward search algorithm. The authors also report a similar set of correlation coefficients when dividing their dataset in terms of gender report correlations of 0.42 (brute), 0.36 (knowledge) and 0.31 (data) for females and 0.40 (brute), 0.39 (knowledge) and 0.39 (data) for males. The weaker performance of the data driven approach in the combined task highlights that using "black-box" techniques to narrow down feature sets can potentially result in a set of features which bear little physiological relevance to the vocal phenomena being modelled (Sundberg et al., 2011).

7.2. Automatic classification of suicidal speech

Several investigations, (France et al., 2000; Ozdas et al., 2004a; Yingthawornsuk et al., 2007, 2006), have investigated the discriminatory characteristics of speech features, in a pairwise classification set up, for determining if a speaker is in either a depressed, suicidal or neutral (control) mental state. Whilst all the papers report promising classification accuracies a major limitation of these papers, as discussed in France et al. (2000) and Ozdas et al. (2004a) and Section 6.4.2, is the fact the results are reported on datasets containing large variance in recording setups making it difficult to assess the true accuracy of the systems reported.

To the best of the author's knowledge, the work undertaken by France et al. (2000) is the first to investigate the automatic classification of suicidal voices using acoustic analysis. In this paper the authors hypothesise that both depressed and suicidal speech is characterised by a flattened spectrum. Using formant and power spectral density measures in combination with a Quadratic Classifier, the authors were able to discriminate between male control voices and suicidal voices with an accuracy of 80%.

In a follow-up study Ozdas et al. (2004a) investigated the discriminatory potential of source measures for differentiating control, depressed and suicidal voices. When using spectral slope in combination with a maximum likelihood binary classifier accuracies of 90% (depressed/control: Sens. 0.90, Spec. 0.90), 75% (depressed/suicidal: Sens. 0.90, Spec.0.60) and 60% (suicidal/control: Sens. 0.50, Spec. 0.70) were reported. For jitter, again in combination with a maximum likelihood classifier, accuracies of 65% (depressed/control: Sens. 0.70, Spec. 0.60), 60% (depressed/suicidal: Sens. 0.50, Spec. 0.70) and 80% (suicidal/control: Sens. 0.70, Spec. 0.90) were also reported. However, when combing the two features only the classification accuracy for the suicidal/control groups improved (85%: Sens. 0.80, Spec. 0.90) beyond the spectral slope results. Given the lack of standardization in the datasets and difficulty in using jitter as a diagnostic measure (Section 6.2) the results on this feature are hard to interpret. However, the strong performance of the spectral slope features combined with statistical analysis reported in the paper showing a decrease in spectral slope with increasing levels of suicidality and depression these results do match the hypothesis of spectral flattening reported in France et al. (2000).

Two papers, led by Yingthawornsuk, from the same research group have examined the spectral properties of male control,[13] depressed and suicidial voices (Yingthawornsuk et al., 2006) and female depressed and suicidal speech (Yingthawornsuk et al., 2007). Yingthawornsuk et al. (2006) analysed the voiced *power distribution spectral density* (PSD) features and reported high accuracy using interview style speech; 94% (depressed/control: Sens. 0.94, Spec. 0.94), 77% (depressed/suicidal: Sens. 0.89, Spec. 0.63) and 85% (suicidal/control: Sens. 0.91, Spec. 0.76). Yingthawornsuk et al. (2007) identified the location and bandwidth of F_2 as well as the PSD between 0–500 Hz and 500–1000 Hz as having good discriminatory power for a database consisting of 10 female depressive and 10 female pre-suicidal voices. Using these features in a logistic regression classifier, classification accuracies of 90% (Sens. 0.89, Spec. 0.92) using read speech samples and 85.75% (Sens. 0.89, Spec. 0.80) using interview style were achieved. It is not clear in either paper whether the cross-validation procedure utilised resulted in speaker-independent testing.[14] Both results presented are consistent with the spectral flattening hypothesis presented in France et al. (2000).

Scherer et al. (2013a), using a set of 16 adolescent speakers (8 suicidal and 8 age and gender matched controls) collected under more controlled recording conditions, compared the accuracy of HMM and SVM classifiers, using leave-one-out speaker cross validation. The HMM's had three states containing three mixtures per state

[13] Control samples were taken from patients in remission from associated suicidal prevention study.

[14] Speech samples from the same speaker present in both training and testing partitions.

employing a full transition matrix. To take advantage of the sequential and dynamic characteristics present in the data, the HMM's were trained using all frames extracted from each utterance. The SVMs, on the other hand, were trained on the median and standard deviations functionals calculated from all frames extracted from each utterance. Using a front-end of 12 extracted prosodic and voice quality features maximum accuracies of 81.25% and 75% for the HMM and SVM classifiers were reported respectively. For both classifiers all misclassifications were false negatives (suicidal adolescent misclassified as non-suicidal).

7.3. Classification and prediction: where to next?

In general, many of the papers discussed in Section 7 use techniques and systems designed and developed originally for automatic speech, speaker or (unrelated) paralinguistic recognition tasks. Whilst this is not an inappropriate approach for depressed and suicidal speech analysis – indeed the papers discussed represent an excellent set of baselines for future studies to compare against and future research using (future) methods from other speech-based disciplines may lead to a set of markers with clinical utility – it is also a potentially limiting approach. An example of this can be highlighted by considering a simple classifier built using a MFCC front-end and a generic back end. As speech affected by depressed and suicidality has a wide range of associated and sometimes conflicting vocal effects (Section 6) and detailed spectral features are representative of all the information contained in speech (Section 6.4), it seems unlikely that a classifier this simple could ever achieve 100% accuracy.

Therefore to increase the likelihood of finding a (set of) speech-based markers with clinical utility for depression and suicidality it is also important for the research community to move into more focused – hypothesis driven – research, where specific aspects of the effects of depression and suicidality on speech are considered in system design. The next section focuses on some of the main future challenges and research directions relating to a more specific approach towards an objective marker of depression or suicidality.

8. Future challenges and research directions

A major limiting factor in the search for a suitable speech based markers of either depression or suicide is the comparatively small, both in terms of number of speakers and duration, sizes of the datasets used to generate the findings discussed in Sections 6 and 7. This size limitation raises issues relating to validity, and generalisability (Douglas-Cowie et al., 2003). These issues are potentially further confounded by issues relating to the relative strength of different sources of variability in speech. Given that linguistic communication is the primary aim of speech and the comparatively high classification accuracies of other speech-based tasks; i.e. speaker recognition and other paralinguistic tasks, such as age and gender recognition (Schuller et al., 2013), when compared to the accuracy of many mental state systems it seems reasonable to assume depression and suicidality have a low degree of variability in speech i.e. between-class (Low or Control versus High level of depression or suicide risk) acoustic variability is diluted by linguistic information and speaker characteristics.

Therefore, future challenges in finding a speech-based marker of either depression or suicide proposed herein relate to improving the generalisability of our findings through greater research collaboration and cooperation (Section 8.1) and in researching techniques and strategies to highlight more of the acoustic variability and discriminatory information due to depression and suicidality in speech (Section 8.2). We note that because it is arguably the more established field, Sections 8.1 and 8.2 are focused more towards depression research; the collection of suitable data to build on the findings discussed in Sections 6 and 7 is the number one challenge in suicidal speech research.

8.1. Need for greater research collaboration and cooperation

To find valid clinical speech-based markers for either depression or suicide there is a need for greater research collaboration and cooperation. Generalising findings across a range of datasets can only help uncover a true objective mental state marker. The authors have identified three major factors to help facilitate this: (1) the need for sharing both datasets and code and the need for greater transparency in published results, (2) a more standardised approach to data collection and (3) collaboration with practicing clinicians and healthcare providers to help validate the usefulness and benefits of behavioural markers in clinical settings.

8.1.1. Sharing and transparency

To facilitate greater research collaboration we encourage sharing of data and code needed to extract novel features. The recent Audio/Visual Emotion Challenge (AVEC) datasets (Section 5) is an excellent collaborative resource, allowing for comparisons of a wide range of multimodal score level prediction approaches under well-defined and strictly comparable conditions (Valstar et al., 2014, 2013). The authors recognise that due to the sensitive and often personal nature of mental state related datasets, ethical constraints make it justifiably hard to share speech data in its raw form. Therefore we recommend that researchers who are starting in the field include a provision in any ethics applications about sharing a minimum of audio features in a form from which the raw data cannot be reconstructed, such as MFCC's or formants with unvoiced frames already removed. Researchers active in the field can also significantly aid collaboration efforts by sharing and running code from other research groups.

As direct research collaboration is not always possible, including a precise description, or full referencing, for the extraction of the audio features used and a full system configuration description when publishing results helps ensure reproducibility by other research groups. System configuration information could include coefficient values, class divisions, classifier (model) used, scoring or classification technique and a clear definition of training and test partitioning or cross-fold-validation technique used including information relating to the speaker independence of class, test and training divisions. When reporting results in terms of classification accuracy, publishing statistical measures of performance such as sensitivity and specificity or F_1 measures is highly recommended.

8.1.2. Standardisation of data collection

Whilst research goals will dictate the final make-up of any data collection paradigm, in this section we discuss issues relating to the standardisation of data collection including issues to consider when constructing a new database (Fig. 3); establishing a "ground truth", a consistent – in terms of equipment – recording set up, eliciting suitable speech samples, and potential advantages of collecting both extra clinical information relating to sub-symptom severity and extra behavioural, physiological and biological markers in parallel with audio data. We conclude this subsection by drawing attention to possible researcher health effects related to the collection and annotation of data.

8.1.2.1. Ground truth: Due to the multi-dimensional nature of depression and suicidal risk assessment tools and differences in symptoms analysed and weighting when calculating overall severity or risk, the choice of assessment methodology will affect the results of any behavioural marker based study of mental state. When choosing a clinical scale for a new project, researchers should select a commonly used scale with clinical validity and be aware of both the strengths and weaknesses of their chosen scale and how these factors could affect their study.

8.1.2.1.1. Clinician-led or self-reported: These are many factors to consider when deciding between either a clinician assessed or self-reported measure. Clinicians are thought to measure depressive severity and suicide risk more accurately, but are more time intensive. Self-rated scales are quicker and potentially more sensitive to detect changes than clinician-rated scales in milder forms of depression; they require minimal reading skills, are open to errors due to patient familiarity and require a patient's willingness and cooperation to complete.

8.1.2.1.2. Class divisions: Another factor relating to clinical assessment is class division for classification or statistical analysis. Given the ordinal nature of mental state scales, we recommend removing *moderate* speakers when forming low or high classes; i.e. if using the HAMD scale, a low class could consist of speech samples from the *Normal* and *Mild* categories (HAMDtotal $\leqslant 13$) and a high class could consist of speech samples from the *Severe* and *Very Severe* categories (HAMDtotal $\geqslant 19$). For a study which seeks to find behavioural markers for the detection of someone as clinically depressed (as opposed to a severity level) the collection of an age and gender matched control group consisting of individuals who are not clinically depressed or at a high suicidal risk with no family history of mental illness could be beneficial.

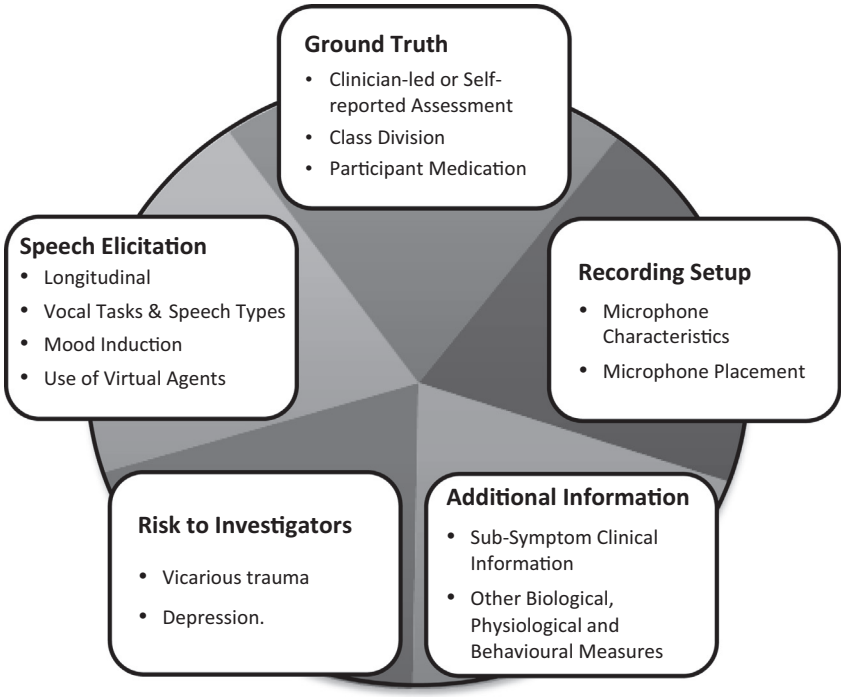

Fig. 3. Issues relating to standardization of data collection; increased standardization can help address validity and generalisability concerns when eliciting speech-based markers of mental state.

8.1.2.1.3. Participant medication: The presence of medication in a speech based dataset is a strong confounding factor. Common side effects which effect speech production associated with antidepressant medication include dry mouth (inadequate saliva production), fatigue, tremors and agitation (Crawford et al., 2014; Kikuchi et al., 2012). Dyskinesia (involuntary muscle movements) is also associated with antipsychotic medication (De Hert et al., 2012). Whilst exclusion of medicated participants is the simple solution to these effects, this might not be in-line with the research aims of a given study, therefore disclosure of medication present or disclosure of unknown medications in a dataset should be given in relevant papers.

8.1.2.2. Recording set-up: To minimise channel related effects, consistency is needed with the microphone and recording equipment used throughout the data collection. The negative effect of varying channel characteristics and techniques to mitigate the effects are widely studied in speaker recognition, (Kenny et al., 2008; Kinnunen and Li, 2009) and discussed further in relation to mental state in Section 8.2. Source and energy based measures are susceptible to errors relating to differing sound pressure levels (Orlikoff and Kahane, 1991) therefore there is a need to help ensure a consistent distance from the speaker's mouth to the microphone. Non wearable microphones such as a desk mounted microphone could be susceptible to errors from slumped body posture as well as other effects relating to psychomotor retardation or increased agitation.

8.1.2.3. Speech elicitation: Speech contains a wide variety of acoustic information variability, which arises primarily due to speaker characteristics, channel effects and phonetic content. Therefore, successfully identifying a speech-based marker of mental state will potentially be influenced in part by the paradigm the speech data was collected under.

8.1.2.3.1. Longitudinal: Due to differences in speaker characteristics and the individuality of speaking styles in combination with the wide profile of either, longitudinal data collection is recommended. Through the collection of multiple recordings per speaker from a wide range of speakers over the period of many months, in combination with clinical treatment, insights could be gained into how individual, as well as group, speech characteristics change in response to treatment. Further, longitudinal data allows collection of potentially multiple classes within the same speaker (Clinical categories associated with assessment tool, Section 2.2.1), allowing reduction of the speaker variability effects (Section 8.2.2).

8.1.2.3.2. Vocal tasks and speech types: The behaviour of speech produced by depressed individuals has been shown to vary due to the negativity of conversational content (Vanger et al., 1992) or cognitive effort required (Calev et al., 1989) regardless of their level of depression. Therefore, read speech and vocal exercises, despite having the advantage of being able to offer specific phonetic extraction with minimal variability, may not strongly reflect the acoustic effects of depression. This effect has been reported in Alpert et al. (2001), Ellgring and Scherer (1996), Nilsonne et al. (1987) and Stassen et al. (1991). Using free speech negates practise effects and potentially allows for a greater range of induced emotional effects, such as asking participants to describe events that has aroused significant emotions, McIntyre et al. (2009) but is more phonetically varied, meaning greater effort may be required to isolate suitable speech segments for acoustic analysis. Many participants in AVEC 2014 found the free-form data offered greater predictor performance over the read passage (Gupta et al., 2014; Pérez et al., 2014; Senoussaoui et al., 2014; Sidorov and Minker, 2014; Valstar et al., 2014). Ideally, as each has a different range of advantages, a mix of read speech, vocal exercises and free response speech should be collected.

8.1.2.3.3. Mood induction: Negative affect and participant engagement potentially play an important role in identifying speech-based markers, whilst an individual's range and ability to exhibit emotional responses are inhibited during depression (McIntyre et al., 2009): depressed patients show a specific failure to impair inhibitions relating to negative information (Goeleven et al., 2006). McIntyre et al. (2009) present a mood induction paradigm that includes watching affective video clips, watching and rating pictures from the International Affective Picture System (Lang et al., 2005), reading sentences containing affective content and asking participants to describe events relating to significant emotions. Negative Priming tasks, in which participants are required to respond to the target stimuli by evaluating it as negative or positive, while ignoring distracter stimuli could also be used, commonly used stimuli include emotional words (Joormann and Gotlib, 2008) or pictures (Goeleven et al., 2006). Note: even *without* a mood induction paradigm it is possible – due to the nature of depression and suicidality – to inadvertently induce anxiety, annoyance or anger within participants during the collection process, reducing the usefulness of results found when transferring them to non-research daily life speaking settings. This could also have an undue influence on longitudinal within-subject studies where improvements in speech are due to increased ease with the recording paradigm as opposed to any reduction in depression or suicidality level. The use of virtual agents in the data collection, to create the exact same experience in different participants is one possible solution to mitigating unwanted mood effects.

8.1.2.3.4. Interviewer bias and virtual human interviewers: Virtual humans hold several advantages over their natural counterparts in regards to eliciting suitable speech data (Hart et al., 2013). The involvement and use of virtual humans increases the available level of control for the investigators or clinical personnel over the assessment process and the presentation of stimuli (e.g. questions with positive or negative affect); the virtual human's behaviour can be controlled to the slightest detail and no behavioural bias is introduced into the interview process. This level of

control bears potential advantages for the comparability between studies and subjects over common human to human assessment interview corpora, which may comprise interviewer-induced biases. Further, findings suggest that virtual humans can reduce the stress and fear associated with the perception of being judged, and thereby, lower emotional barriers to seeking help (Hart et al., 2013). Virtual humans have also been studied within the context of schizophrenia, depression, and autism (Kim et al., 2009; Ku et al., 2006; Robins et al., 2005; Scherer et al., 2013d).

Another potential benefit of using virtual human interviewers is that researchers may be able to get more, or richer, samples of speech than with real human interviewers. Interacting with a virtual human can increase participants' willingness to say more. In particular, an investigation of the effects of framing the character as human-controlled or autonomous showed that participants felt more comfortable disclosing personal information with a character that was framed as autonomous than when it was framed as human-controlled (Gratch et al., 2014). Specifically, participants reported experiencing lower fear of negative evaluation and engaged in less impression management when the character was framed as autonomous than when it was framed as human-controlled (Gratch et al., 2014). In fact, the actual method of data collection (i.e. human-controlled versus automated agent interviews) had no impact on fear of negative evaluation or impression management, but participants who believed they were interacting with a human versus a computer felt the effect of both fear of negative evaluation and impression management.

In recent investigations, DeVault et al. (2014), a virtual human has already been used in depression screening interviews. These virtual human interviews comprise the distress assessment interview corpus (DAIC) that has been used in several investigations to assess nonverbal behaviour indicative of depression. In particular, researchers have thoroughly investigated visual nonverbal behaviour (Scherer et al., 2013d; Stratou et al., 2014), voice quality characteristics (Scherer et al., 2013b), verbal and temporal characteristics (DeVault et al., 2013) as well as multimodal behaviour descriptors of depression and PTSD (Scherer et al., 2013c; Zhou et al., 2013).

8.1.2.4. Additional clinical information and modalities: Given the wide clinical profile of both conditions, significant benefits in could be found through either the fusion of a multimodal set of behavioural markers - provided of course that the different modalities are sufficiently diverse – or through analysing the relationships between different vocal features and individual symptoms of depression or suicidality may prove advantageous (Section 8.2.2) in which case it may be advantageous to collect clinical information relating to sub symptoms of either condition.

8.1.2.4.1. Sub-symptom clinical information: Assessment scales such as CORE (Parker and Hadzi-Pavlovic, 1996) or the Beck Anxiety Inventory (Beck et al., 1988), which asses an individual's level psychomotor retardation (PMR) and anxiety respectively, could be collected in concurrence with depression assessment information. There are a range of objective tests to independently assess a patient's level of PMR and cognitive load; such as the trail making task or digit symbol substitution test drawing tasks for PMR (Buyukdura et al., 2011) and the Stroop test (Stroop, 1935), Sternberg recognition task (Oberauer, 2001) or memory span tests (Conway et al., 2005) for cognitive load. The Cambridge Automated Neuropsychological Test Battery (Buyukdura et al., 2011) provides of a range of objective motor performance and cognitive test which have been shown to be useful for evaluating neurocognitive disturbances in depression (Sweeney et al., 2000; Weiland-Fiedler et al., 2004).

8.1.2.4.2. Other biological, physiological and behavioural measures: Multimodal affect recognition is a well-established field; D'Mello and Kory (2012) analysed 30 different published studies which analysing facial expressions, acoustic–prosodic cues, body movements, postures, gestures, physiological signals such as EEG and biosignals for both unimodal and multimodal affect detection. Results indicated that, on average, multimodal systems offer an 8% relative improvement over unimodal systems. Initial studies have indicated the advantages of combining speech with visual information (Cummins et al., 2013c; Joshi et al., 2013; Kaya and Salah, 2014; Scherer et al., 2013d). Many participants in AVEC 2014 found fusing audio and visual information improved their predictor accuracy (Gupta et al., 2014; Kaya et al., 2014a; Pérez et al., 2014; Senoussaoui et al., 2014; Sidorov and Minker, 2014; Williamson et al., 2014).

Further advantages may be found in fusing non-acoustic and prosodic features present in depressed or suicidal speech. Linguistic analysis reveals more negative content in depressed speech (Segrin, 2000), whilst a wide range of disfluencies, changes in non-linguistic verbalizations such as laughter and increases in speaker-switch disfluencies during clinical interviews are all reported with increasing levels of speaker depression (Alghowinem et al., 2012; Cohn et al., 2009; DeVault et al., 2013; Navarro et al., 2014; Yang et al., 2012). A small number of studies have shown that fusing bio-signals with acoustic information (Kim, 2007) or prosodic information (Fragopanagos and Taylor, 2005) improves emotion recognition accuracy, beyond that found using each modality separately. As well as visual information (including information relating to facial, head, body and eye movements), blood samples, saliva samples and bio-signals such as electromyogram, electrocardiogram and skin conductivity could also be considered.

8.1.2.5. Warning: risk to investigators: There is a range of potential health risks *to investigators* associated with collection of severely depressed and suicidal data. Direct interactions with depressed and suicidal individuals during

collection or subsequent exposure to recorded data during tasks such as annotation can lead to researcher health risks including vicarious trauma and depression. This risk is magnified in researchers from non-clinical backgrounds, who might be unfamiliar with either condition. There is considerable anecdotal evidence suggesting that a range of adverse effects can be experienced by investigators following exposure to collected depressed and particularly pre-suicidal speech. Current best practise the authors are aware of includes the following risk reduction measures: (i) explaining risks upfront to investigators; (ii) minimising listening exposure to raw speech data as much as possible, particularly for suicidal speech; (iii) commencing any research involving listening to the data with one or two typical sessions of listening, followed by a consultation with a suitably qualified health professional, e.g. a trauma psychologist. This professional should then provide guidance and restrictions on any ensuing listening; (iv) limiting the amount of listening on any given day; (v) providing qualified clinical assistance (e.g. counselling, general practitioner or psychologist) on an as-needed basis to investigators; (vi) avoiding the use of headphones while listening; and (vii) logging the hours spent listening and discussing the listening regularly with other investigators. These practices may be even more imperative where the data comprise video rather than audio-only recordings. These practices may be slightly less imperative when the recorded language is unfamiliar to the listener, however caution is still urged.[15]

8.1.3. Dissemination

Future research should additionally seek to incorporate the collaboration with practicing clinicians and healthcare providers, who are potential customers of the proposed technology. Hence, future research endeavours should actively seek to validate the usefulness and benefit on clinical assessment accuracy and clinical training of the quantified acoustic, multimodal, dyadic, and longitudinal behavioural observations in clinician centred user studies.

Researchers should not lose focus on the use of the developed technology that should ultimately aim to *support clinicians* during practice with objective observations about a patient's behaviour and the behavioural development over time. For example, this new technology could assist doctors during remote telemedicine sessions that lack the communication cues provided in face-to-face interactions. Automatic behaviour analysis can further add quantitative information to the interactions such as behaviour dynamics and intensities. These quantitative data can improve both post-session and online analysis. Further, this technology can lay the ground for novel *educational and training* material to clinicians in training through access to complementary information.

8.2. Nuisance factors

As previously discussed, it seems reasonable to assume that depression and suicidality have a low degree of variability in speech compared to linguistic information and speaker characteristics. It can be seen from emotion recognition research that these unwanted forms of variability, herein referred to as nuisance factors, place an upper limit on the accuracy of a speech based emotion recognition system (Sethu et al., 2014); again it seems reasonable to assume this limit would apply in depression and suicidality systems (Section 7.3). Given this potential limitation, a key area of future research is investigating nuisance mitigation approaches specific to depressed and suicidal speech.

8.2.1. Sources of nuisance variability

In general when eliciting a speech based marker of depression or suicidality the following sources of nuisance variability can be distinguished: (i) *Biological Trait Primitives* such as race, ethnicity, gender, age, height, and weight; (ii) *Cultural Trait Primitives* such as first language, dialect, and sociolect; (iii) *Emotional Signals* such as anger, fear, and energetic states; (iv) *Social Signals* such as conversational sounds, intimacy and dominance; and (v) *Voice Pathology* such as speech disorders, intoxication and respiratory tract infection.

For depression recognition, specific nuisance factors such as speaker characteristics, phonetic content and inter-session (recording setup) variability have all been shown to be strong confounders (Cummins et al., 2014a, 2013c, 2011; Sturim et al., 2011). Another potential source of unwanted variability is the co-existence of other forms of paralinguistic information, relating to symptoms of depression in speech. An example of this nuisance variability can be seen in the AVEC 2013 and 2014 corpora: the challenge organisers were able to run two separate sub-challenges on the same depression dataset, a depression recognition challenge and a prediction of changing affective state challenge (Valstar et al., 2014, 2013).

Three papers led by Cummins highlight the confounding effects of both phonetic content and speaker characteristics. Results presented in Cummins et al. (2011) indicate that feature warping, as a per-speaker feature space normalisation technique (Pelecanos and Sridharan, 2001), offers little to no improvement in a presence of depression classifier. The authors speculate a possible reason for this is the low degree of variability depression has in a speech signal when compared to individual speaker characteristics. Cummins et al. (2013c) demonstrate the effect of the large degree of phonetic variability captured in AVEC 2013 development set. By generating multiple *sub-utterances* per file, with each sub-utterances differing in phonetic content, the authors show that a wide range of prediction

[15] The authors from the University of New South Wales are in the process of devising an Occupational Health and Safety scheme in relation to exposur recording of individuals suffering a mental illness. Please contact Julien Epps for further details. Other researchers improving on these practices are encouraged to share their experiences with the research community.

Table 11
Examples of *symptoms* of depression and suicidality that affect speech production and are measurable using many of the features and techniques discussed in this paper.

Effect	Depression	Suicide	Reference
Anxiety	√	√	Harrigan et al. (2004) Weeks et al. (2012)
Agitation	√	√	Petrushin (1999) Hönig et al. (2014)
Fatigue/Insomnia	√	√	Krajewski et al. (2012), Vogel et al. (2010)
Working memory impairments	√	√	Yap et al. (2010)
PMR	√	✗	Cummins et al. (2013a) Quatieri and Malyska (2012)
Low mood	√	√	Horwitz et al. (2013) Trevino et al. (2011)
Strong negative affect *or*	√	√	Cowie et al. (2001)
Intense affective states	✗	√	Ververidis and Kotropoulos (2006)

scores can be generated for each file. Cummins et al. (2014a) provide visualisations of an i-vector feature space (Dehak et al., 2011) before and after variability compensation, which suggest that speaker variability is a strong confounding effect when distinguishing between low or high levels of speaker depression.

A further source of nuisance variability stems from the heterogeneous clinical profile of both depression and suicide. Speech, a highly complex cognitive and muscular action, Section 4.1, is sensitive to not only changes in depressive or suicidality status, but also changes in a range of symptoms commonly associated with both conditions (Table 11). Given that many of the speech features discussed in Section 6 and classification techniques discussed in Section 7 can be used to elicit speech-based information relating to these effects, care is needed in speech based analysis to ensure that the speech changes are due to either depression or suicidality, not just a marker of a related symptom.

8.2.2. Nuisance mitigating

Many nuisance mitigation techniques exist from speech and speaker recognition that can help improve robustness against non-ideal acoustic conditions and confounding forms of variability, such as Nuisance Attribute Projection (NAP), i-vectors and Joint Factor Analysis (JFA) (Campbell et al., 2006; Dehak et al., 2011; Kenny et al., 2008). Whilst many of these have had some form of success in other paralinguistic tasks such as factor analysis in age, gender and emotion classification (Li et al., 2013, 2012; Sethu et al., 2013; Xia and Liu, 2012) there have been limited investigations into advantages they might offer depressed classification (Cummins et al., 2014a, 2013c; Senoussaoui et al., 2014).

The application of these techniques may be complicated for both depressed and suicidal speech by comparatively small databases (Sections 5 and 7). Further, depression and suicidal databases often only have examples of one class (i.e. low or high depression but not both) from a single speaker among the training/development data (Cummins et al., 2013b). By comparison, in emotion or speaker recognition, (often balanced) training databases exist with examples of many emotions or recording conditions per speaker (Sethu et al., 2013). This potentially limits the successful application of many standard nuisance mitigation techniques as within-subject effects cannot be systematically measured.

Techniques are needed to improve variability in the paralinguistic information that relates specifically to depression and suicidality; it is this variability that contains the discriminatory information needed to build an accurate classifier or predictor. As discussed in Section 7.3 we encourage hypothesis driven research to improve variability; whilst brute force machine learning approaches offers a potential solution they create systems which are essentially black boxes. Recent research highlights the benefits of a more depression focused approach; both Sturim et al. (2011) and Cummins et al. (2014a) introduce new approaches for analysing the effect of depression in the MFCC/GMM representation of acoustic space. Williamson et al. (2013) introduced a novel and promising approach to analysing articulator coordination, winning the AVEC 2013 Depression sub challenge.

Sturim et al. (2011) enhanced a MFCC/GMM system by using a Wiener filtering factor analysis method to successfully reduce the effects of both speaker and intersession variability when classifying the severity of depression. Using the Mundt 2007 dataset in a 2-class gender independent set up, the authors report absolute reductions in EER of ∼21% and ∼29% for the male and female systems respectively when using this technique compared to a uncompensated equivalent.[16] Cummins et al. (2014a) proposed a novel method for measuring acoustic volume – a GMM-based measure that is reflective of the decreasing spectral variability commonly reported with depressed speech (Section 6.4.1). The authors use this technique to demonstrate, with statistical significance, that MFCC

[16] Equal Error Rate; the rate at which both acceptance (recognising a condition as present when it is not) and rejection (failing to recognise condition present) errors are equal.

feature space becomes more tightly concentrated with increasing levels of speaker depression.

Williamson et al. (2013), who investigated changes in correlation that occur at different time scales across both formant frequencies and channels of the delta-mel-cepstrum (Section 7.1.3), employed an approach that provides information relating to coordination in vocal tract articulation whilst reducing the effects of a slowly-varying linear channel introduced by intersession variability. Further, their Gaussian "staircase" regression system, by first assigning a given test vector a probability relating to a partition of BDI before using univariate regression with the relating test statistics to assign a final score, offers a potential solution to dealing with the ordinal nature of mental state assessments (Section 7). Finally the system is composed of multiple Gaussian's it provides a familiar platform for feature fusion and Bayesian Adaption.

Another promising nuisance mitigating approach is to use the variability relating to the homogeneous clinical profile of depression and perform fine grained analysis of the relationships between different prosodic and acoustic features at the symptoms level of depression (Horwitz et al., 2013; Quatieri and Malyska, 2012; Trevino et al., 2011). The effects of PMR on energy and spectral-based features is a well-established area of research (Section 6.4.1), however, recent studies have revealed potentially strong relationships between prosodic, source and formant features and *individual* HAMD items. Trevino et al. (2011) report significantly stronger correlations between their combined phoneme rate measure, Section 6.1.1, and individual HAMD items such as PMR and low mood, when compared to correlating the same measure with the total HAMD score. Quatieri and Malyska (2012) report significant positive correlation between jitter, shimmer and pitch variance with HAMD-PMR and significant negative correlations with aspiration and energy variance again with HAMD-PMR. Similarly Horwitz et al. (2013) report strong negative correlation between speech rate and formant measures with low mood and for $F1$ velocity, in particular significant negative correlation with PMR and positive correlation with agitation. Given these strong relationships, continued research into the behaviour of speech at the symptom levels can potentially provide a set of speech markers which emcompass all manifestations of depression.

9. Concluding remarks

Both the diagnosis of depression and assessment of suicide risk are difficult and time consuming tasks; success often relies on the skill and experience of a clinician in eliciting suitable diagnostic information from a patient whom by definition will have impaired outlook and motivation. There is a real need for a simple, low cost, automated and objective diagnostic aid for use in both primary care settings and specialist clinics. Such a tool could be a game changing in terms of patient monitoring; providing immediate feedback and therapeutic mental health advice reducing the strain current diagnostic methods place on medical systems. Further, it could help in improving the quality of life of someone suffering from or susceptible to mental illness. Given the wide clinical profile of both conditions seems unlikely that a single biological, physiological and behavioural marker will be found; the best solution is most likely a multifaceted approach carefully built classifiers sensitive to individual items from a depression or suicidal scale. The findings of this review suggest that speech will be a key component in the search for an objective marker.

The primary aim of this paper was to discuss the effects of depression and suicidality on speech characteristics and review investigations into the automatic analysis of speech as a predictor of either condition. To achieve this we also discussed; current diagnostic and assessment methods, a range of potential (non-speech) biological, physiological and behavioural markers and the expected physiological and cognitive effects of either condition on speech production. As the validity and generalisability of findings discussed in this review rest upon them, we also reviewed the characteristics of depressed and suicidal speech databases currently used in this field of research. We finished by discussing future challenges and directions associated with finding a speech-based marker of either depression or suicide focusing on the need for greater research collaboration including dissemination into clinical practise, the standardisation of depressed and suicidal speech data collection and investigating into nuisance mitigation approaches specific to depressed and suicidal speech.

The smaller sizes of many depression and suicidal speech data represent a major challenge in improving the validity and generalisability of many of the findings discussed in this review. As database sizes increases our understanding of the limit in accuracy a speech-based system can achieve, as well as the system design needed to achieve this limit, will also increase. If speech can be used as a useful clinical marker the then it can be deployed in a lot of contexts such as remote monitoring using a telephone or Voice-over-Internet Protocol based system for susceptible individuals whilst dissemination into clinical practise could open up new areas of clinical research such as new diagnostic aids and training aids. A combination of growing size of databases, increases in research collaboration – especially dissemination into clinical practise – and more hypothesis driven research should lead to significant advances in this fascinating and potentially lifesaving application of speech processing.

Acknowledgements

The work of Nicholas Cummins and Julien Epps is supported by National ICT Australia, funded by the Australian Government as represented by the Department of Broadband, Communication and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program. Julien Epps is also supported by the Australian Research Council through Discovery Projects DP110105240 and DP120100641. The work of

Jarek Krajewski and Sebastian Schnieder is partly funded by the German Research Foundation (KR3698/4-1). MIT Lincoln Laboratory work is sponsored by the Assistant Secretary of Defence for Research & Engineering under Air Force contract #FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

Appendix A. Although the HAMD form lists 21 items, the scoring can be based on the first 17 items.

0–7 = Normal
8–13 = Mild Depression
14–18 = Moderate Depression
19–22 = Severe Depression
>23 = Very Severe Depression

1. Depressed mood
(Gloomy attitude, pessimism about the future, feeling of sadness, tendency to weep)
0 = Absent
1 = Sadness, etc.
2 = Occasional weeping
3 = Frequent weeping
4 = Extreme symptoms

2. Feelings of guilt
0 = Absent
1 = Self-reproach, feels he/she has let people down
2 = Ideas of guilt
3 = Present illness is a punishment; delusions of guilt
4 = Hallucinations of guilt

3. Suicide
0 = Absent
1 = Feels life is not worth living
2 = Wishes he/she were dead
3 = Suicidal ideas or gestures
4 = Attempts at suicide

4. Insomnia – Initial
(Difficulty in falling asleep)
0 = Absent
1 = Occasional
2 = Frequent

5. Insomnia – Middle
(Complains of being restless and disturbed during the night. Waking during the night)
0 = Absent
1 = Occasional
2 = Frequent

6. Insomnia – Delayed
(Waking in early hours of the morning and unable to fall asleep again)
0 = Absent
1 = Occasional
2 = Frequent

7. Work and interests
0 = No difficulty
1 = Feelings of incapacity, listlessness, indecision and vacillation
2 = Loss of interest in hobbies, decreased social activities
3 = Productivity decreased
4 = Unable to work. Stopped working because of present illness only.

8. Retardation
(Slowness of thought, speech, and activity; apathy; stupor.)
0 = Absent
1 = Slight retardation at interview
2 = Obvious retardation at interview
3 = Interview difficult
4 = Complete stupor

9. Agitation
(Restlessness associated with anxiety.)
0 = Absent
1 = Occasional
2 = Frequent

10. Anxiety – psychic
0 = No difficulty
1 = Tension and irritability
2 = Worrying about minor matters
3 = Apprehensive attitude
4 = Fears

11. Anxiety – somatic
(Gastrointestinal, indigestion, Cardiovascular, palpitation, Headaches Respiratory, Genito-urinary, etc.)
0 = Absent
1 = Mild
2 = Moderate
3 = Severe
4 = Incapacitating

12. Somatic symptoms – gastrointestinal
(Loss of appetite, heavy feeling in abdomen; constipation)
0 = Absent
1 = Mild
2 = Severe

13. Somatic symptoms – general
(Heaviness in limbs, back or head; diffuse backache; loss of energy and fatiguability)
0 = Absent
1 = Mild
2 = Severe

14. Genital symptoms
(Loss of libido, menstrual disturbances)
0 = Absent
1 = Mild
2 = Severe

15. Hypochondriasis
0 = Not present
1 = Self-absorption (bodily)
2 = Preoccupation with health
3 = Querulous attitude
4 = Hypochondriacal delusions

16. Weight loss
0 = No weight loss
1 = Slight
2 = Obvious or severe

17. Insight
(Insight must be interpreted in terms of patient's understanding and background.)
0 = No loss
1 = Partial or doubtfull loss
2 = Loss of insight

Total items 1–17: _____

18. Diurnal variation
(Symptoms worse in morning or evening. Note which it is.)
0 = No variation
1 = Mild variation; AM ( ) PM ( )
2 = Severe variation; AM ( ) PM ( )

19. Depersonalization and derealisation
(feelings of unreality, nihilistic ideas)
0 = Absent
1 = Mild
2 = Moderate
3 = Severe
4 = Incapacitating

20. Paranoid symptoms
(Not with a depressive quality)
0 = None
1 = Suspicious
2 = Ideas of reference
3 = Delusions of reference and persecution
4 = Hallucinations, persecutory

21. Obsessional symptoms
(Obsessive thoughts and compulsions against which the patient struggles)
0 = Absent
1 = Mild
2 = Severe

Total items 1–21: _____

Appendix B.

Best Practice in Suicide Assessment reproduced from Bisconer and Gross (2007)

1. Conduct objective rating of suicide behaviour using a standardised instrument.
2. Conduct mental status exam (e.g., assessment of general appearance, behaviour, alertness, orientation, memory, mood, affect, hallucinations, thought processes, thought content, reality testing, speech, eye contact, and sleep and appetite patterns).
3. Evaluate history of treatment for psychiatric illness and compliance with psychiatric medication and treatment.
4. Evaluate substance use history and current use.
5. Evaluate suicide ideation, intention, plan, lethality of plan, means, immediacy, history of gestures, lethality of previous gestures, family history of suicide, and access to means (e.g., firearms, pills).
6. Evaluate history of parasuicidal behaviour found with persons with borderline personality disorder (e.g., cutting, burning).
7. Evaluate recent losses (e.g., death, relationship, job, recent geographical move).
8. Evaluate social support systems (e.g., social isolation, stressful or dysfunctional relationships).
9. Evaluate health problems, legal problems, and problems with current job or unemployment.
10. Repeat standardised assessment and interview prior to discharge from inpatient and outpatient services.
11. The clinician should consider whether the individual, by nature of his or her gender, ethnic group, age, psychiatric diagnoses, and life circumstances, falls into high-risk statistical categories.

References

Åsberg, M., 1997. Neurotransmitters and suicidal behavior. Ann. NY Acad. Sci. 836, 158–181.

Abel, L.A., Friedman, L., Jesberger, J., Malki, A., Meltzer, H.Y., 1991. Quantitative assessment of smooth pursuit gain and catch-up saccades in schizophrenia and affective disorders. Biol. Psych. 29, 1063–1072.

Airas, M., 2008. TKK Aparat: an environment for voice inverse filtering and parameterization. Logop. Phoniatr. Vocology 33, 49–64.

Alghowinem, S., Goecke, R., Wagner, M., Epps, J., Breakspear, M., Parker, G., 2012. From joyous to clinically depressed: mood detection using spontaneous speech. In: Twenty-Fifth International FLAIRS Conference. Marco Island, Florida, pp. 141–146.

Alghowinem, S., Goecke, R., Wagner, M., Epps, J., 2013a. Detecting depression: a comparison between spontaneous and read speech. In: Proceedings of ICASSP. Vancouver, Canada, pp. 7547–7551.

Alghowinem, S., Goecke, R., Wagner, M., Epps, J., 2013b. A comparative of different classifiers for detecting depression from spontaneous speech. In: Proceedings of ICASSP. Vancouver, Canada, pp. 8022–8026.

Alghowinem, S., Goecke, R., Wagner, M., Parker, G., Breakspear, M., 2013c. Eye movement analysis for depression detection. 2013 IEEE International Conference on Image Processing (ICIP2013). IEEE, Melbourne, Australia, pp. 4220–4224.

Alku, P., 1992. Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering. Speech Commun. 11, 109–118.

Alpert, M., Pouget, E.R., Silva, R.R., 2001. Reflections of depression in acoustic measures of the patient's speech. J. Affect. Disord. 66, 59–69.

American-Psychiatric-Association, 2013. Diagnostic and statistical manual of mental disorders? DSM-V., 5th ed. American Psychiatric Association, Washington, DC.

Baddeley, A., 2003. Working memory and language: an overview. J. Commun. Disord. 36, 189–208.

Bagby, R.M., Ryder, A.G., Schuller, D.R., Marshall, M.B., 2004. The Hamilton depression rating scale: has the gold standard become a lead weight? Am. J. Psych. 161, 2163–2177.

Balsters, M.J.H., Krahmer, E.J., Swerts, M.G.J., Vingerhoets, A.J.J.M., 2012. Verbal and nonverbal correlates for depression: a review. Curr. Psych. Rev. 8, 227–234.

Bech, P., Allerup, P., Gram, L.F., Reisby, N., Rosenberg, R., Jacobsen, O., Nagy, A., 1981. The Hamilton depression scale: evaluation of objectivity using logistic models. Acta Psych. Scand. 63, 290–299.

Beck, A.T., Alford, B.A., 2008. Depression: Causes and Treatment. University of Pennsylvania Press.

Beck, A.T., Steer, R.A., 1988. Beck Hopelessness Scale. Psychol. Corp., San Antonio.

Beck, A.T., Epstein, N., Brown, G., Steer, R.A., 1988. An inventory for measuring clinical anxiety: psychometric properties. J. Consult. Clin. Psychol. 56, 893.

Beck, A.T., Steer, R.A., Ball, R., Ranieri, W.F., 1996. Comparison of beck depression inventories-ia and-ii in psychiatric outpatients. J. Pers. Assess. 67, 588–597.

Beck, A.T., Brown, G.K., Steer, R.A., Dahlsgaard, K.K., Grisham, J.R., 1999. Suicide ideation at its worst point: a predictor of eventual suicide in psychiatric outpatients. Suicide Life-Threaten. Behav. 29, 1–9.

Bisconer, S.W., Gross, D.M., 2007. Assessment of suicide risk in a psychiatric hospital. Prof. Psychol. Res. Pract. 38, 143–149.

Bolton, J.M., Pagura, J., Enns, M.W., Grant, B., Sareen, J., 2010. A population-based longitudinal study of risk factors for suicide attempts in major depressive disorder. J. Psych. Res. 44, 817–826.

Bos, E.H., Geerts, E., Bouhuys, A.L., 2002. Non-verbal interaction involvement as an indicator of prognosis in remitted depressed subjects. Psych. Res. 113, 269–277.

Brendel, R.W., Wei, M., Lagomasino, I.T., Perlis, R.H., Stern, T.A., 2010. Care of the suicidal patient. Massachusetts General Hospital Handbook of General Hospital Psychiatry, 6th ed. W.B. Saunders, Saint Louis, pp. 541–554.

Breznitz, Z., 1992. Verbal indicators of depression. J. Gen. Psychol. 119, 351–363.

Brooks, S.J., Nilsson, E.K., Jacobsson, J.A., Stein, D.J., Fredriksson, R., Lind, L., Schiöth, H.B., 2014. BDNF polymorphisms are linked to poorer working memory performance, reduced cerebellar and hippocampal volumes and differences in prefrontal cortex in a Swedish elderly population. PLoS One 9, e82707.

Brown, G.K., Beck, A.T., Steer, R.A., Grisham, J.R., 2000. Risk factors for suicide in psychiatric outpatients: a 20-year prospective study. J. Consult. Clin. Psychol. 68, 371–377.

Brown, T.A., Di Nardo, P.A., Lehman, C.L., Campbell, L.A., 2001. Reliability of DSM-IV anxiety and mood disorders: implications for the classification of emotional disorders. J. Abnorm. Psychol. 110 (1), 49–58.

Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Discov. 2, 121–167.

Buyukdura, J.S., McClintock, S.M., Croarkin, P.E., 2011. Psychomotor retardation in depression: biological underpinnings, measurement, and treatment. Prog. Neuro-Psychopharmacol. Biol. Psych. 35, 395–409.

Calev, A., Nigal, D., Chazan, S., 1989. Retrieval from semantic memory using meaningful and meaningless constructs by depressed, stable bipolar and manic patients. Br. J. Clin. Psychol. 28, 67–73.

Campbell, W.M., Sturim, D.E., Reynolds, D.A., Solomonoff, A., 2006. SVM based speaker verification using a GMM supervector kernel and NAP variability compensation. In: Proceedings of ICASSP, pp. 97–100.

Cannizzaro, M., Harel, B., Reilly, N., Chappell, P., Snyder, P.J., 2004. Voice acoustical measurement of the severity of major depression. Brain Cognit. 56, 30–35.
Cantor, C.H., 2008. Suicide in the western world. The International Handbook of Suicide and Attempted Suicide. John Wiley & Sons, Ltd, pp. 9–28.
Carney, R.M., Freedland, K.E., Veith, R.C., 2005. Depression, the autonomic nervous system, and coronary heart disease. Psychosom. Med. 67, S29–S33.
Chang, C.-C., Lin, C.-J., 2011. LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. 2, 27.
Christopher, G., MacDonald, J., 2005. The impact of clinical depression on working memory. Cognit. Neuropsych. 10, 379–399.
Cochrane-Brink, K.A., Lofchy, J.S., Sakinofsky, I., 2000. Clinical rating scales in suicide risk assessment. Gen. Hosp. Psych. 22, 445–451.
Cohn, J.F., Kruez, T.S., Matthews, I., Ying, Y., Minh Hoai, N., Padilla, M.T., Feng, Z., De la Torre, F., 2009. Detecting depression from facial actions and vocal prosody. 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, 2009, ACII 2009. IEEE, pp. 1–7.
Conway, A.R.A., Kane, M.J., Bunting, M.F., Hambrick, D.Z., Wilhelm, O., Engle, R.W., 2005. Working memory span tasks: a methodological review and user's guide. Psychon. Bull. Rev. 12, 769–786.
Coryell, W., Young, E.A., 2005. Clinical predictors of suicide in primary major depressive disorder. J. Clin. Psych. 66, 412–417.
Costanza, A., D'Orta, I., Perroud, N., Burkhardt, S., Malafosse, A., Mangin, P., Harpe, R., 2014. Neurobiology of suicide: do biomarkers exist? Int. J. Legal Med. 128, 73–82.
Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.G., 2001. Emotion recognition in human–computer interaction. IEEE Signal Process. Magaz. 18, 32–80.
Crawford, T.J., Haeger, B., Kennard, C., Reveley, M.A., Henderson, L., 1995. Saccadic abnormalities in psychotic patients. I: Neuroleptic-free psychotic patients. Psychol. Med. 25, 461–471.
Crawford, A.A., Lewis, S., Nutt, D., Peters, T.J., Cowen, P., O'Donovan, M.C., Wiles, N., Lewis, G., 2014. Adverse effects from antidepressant treatment: randomised controlled trial of 601 depressed individuals. Psychopharmacology (Berl). 231, 2921–2931.
Croarkin, P.E., Levinson, A.J., Daskalakis, Z.J., 2011. Evidence for GABAergic inhibitory deficits in major depressive disorder. Neurosci. Biobehav. Rev. 35, 818–825.
Cull, J.G., Gill, W.S., 1982. Suicide probability scale. In: Western Psychological Services. Western Psychological Services, Los Angeles, CA, pp. 1997–2005.
Cummins, N., Epps, J., Breakspear, M., Goecke, R., 2011. An investigation of depressed speech detection: features and normalization. Proceedings of Interspeech. ISCA, Florence, Italy, pp. 2997–3000.
Cummins, N., Epps, J., Ambikairajah, E., 2013a. Spectro-temporal analysis of speech affected by depression and psychomotor retardation. Proceedings of ICASSP. IEEE, Vancouver, Canada, pp. 7542–7546.
Cummins, N., Epps, J., Sethu, V., Breakspear, M., Goecke, R., 2013b. Modeling spectral variability for the classification of depressed speech. Proceedings of Interspeech. ISCA, Lyon, France, pp. 857–861.
Cummins, N., Joshi, J., Dhall, A., Sethu, V., Goecke, R., Epps, J., 2013c. Diagnosis of depression by behavioural signals: a multimodal approach. In: Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge. Barcelona, Spain, pp. 11–20.
Cummins, N., Epps, J., Sethu, V., Krajewski, J., 2014a. Variability compensation in small data: oversampled extraction of i-vectors for the classification of depressed speech. Proceedings of ICASSP. IEEE, Florence, Italy, pp. 970–974.
Cummins, N., Sethu, V., Epps, J., Krajewski, J., 2014b. Probabilistic acoustic volume analysis for speech affected by depression. Proceedings of Interspeech. ISCA, Singapore, pp. 1238–1242.
Cusin, C., Yang, H., Yeung, A., Fava, M., 2010. Rating scales for depression. In: Baer, L., Blais, M.A. (Eds.), Handbook of Clinical Rating Scales and Assessment in Psychiatry and Mental Health SE – 2, Current Clinical Psychiatry. Humana Press, pp. 7–35.
Darby, J.K., Hollien, H., 1977. Vocal and speech patterns of depressive patients. Folia Phoniatr. (Basel). 29 (4), 279–291.
Darby, J.K., Simmons, N., Berger, P.A., 1984. Speech and voice parameters of depression: a pilot study. J. Commun. Disord. 17, 75–85.
Davidson, R.J., Pizzagalli, D., Nitschke, J.B., Putnam, K., 2002. Depression: perspectives from affective neuroscience. Annu. Rev. Psychol. 53, 545–574.
Deckersbach, T., Dougherty, D.D., Rauch, S.L., 2006. Functional imaging of mood and anxiety disorders. J. Neuroimag. 16, 1–10.
Dehak, N., Kenny, P.J., Dehak, R., Dumouchel, P., Ouellet, P., 2011. Front-end factor analysis for speaker verification. IEEE Trans. Audio. Speech. Lang. Process. 19, 788–798.
De Hert, M., Detraux, J., van Winkel, R., Yu, W., Correll, C.U., 2012. Metabolic and cardiovascular adverse effects associated with antipsychotic drugs. Nat. Rev. Endocrinol. 8, 114–126.
DeVault, D., Georgilia, K., Artstein, R., Morbini, F., Traum, D., Scherer, S., Rizzo, A., Morency, L.P., 2013. Verbal indicators of psychological distress in interactive dialogue with a virtual human. Proceedings of SigDial 2013. Association for Computational Linguistics, pp. 193–202..
DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., Georgila, K., Gratch, J., Hartholt, A., Lhommet, M., Lucas, G., Marsella, S., Morbini, F., Nazarian, A., Scherer, S., Stratou, G., Suri, A., Traum, D., Wood, R., Xu, Y., Rizzo, A., Morency, L.P., 2014. SimSensei kiosk: a virtual human interviewer for healthcare decision support. In: Proceedings of AAMAS '14. International Foundation for Autonomous Agents and Multiagent Systems, Paris, France, pp. 1061–1068.
D'Mello, S., Kory, J., 2012. Consistent but modest: a meta-analysis on unimodal and multimodal affect detection accuracies from 30 studies. In: 14th ACM Internationla Conference on Multimodal Interaction (ICMI '12), pp. 31–38.
Domenici, E., Willé, D.R., Tozzi, F., Prokopenko, I., Miller, S., McKeown, A., Brittain, C., Rujescu, D., Giegling, I., Turck, C.W., Holsboer, F., Bullmore, E.T., Middleton, L., Merlo-Pich, E., Alexander, R.C., Muglia, P., 2010. Plasma protein biomarkers for depression and schizophrenia by multi analyte profiling of case-control collections. PLoS One 5 (2), e9166.
Douglas-Cowie, E., Campbell, N., Cowie, R., Roach, P., 2003. Emotional speech: towards a new generation of databases. Speech Commun. 40, 33–60.
Doval, B., d'Alessandro, C., Henrich, N., 2006. The spectrum of glottal flow models. Acta Acust. united with Acust. 92, 1026–1046.
Drugman, T., Bozkurt, B., Dutoit, T., 2010. A comparative study of glottal source estimation techniques. Comput. Speech Lang. 26, 20–34.
Dwivedi, Y., Rizavi, H.S., Conley, R.R., Roberts, R.C., Tamminga, C.A., Pandey, G.N., 2003. Altered gene expression of brain-derived neurotrophic factor and receptor tyrosine kinase B in postmortem brain of suicide subjects. Arch. Gen. Psych. 60, 804–815.
Ekman, P., Freisen, W.V., Ancoli, S., 1980. Facial signs of emotional experience. J. Pers. Soc. Psychol. 39, 1125–1134.
Ellgring, H., Scherer, K., 1996. Vocal indicators of mood change in depression. J. Nonverbal Behav. 20, 83–110.
Enge, S., Fleischhauer, M., Lesch, K.-P., Reif, A., Strobel, A., 2011. Serotonergic modulation in executive functioning: linking genetic variations to working memory performance. Neuropsychologia 49, 3776–3785.
Evans, K.C., Dougherty, D.D., Pollack, M.H., Rauch, S.L., 2006. Using neuroimaging to predict treatment response in mood and anxiety disorders. Am. Acad. Clin. Psychiatr. 18, 33–42.
Faries, D., Herrera, J., Rayamajhi, J., DeBrota, D., Demitrack, M., Potter, W.Z., 2000. The responsiveness of the Hamilton depression rating scale. J. Psych. Res. 34, 3–10.
Flint, A.J., Black, S.E., Campbell-Taylor, I., Gailey, G.F., Levinton, C., 1993. Abnormal speech articulation, psychomotor retardation, and subcortical dysfunction in major depression. J. Psych. Res. 27, 309–319.

Florentine, J.B., Crane, C., 2010. Suicide prevention by limiting access to methods: a review of theory and practice. Soc. Sci. Med. 70, 1626–1632.
Fragopanagos, N., Taylor, J.G., 2005. Emotion recognition in human–computer interaction. Neural Netw. 18, 389–405.
France, D.J., Shiavi, R.G., Silverman, S., Silverman, M., Wilkes, M., 2000. Acoustical properties of speech as indicators of depression and suicidal risk. IEEE Trans. Bio-Eng. 47, 829–837.
Frodl, T., Schüle, C., Schmitt, G., Al, E., 2007. Association of the brain-derived neurotrophic factor val66met polymorphism with reduced hippocampal volumes in major depression. Arch. Gen. Psych. 64, 410–416.
Gaebel, W., Wölwer, W., 2004. Facial expressivity in the course of schizophrenia and depression. Eur. Arch. Psych. Clin. Neurosci. 254, 335–342.
Gatt, J.M., Nemeroff, C.B., Dobson-Stone, C., Paul, R.H., Bryant, R.A., Schofield, P.R., Gordon, E., Kemp, A.H., Williams, L.M., 2009. Interactions between BDNF Val66Met polymorphism and early life stress predict brain and arousal pathways to syndromal depression and anxiety. Mol. Psych. 14, 681–695.
Gibbons, R.D., Clark, D.C., Kupfer, D.J., 1993. Exactly what does the Hamilton depression rating scale measure? J. Psych. Res. 27, 259–273.
Girard, J.M., Cohn, J.F., Mahoor, M.H., Mavadati, S.M., Hammal, Z., Rosenwald, D.P., 2013. Nonverbal social withdrawal in depression: evidence from manual and automatic analyses. Image Vis. Comput. 32, 641–647.
Gobl, C., Ní Chasaide, A., 2003. The role of voice quality in communicating emotion, mood and attitude. Speech Commun. 40, 189–212.
Godfrey, H.P., Knight, R.G., 1984. The validity of actometer and speech activity measures in the assessment of depressed patients. Br. J. Psych. 145, 159–163.
Goeleven, E., De Raedt, R., Baert, S., Koster, E.H.W., 2006. Deficient inhibition of emotional information in depression. J. Affect. Disord. 93, 149–157.
Goldney, R.D., 2008. Prediction of suicide and attempted suicide. In: Hawton, K., van Heeringen, K. (Eds.), The International Handbook of Suicide and Attempted Suicide. John Wiley & Sons, Ltd, Ltd, West Sussex, England, pp. 585–595.
Gratch, J., Lucas, G.M., King, A.A., Morency, L.P., 2014. It is only a computer: the impact of human-agent interaction in clinical interviews. In: Proceedings of the 13th Annual Conference on Autonomous Agents and Multi-Agent Systems, pp. 85–92.
Greden, J.F., Carroll, B.J., 1980. Decrease in speech pause times with treatment of endogenous depression. Biol. Psych. 15, 575–587.
Greden, J.F., Albala, A.A., Smokler, I.A., 1981. Speech pause time: a marker of psychomotor retardation among endogenous depressives. Biol. Psych. 16, 851–859.
Gupta, R., Malandrakis, N., Xiao, B., Guha, T., Van Segbroeck, M., Black, M., Potamianos, A., Narayanan, S., 2014. Multimodal prediction of affective dimensions and depression in human–computer interactions. Proceedings of the 4th ACM International Workshop on Audio/Visual Emotion Challenge (AVEC '14). ACM, Orlando, Florida, USA, pp. 33–40.
Hall, J.A., Harrigan, J.A., Rosenthal, R., 1995. Nonverbal behavior in clinician—patient interaction. Appl. Prev. Psychol. 4, 21–37.
Hamilton, H., 1960. HAMD: a rating scale for depression. Neurosurg. Psych. 23, 56–62.
Hardy, P., Jouvent, R., Widlöcher, D., 1984. Speech pause time and the retardation rating scale for depression (ERD): towards a reciprocal validation. J. Affect. Disord. 6, 123–127.
Harrigan, J.A., Wilson, K., Rosenthal, R., 2004. Detecting state and trait anxiety from auditory and visual cues: a meta-analysis. Personal. Soc. Psychol. Bull. 30, 56–66.
Hart, J., Gratch, J., Marsella, S., 2013. How virtual reality training can win friends and influence people. In: Best, C., Galanis, G., Kerry, J., Sottilare, R. (Eds.), Human Factors in Defence. Ashgate, pp. 235–249.
Hashim, N.W., Wilkes, M., Salomon, R., Meggs, J., 2012. Analysis of timing pattern of speech as possible indicator for near-term suicidal risk and depression in male patients. In: 2012 International Conference on Signal Processing Systems (ICSPS 2012), pp. 6–13.
Hasler, G., Northoff, G., 2011. Discovering imaging endophenotypes for major depression. Mol Psych. 16, 604–619.
Hasler, G., Drevets, W.C., Manji, H.K., Charney, D.S., 2004. Discovering endophenotypes for major depression. Neuropsychopharmacology 29, 1765–1781.
Hawton, K., van Heeringen, K., 2009. Suicide. Lancet 373, 1372–1381.
Hawton, K., Casañas I Comabella, C., Haw, C., Saunders, K., 2013. Risk factors for suicide in individuals with depression: a systematic review. J. Affect. Disord. 147, 17–28.
Helfer, B.S., Quatieri, T.F., Williamson, J.R., Mehta, D.D., Horwitz, R., Yu, B., 2013. Classification of depression state based on articulatory precision. Proceedings of Interspeech. ISCA, Lyon, France, pp. 2172–2176.
Heman-Ackah, Y.D., Heuer, R.J., Michael, D.D., Ostrowski, R., Horman, M., Baroody, M.M., Hillenbrand, J., Sataloff, R.T., 2003. Cepstral peak prominence: a more reliable measure of dysphonia. Ann. Otol. Rhinol. Laryngol. 112, 324–333.
Hendin, H., Maltsberger, J.T., Lipschitz, A., Haas, A.P., Kyle, J., 2001. Recognizing and responding to a suicide crisis. Ann. NY Acad. Sci. 932, 169–187.
Hendin, H., Maltsberger, J.T., Szanto, K., 2007. The role of intense affective states in signaling a suicide crisis. J. Nerv. Ment. Dis. 195, 363–368.
Hollien, H., 1980. Vocal indicators of psychological stress. Ann. NY Acad. Sci. 347, 47–72.
Hönig, F., Batliner, A., Nöth, E., Schnieder, S., Krajewski, J., 2014. Automatic modelling of depressed speech: relevant features and relevance of gender. In: Proceedings of Interspeech. Singapore, pp. 1248–1252.
Horwitz, R., Quatieri, T.F., Helfer, B.S., Yu, B., Williamson, J.R., Mundt, J., 2013. On the relative importance of vocal source, system, and prosody in human depression. in: IEEE International Conference on Body Sensor Networks (BSN), 2013. Cambridge, MA, USA, pp. 1–6.
Joiner, T.E., Brown, J.S., Wingate, L.R., 2005. The psychology and neurobiology of suicidal behavior. Ann. Rev. Psychol. 56, 287–314.
Jokinen, J., Mårtensson, B., Nordström, A.-L., Nordström, P., 2008. CSF 5-HIAA and DST non-suppression – independent biomarkers in suicide attempters? J. Affect. Disord. 105, 241–245.
Joormann, J., Gotlib, I.H., 2008. Updating the contents of working memory in depression: interference from irrelevant negative material. J. Abnorm. Psychol. 117, 182–192.
Joshi, J., Goecke, R., Alghowinem, S., Dhall, A., Wagner, M., Epps, J., Parker, G., Breakspear, M., 2013. Multimodal assistive technologies for depression diagnosis and monitoring. J. Multimodal User Interf. 7, 217–228.
Kamphuis, J.H., Noordhof, A., 2009. On categorical diagnoses in DSM-V: cutting dimensions at useful points? Psychol. Assess. 21 (September), 294–301.
Kane, J., Gobl, C., 2011. Identifying regions of non-modal phonation using features of the wavelet transform. In: Proceedings of Interspeech. Florence, Italy, pp. 177–180.
Kane, J., Yanushevskaya, I., Dalton, J., Gobl, C., Chasaide, A.N., 2013. Using phonetic feature extraction to determine optimal speech regions for maximising the effectiveness of glottal source analysis. Proceedings of Interspeech. ISCA, Portland, USA, pp. 29–33.
Kane, J., Aylett, M., Yanushevskaya, I., Gobl, C., 2014. Phonetic feature extraction for context-sensitive glottal source processing. Speech Commun. 59, 10–21.
Kaya, H., Salah, A.A., 2014. Eyes whisper depression: a CCA based multimodal approach. In: Proceedings of the ACM International Conference on Multimedia - MM '14. ACM Press, New York, USA, pp. 961–964.
Kaya, H., Çilli, F., Salah, A., 2014a. Ensemble CCA for continuous emotion prediction. Proceedings of the 4th ACM International Workshop on Audio/Visual Emotion Challenge (AVEC '14). ACM, Orlando, Florida, USA, pp. 19–26.

Kaya, H., Eyben, F., Salah, A.A., 2014b. CCA based feature selection with application to continuous depression recognition from acoustic speech features. In: Proceedings of ICASSP. Florence, Italy, pp. 3757–3761.

Kemp, A.H., Gordon, E., Rush, A.J., Williams, L.M., 2008. Improving the prediction of treatment response in depression: integration of clinical, cognitive, psychophysiological, neuroimaging, and genetic measures. CNS Spectr. 13, 1066–1068.

Kenny, P., Ouellet, P., Dehak, N., Gupta, V., Dumouchel, P., 2008. A study of inter-speaker variability in speaker verification. IEEE Trans. Audio, Speech, Lang. Process. 16, 980–988.

Kent, R.D., 2000. Research on speech motor control and its disorders: a review and prospective. J. Commun. Disord. 33, 391–428.

Keskinpala, H.K., Yingthawornsuk, T., Wilkes, D.M., Shiavi, R.G., Salomon, R.M., 2007. Screening for high risk suicidal states using mel-cepstral coefficients and energy in frequency bands. Eur. Signal Process. Conf. Pozn. Pol., 2229–2233

Kessler, R.C., Berglund, P., Nock, M., Wang, P.S., 2005. Trends in suicide ideation, plans, gestures, and attempts in the United States, 1990–1992 to 2001–2003. JAMA 293, 2487–2495.

Kikuchi, T., Suzuki, T., Uchida, H., Watanabe, K., Mimura, M., 2012. Coping strategies for antidepressant side effects: an Internet survey. J. Affect. Disord. 143, 89–94.

Kim, J., 2007. Emotion recognition using speech and physiological changes. Robust Speech Recognition and Understanding. I-Tech Education and Publishing, pp. 265–280.

Kim, E., Ku, J., Kim, J.-J., Lee, H., Han, K., Kim, S.I., Cho, H.-S., 2009. Nonverbal social behaviors of patients with bipolar mania during interactions with virtual humans. J. Nerv. Ment. Dis. 197, 412–418.

Kinnunen, K., Li, H., 2009. An overview of text-independent speaker recognition: from features to supervectors. Elsevier Speech Commun. 52, 12–40.

Kinnunen, T., Lee, K.A., Li, H., 2008. Dimension reduction of the modulation spectrogram for speaker verification. Proceedings of Speaker Odyssey. ISCA, Stellenbosch, South Africa, p. 30.

Klatt, D.H., Klatt, L.C., 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. J. Acoust. Soc. Am. 87, 820–857.

Kotov, R., Gamez, W., Schmidt, F., Watson, D., 2010. Linking "big" personality traits to anxiety, depressive, and substance use disorders: a meta-analysis. Psychol. Bull. 136, 768–821.

Kraepelin, E., 1921. Manic depressive insanity and paranoia. J. Nerv. Ment. Dis. 53, 350.

Krajewski, J., Schnieder, S., Sommer, D., Batliner, A., Schuller, B., 2012. Applying multiple classifiers and non-linear dynamics features for detecting sleepiness from speech. Neurocomputing 84, 65–75.

Kreibig, S.D., 2010. Autonomic nervous system activity in emotion: a review. Biol. Psychol. 84, 394–421.

Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9: validity of a brief depression severity measure. J. Gen. Int. Med. 16, 606–613.

Ku, J., Jang, H., Kim, K., Park, S., Kim, J., Kim, C., SW, N., Kim, I., Kim, S., 2006. Pilot study for assessing the behaviors of patients with schizophrenia towards a virtual avatar. CyberPsychol. Behav. 9, 531–539.

Kuny, S., Stassen, H.H., 1993. Speaking behavior and voice sound characteristics in depressive patients during recovery. J. Psych. Res. 27, 289–307.

Landau, M., Yingthawornsuk, T., Wilkes, D.M., Shiavi, R.G., Salomon, R.M., 2007. Predicting severity of mental state using vocal output characteristics. MAVEBA-2007, 153–156.

Lang, P.J., Bradley, M.M., Cuthbert, B.N., 2005. International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual. NIMH, Center for the Study of Emotion & Attention.

Laukkanen, A.-M., Björkner, E., Sundberg, J., 2006. Throaty voice quality: subglottal pressure, voice source, and formant characteristics. J. Voice 20, 25–37.

Laver, J., Hiller, S., Beck, J.M., 1992. Acoustic waveform perturbations and voice disorders. J. Voice 6, 115–126.

Le-Niculescu, H., Levey, D.F., Ayalew, M., Palmer, L., Gavrin, L.M., Jain, N., Winiger, E., Bhosrekar, S., Shankar, G., Radel, M., Bellanger, E., Duckworth, H., Olesek, K., Vergo, J., Schweitzer, R., Yard, M., Ballew, A., Shekhar, A., Sandusky, G.E., Schork, N.J., Kurian, S.M., Salomon, D.R., Niculescu, A.B., 2013. Discovery and validation of blood biomarkers for suicidality. Mol. Psych. 18, 1249–1264.

Lépine, J.-P., Briley, M., 2011. The increasing burden of depression. Neuropsych. Dis. Treat. 7, 3–7.

Levelt, W.J., Roelofs, a., Meyer, A.S., 1999. A theory of lexical access in speech production. Behav. Brain Sci. 22, 1–38 (discussion 38–75).

Li, M., Metallinou, A., Bone, D., Narayanan, S., 2012. Speaker states recognition using latent factor analysis based Eigenchannel factor vector modeling. Proceedings of ICASSP. IEEE, Kyoto, Japan, pp. 1937–1940.

Li, M., Han, K.J., Narayanan, S., 2013. Automatic speaker age and gender recognition using acoustic and prosodic level information fusion. Comput. Speech Lang. 27, 151–167.

Linehan, M.M., Goodstein, J.L., Nielsen, S.L., Chiles, J.A., 1983. Reasons for staying alive when you are thinking of killing yourself: the reasons for living inventory. J. Consult. Clin. Psychol. 51, 276.

Lipton, R.B., Levin, S., Holzman, P.S., 1980. Horizontal and vertical pursuit eye movements, the oculocephalic reflex, and the functional psychoses. Psych. Res. 3, 193–203.

Low, L.S.A., Maddage, N.C., Lech, M., Allen, N., 2009. Mel frequency cepstral feature and Gaussian Mixtures for modeling clinical depression in adolescents. In: 8th IEEE International Conference on Cognitive Informatics, 2009 (ICCI '09), pp. 346–350.

Low, L.S.A., Maddage, N.C., Lech, M., Sheeber, L., Allen, N., 2010. Influence of acoustic low-level descriptors in the detection of clinical depression in adolescents. Proceedings of ICASSP. IEEE, Dallas, Texas, pp. 5154–5157.

Low, L.S.A., Maddage, M.C., Lech, M., Sheeber, L.B., Allen, N.B., 2011. Detection of clinical depression in adolescents; speech during family interactions. IEEE Trans. Biomed. Eng. 58, 574–586.

Luscher, B., Shen, Q., Sahir, N., 2011. The GABAergic deficit hypothesis of major depressive disorder. Mol. Psych. 16, 383–406.

Lux, V., Kendler, K.S., 2010. Deconstructing major depression: a validation study of the DSM-IV symptomatic criteria. Psychol. Med. 40, 1679–1690.

MacQueen, G., Frodl, T., 2011. The hippocampus in major depression: evidence for the convergence of the bench and bedside in psychiatric research? Mol. Psych. 16, 252–264.

Maddage, N.C., Senaratne, R., Low, L.S.A., Lech, M., Allen, N., 2009. Video-based detection of the clinical depression in adolescents. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009, EMBC 2009, pp. 3723–3726.

Mandrusiak, M., Rudd, M.D., Joiner, T.E., Berman, A.L., Van Orden, K.A., Witte, T., 2006. Warning signs for suicide on the internet: a descriptive study. Suicide Life-Threaten. Behav. 36, 263–271.

Manea, L., Gilbody, S., McMillan, D., 2012. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. CMAJ 184, E191–E196.

Mann, J.J., 2003. Neurobiology of suicidal behaviour. Nat. Rev. Neurosci. 4, 819–828.

Mann, J., Apter, A., Bertolote, J., 2005. Suicide prevention strategies: a systematic review. JAMA J. Am. Med. Assoc. 294, 2064–2074.

Maragos, P., Kaiser, J.F., Quatieri, T.F., 1993. Energy separation in signal modulations with application to speech analysis. IEEE Trans. Signal Process. 41, 3024–3051.

Maris, R.W., 2002. Suicide. Lancet 360, 319–326.

Mathers, C.D., Loncar, D., 2006. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med. 3, 2011–2030.

Mattisson, C., Bogren, M., Horstmann, V., Munk-Jörgensen, P., Nettelbladt, P., 2007. The long-term course of depressive disorders in the Lundby study. Psychol. Med. 37, 883–891.

Maust, D., Cristancho, M., Gray, L., Rushing, S., Tjoa, C., Thase, M.E., 2012. Chapter 13 – psychiatric rating scales. In: Michael, J., Aminoff,

F.B., Dick, F.S. (Eds.), Handbook of Clinical Neurology. Elsevier, pp. 227–237.
Mayberg, H.S., Lozano, A.M., Voon, V., McNeely, H.E., Seminowicz, D., Hamani, C., Schwalb, J.M., Kennedy, S.H., 2005. Deep brain stimulation for treatment-resistant depression. Neuron 45, 651–660.
McAllister, A., Sederholm, E., Ternström, S., Sundberg, J., 1996. Perturbation and hoarseness: a pilot study of six children's voices. J. Voice 10, 252–261.
McGirr, A., Renaud, J., Seguin, M., Alda, M., Benkelfat, C., Lesage, A., Turecki, G., 2007. An examination of DSM-IV depressive symptoms and risk for suicide completion in major depressive disorder: a psychological autopsy study. J. Affect. Disord. 97, 203–209.
McIntosh, J.L., 2009. USA Suicide 2006: Official final data. [WWW Document]. Am. Assoc. Suicidol <http://www.suicidology.org> (accessed 28.02.14).
McIntyre, G., Goecke, R., Hyett, M., Green, M., Breakspear, M., 2009. An approach for automatically measuring facial activity in depressed subjects. In: 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, 2009 (ACII '09), pp. 1–8.
Memon, S., Maddage, N., Lech, M., Allen, N., 2009. Effect of clinical depression on automatic speaker identification. In: 3rd International Conference on Bioinformatics and Biomedical Engineering, 2009 (ICBBE '09), pp. 1–4.
Mitchell, A.J., Vaze, A., Rao, S., 2009. Clinical diagnosis of depression in primary care: a meta-analysis. Lancet 374, 609–619.
Mitra, V., Shriberg, E., McLaren, M., Kathol, A., Richey, C., Vergyri, D., Graciarena, M., 2014. The SRI AVEC-2014 evaluation system. Proceedings of the 4th ACM International Workshop on Audio/Visual Emotion Challenge (AVEC '14). ACM, Orlando, Florida, USA, pp. 93–101.
Montgomery, S.A., Asberg, M., 1979. A new depression scale designed to be sensitive to change. Br. J. Psych. 134, 382–389.
Moore, E., Clements, M., Peifer, J., Weisser, L., 2003. Analysis of prosodic variation in speech for clinical depression. In: Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 2925–2928.
Moore, E., Clements, M., Peifer, J., Weisser, L., 2004. Comparing objective feature statistics of speech for classifying clinical depression. In: 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2004, IEMBS '04, pp. 17–20.
Moore, E., Clements, M.A., Peifer, J.W., Weisser, L., 2008. Critical analysis of the impact of glottal features in the classification of clinical depression in speech. IEEE Trans. Biomed. Eng. 55, 96–107.
Mundt, J.C., Snyder, P.J., Cannizzaro, M.S., Chappie, K., Geralts, D.S., 2007. Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. J. Neuroling. 20, 50–64.
Mundt, J.C., Vogel, A.P., Feltner, D.E., Lenderking, W.R., 2012. Vocal acoustic biomarkers of depression severity and treatment response. Biol. Psych. 72, 580–587.
Murphy, F.C., Sahakian, B.J., Rubinsztein, J.S., Michael, A., Rogers, R.D., Robbins, T.W., Paykel, E.S., 1999. Emotional bias and inhibitory control processes in mania and depression. Psychol. Med. 29, 1307–1321.
Navarro, J., del Moral, R., Alonso, M.F., Loste, P., Garcia-Campayo, J., Lahoz-Beltra, R., Marijuán, P.C., 2014. Validation of laughter for diagnosis and evaluation of depression. J. Affect. Disord. 160, 43–49.
Nestler, E.J., Barrot, M., DiLeone, R.J., Eisch, A.J., Gold, S.J., Monteggia, L.M., 2002. Neurobiology of depression. Neuron 34, 13–25.
Niemiec, A.J., Lithgow, B.J., 2005. Alpha-band characteristics in EEG spectrum indicate reliability of frontal brain asymmetry measures in diagnosis of depression. In: 27th Annual International Conference of the Engineering in Medicine and Biology Society, 2005, IEEE-EMBS 2005, pp. 7517–7520.
Nilsonne, A., 1987. Acoustic analysis of speech variables during depression and after improvement. Acta Psych. Scand. 76, 235–245.
Nilsonne, A., 1988. Speech characteristics as indicators of depressive illness. Acta Psych. Scand. 77, 253–263.
Nilsonne, A., Sundberg, J., 1985. Differences in ability of musicians and nonmusicians to judge emotional state from the fundamental frequency of voice samples. Music Percept. 2, 507–516.
Nilsonne, A., Sundberg, J., Ternstrom, S., Askenfelt, A., 1987. Measuring the rate of change of voice fundamental frequency in fluent speech during mental depression. J. Acoust. Soc. Am. 83, 716–728.
Nock, M.K., Borges, G., Bromet, E.J., Cha, C.B., Kessler, R.C., Lee, S., 2008. Suicide and suicidal behavior. Epidemiol. Rev. 30, 133–154.
Nuevo, R., Lehtinen, V., Reyna-Liberato, P.M., Ayuso-Mateos, J.L., 2009. Usefulness of the Beck Depression Inventory as a screening method for depression among the general population of Finland. Scand. J. Public Heal. 37, 28–34.
Oberauer, K., 2001. Removing irrelevant information from working memory: a cognitive aging study with the modified Sternberg task. J. Exp. Psychol. Learn. Mem. Cogn. 27, 948.
Olesen, J., Gustavsson, A., Svensson, M., Wittchen, H.-U., Jönsson, B., 2012. The economic cost of brain disorders in Europe. Eur. J. Neurol. 19, 155–162.
Ooi, K.E.B., Low, L.S.A., Lech, M., Allen, N., 2012. Early prediction of major depression in adolescents using glottal wave characteristics and Teager Energy parameters. Proceedings of ICASSP. IEEE, Kyoto, Japan, pp. 4613–4616.
Ooi, K.E.B., Lech, M., Allen, N.B., 2013. Multichannel weighted speech classification system for prediction of major depression in adolescents. IEEE Trans. Biomed. Eng. 60, 497–506.
Oquendo, M.A.M.D., Baca-Garcia, E.M.D.P.D., Mann, J.J.M.D., Giner, J.M.D.P.D., 2008. Issues for DSM-V: suicidal behavior as a separate diagnosis on a separate axis. Am. J. Psych. 165, 1383–1384.
Orlikoff, R.F., Kahane, J.C., 1991. Influence of mean sound pressure level on jitter and shimmer measures. J. Voice 5, 113–119.
O'Shaughnessy, D., 1999. Speech Communications: Human and Machine. Institute of Electrical and Electronics Engineers, New York.
Østergaard, S.D., Jensen, S.O.W., Bech, P., 2011. The heterogeneity of the depressive syndrome: when numbers get serious. Acta Psych. Scand. 124, 495–496.
Owens, M., Herbert, J., Jones, P.B., Sahakian, B.J., Wilkinson, P.O., Dunn, V.J., Croudace, T.J., Goodyer, I.M., 2014. Elevated morning cortisol is a stratified population-level biomarker for major depression in boys only with high depressive symptoms. Proc. Natl. Acad. Sci. 111, 3638–3643.
Ozdas, A., Shiavi, R.G., Silverman, S.E., Silverman, M.K., Wilkes, D.M., 2000. Analysis of fundamental frequency for near term suicidal risk assessment. In: 2000 IEEE International Conference on Systems, Man, and Cybernetics, pp. 1853–1858.
Ozdas, A., Shiavi, R.G., Silverman, S.E., Silverman, M.K., Wilkes, D.M., 2004a. Investigation of vocal jitter and glottal flow spectrum as possible cues for depression and near-term suicidal risk. IEEE Trans. Bio-Eng. 51, 1530–1540.
Ozdas, A., Shiavi, R.G., Wilkes, D.M., Silverman, M.K., Silverman, S.E., 2004b. Analysis of vocal tract characteristics for near-term suicidal risk assessment. Methods Inf. Med. 43, 36–38.
Parker, G., Hadzi-Pavlovic, D., 1996. Melancholia: A Disorder of Movement and Mood: A Phenomenological and Neurobiological Review. Cambridge University Press.
Parker, G., Hadzi-Pavlovic, D., Boyce, P., Wilhelm, K., Brodaty, H., Mitchell, P., Hickie, I., Eyers, K., 1990. Classifying depression by mental state signs. Br. J. Psych. 157, 55–65.
Pelecanos, J., Sridharan, S., 2001. Feature warping for robust speaker verification. Proceedings of Speaker Odyssey. ICSA, Crete, Greece, pp. 213–218.
Pérez, H., Escalante, H.J., Villaseñor-Pineda, L., Montes-y-Gómez, M., Pinto-Avedaño, D., Reyes-Meza, V., 2014. Fusing affective dimensions and audio-visual features from segmented video for depression recognition. Proceedings of the 4th ACM International Workshop on Audio/Visual Emotion Challenge (AVEC '14). ACM, Orlando, Florida, USA, pp. 49–55.

Petrushin, V., 1999. Emotion in speech: recognition and application to call centers. In: Artificial Neural Networks in Engineering (ANNIE '99). St. Louis, MO, USA, pp. 7–10.
Postma, A., 2000. Detection of errors during speech production: a review of speech monitoring models. Cognition 77, 97–132.
Poulter, M.O., Du, L., Weaver, I.C.G., Palkovits, M., Faludi, G., Merali, Z., Szyf, M., Anisman, H., 2008. GABAA receptor promoter hypermethylation in suicide brain: implications for the involvement of epigenetic processes. Biol. Psych. 64, 645–652.
Prinstein, M.J., 2008. Introduction to the special section on suicide and nonsuicidal self-injury: a review of unique challenges and important directions for self-injury science. J. Consult. Clin. Psychol. 76, 1–8.
Quatieri, T.F., 2001. Discrete-Time Speech-Signal Processing: Principles and Practice. Prentice Hall, Upper Saddle River, NJ, 07458.
Quatieri, T.F., Malyska, N., 2012. Vocal-source biomarkers for depression: a link to psychomotor activity. Proceedings of Interspeech. ICSA, Portland, USA, pp. 1059–1062.
Raust, A., Slama, F., Mathieu, F., Roy, I., Chenu, A., Koncke, D., Fouques, D., Jollant, F., Jouvent, E., Courtet, P., Leboyer, M., Bellivier, F., 2007. Prefrontal cortex dysfunction in patients with suicidal behavior. Psychol. Med. 37, 411–419.
Reynolds, D.A., Rose, R.C., 1995. Robust text-independent speaker identification using Gaussian mixture speaker models. IEEE Trans. Speech Audio Process 3, 72–83.
Reynolds, D.A., Quatieri, T.F., Dunn, R.B., 2000. Speaker verification using adapted Gaussian mixture models. Digit. Signal Process. 10, 19–41.
Ring, H.A., 2002. Neuropsychiatry of the basal ganglia. J. Neurol. Neurosurg. Psych. 72, 12–21.
Robins, B., Dautenhahn, K., Te Boekhorst, R., Billard, A., 2005. Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills? Univers. Access Inf. Soc. 4, 105–120.
Roy, A., Nielsen, D., Rylander, G., Sarchiapone, M., 2008. The genetics of suicidal behaviour. The International Handbook of Suicide and Attempted Suicide. John Wiley & Sons, Ltd., pp. 209–221.
Roy, N., Nissen, S.L., Dromey, C., Sapir, S., 2009. Articulatory changes in muscle tension dysphonia: evidence of vowel space expansion following manual circumlaryngeal therapy. J. Commun. Disord. 42, 124–135.
Rudd, M.D., Berman, A.L., Joiner, T.E., Nock, M.K., Silverman, M.M., Mandrusiak, M., Van Orden, K., Witte, T., 2006. Warning signs for suicide: theory, research, and clinical applications. Suicide Life-Threaten. Behav. 36, 255–262.
Rush, J.A., Trivedi, M.H., Ibrahim, H.M., Carmody, T.J., Arnow, B., Klein, D.N., Markowitz, J.C., Ninan, P.T., Kornstein, S., Manber, R., 2003. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biol. Psych. 54, 573–583.
Ryding, E., Ahnlide, J.-A., Lindström, M., Rosén, I., Träskman-Bendz, L., 2006. Regional brain serotonin and dopamine transporter binding capacity in suicide attempters relate to impulsiveness and mental energy. Psych. Res. 148, 195–203.
Schelde, J.T.M., 1998. Major depression: behavioral markers of depression and recovery. J. Nerv. Ment. Dis. 186, 133–140.
Scherer, K.R., 1986. Vocal affect expressions – a review and a model for future-research. Psychol. Bull. 99, 143–165.
Scherer, S., Pestian, J., Morency, L.P., 2013a. Investigating the speech characteristics of suicidal adolescents. In: IEEE (Ed.), Proceedings of ICASSP. Vancouver, Canada, pp. 709–713.
Scherer, S., Stratou, G., Gratch, J., Morency, L., 2013b. investigating voice quality as a speaker-independent indicator of depression and PTSD. Proceedings of Interspeech. ISCA, Lyon, France, pp. 847–851.
Scherer, S., Stratou, G., Mahmoud, M., Boberg, J., Gratch, J., 2013c. Automatic behavior descriptors for psychological disorder analysis. 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2013. IEEE, Shanghai, pp. 1–8.
Scherer, S., Stratou, G., Morency, L.P., 2013d. Audiovisual behavior descriptors for depression assessment. In: Proceedings of the 15th ACM on International Conference on Multimodal Interaction (ICMI). Sydney, Australia, pp. 135–140.
Schmidt, H.D., Shelton, R.C., Duman, R.S., 2011. Functional biomarkers of depression: diagnosis, treatment, and pathophysiology. Neuropsychopharmacology 36, 2375–2394.
Schneider, D., Regenbogen, C., Kellermann, T., Finkelmeyer, A., Kohn, N., Derntl, B., Schneider, F., Habel, U., 2012. Empathic behavioral and physiological responses to dynamic stimuli in depression. Psych. Res. 200, 294–305.
Schuller, B., Wimmer, M., Mosenlechner, L., Kern, C., Arsic, D., Rigoll, G., Lorenz, M., 2008. Brute-forcing hierarchical functionals for paralinguistics: a waste of feature space? Proceedings of ICASSP. IEEE, Las Vegas, USA, pp. 4501–4504.
Schuller, B., Batliner, A., Steidl, S., Seppi, D., 2011. Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. Speech Commun. 53, 1062–1087.
Schuller, B., Steidl, S., Batliner, A., Burkhardt, F., Devillers, L., Müller, C., Narayanan, S., 2013. Paralinguistics in speech and language; State-of-the-art and the challenge. Comput. Speech Lang. 27, 4–39.
Schumann, I., Schneider, A., Kantert, C., Löwe, B., Linde, K., 2012. Physicians' attitudes, diagnostic process and barriers regarding depression diagnosis in primary care: a systematic review of qualitative studies. Fam. Pract. 29, 255–263.
Segrin, C., 2000. Social skills deficits associated with depression. Clin. Psychol. Rev. 20, 379–403.
Senoussaoui, M., Sarria-Paja, M., Santos, J.F., Falk, T.H., 2014. Model fusion for multimodal depression classification and level detection. Proceedings of the 4th ACM International Workshop on Audio/Visual Emotion Challenge (AVEC '14). ACM, Orlando, Florida, USA, pp. 57–63.
Sequeira, A., Klempan, T., Canetti, L., 2007. Patterns of gene expression in the limbic system of suicides with and without major depression. Mol. Psych. 12, 640–655.
Sethu, V., Ambikairajah, E., Epps, J., 2008. Phonetic and speaker variations in automatic emotion classification. Proceedings of Interspeech. ICSA, Brisbane, Australia, pp. 617–620.
Sethu, V., Epps, J., Ambikairajah, E., 2013. Speaker variability in speech based emotion models – analysis and normalisation. Proceedings of ICASSP. IEEE, Vancouver, Canada, pp. 7522–7526.
Sethu, V., Epps, J., Ambikairajah, E., 2014. Speech based emotion recognition. In: Ogunfunmi, T., Togneri, R., Narasimhai, M. (Eds.), Speech and Audio Processing for Coding Enhancement and Recognition. Springer, New York, pp. 197–228.
Sharp, T., Cowen, P.J., 2011. 5-HT and depression: is the glass half-full? Curr. Opin. Pharmacol. 11, 45–51.
Sheline, Y.I., 2003. Neuroimaging studies of mood disorder effects on the brain. Biol. Psych. 54, 338–352.
Sidorov, M., Minker, W., 2014. Emotion recognition and depression diagnosis by acoustic and visual features: a multimodal approach. Proceedings of the 4th ACM International Workshop on Audio/Visual Emotion Challenge (AVEC '14). ACM, Orlando, Florida, USA, pp. 81–86.
Sikorski, C., Luppa, M., König, H.-H., van den Bussche, H., Riedel-Heller, S.G., 2012. Does GP training in depression care affect patient outcome? – a systematic review and meta-analysis. BMC Health Serv. Res. 12, 10.
Silverman, S.E., 1992. Vocal parameters as predictors of near-term suicidal risk.
Silverman, S.E., Silverman, M.K., 2006. Methods and apparatus for evaluating near-term suicidal risk using vocal parameters.
Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14, 199–222.
Sobin, C., Sackeim, H.A., 1997. Psychomotor symptoms of depression. Am. J. Psych. 154, 4–17.
Stassen, H.H., Bomben, G., Gunther, E., 1991. Speech characteristics in depression. Psychopathology 24, 88–105.

Stassen, H.H., Kuny, S., Hell, D., 1998. The speech analysis approach to determining onset of improvement under antidepressants. Eur. Neuropsychopharmacol. 8, 303–310.
Steiger, A., Kimura, M., 2010. Wake and sleep EEG provide biomarkers in depression. J. Psych. Res. 44, 242–252.
Stein, D.J., Phillips, K.A., Bolton, D., Fulford, K.W.M., Sadler, J.Z., Kendler, K.S., 2010. What is a mental/psychiatric disorder? From DSM-IV to DSM-V. Psychol. Med. 40, 1759–1765.
Stewart, W.F., Ricci, J.A., Chee, E., Hahn, S.R., Morganstein, D., 2003. Cost of lost productive work time among us workers with depression. JAMA 289, 3135–3144.
Stratou, G., Scherer, S., Gratch, J., Morency, L.-P., 2014. Automatic nonverbal behavior indicators of depression and PTSD: the effect of gender. J. Multimodal User Interf., 1–13
Stroop, J.R., 1935. Studies of interference in serial verbal reactions. J. Exp. Psychol. 18, 643.
Sturim, D., Torres-Carrasquillo, P.A., Quatieri, T.F., Malyska, N., McCree, A., 2011. Automatic detection of depression in speech using Gaussian mixture modeling with factor analysis. Proceedings of Interspeech. ISCA, Florence, Italy, pp. 2983–2986.
Sundberg, J., Patel, S., Bjorkner, E., Scherer, K.R., 2011. Interdependencies among voice source parameters in emotional speech. IEEE Trans. Affect. Comput. 2, 162–174.
Sweeney, J.A., Strojwas, M.H., Mann, J.J., Thase, M.E., 1998. Prefrontal and cerebellar abnormalities in major depression: evidence from oculomotor studies. Biol. Psych. 43, 584–594.
Sweeney, J.A., Kmiec, J.A., Kupfer, D.J., 2000. Neuropsychologic impairments in bipolar and unipolar mood disorders on the CANTAB neurocognitive battery. Biol. Psych. 48, 674–684.
Szabadi, E., Bradshaw, C.M., Besson, J.A., 1976. Elongation of pause-time in speech: a simple, objective measure of motor retardation in depression. Br. J. Psych. 129, 592–597.
Teager, H.M., Teager, S.M., 1990. Evidence for nonlinear sound production mechanisms in the vocal tract. Speech Prod. Speech Model. NATO ASI Ser. 55, 241–261.
Teasdale, J.D., Fogarty, S.J., Williams, J.M.G., 1980. Speech rate as a measure of short-term variation in depression. Br. J. Soc. Clin. Psychol. 19, 271–278.
Tolkmitt, F., Helfrich, H., Standke, R., Scherer, K.R., 1982. Vocal indicators of psychiatric treatment effects in depressives and schizophrenics. J. Commun. Disord. 15, 209–222.
Trevino, A., Quatieri, T., Malyska, N., 2011. Phonologically-based biomarkers for major depressive disorder. EURASIP J. Adv. Signal Process. 2011, 1–18.
Valstar, M., Schuller, B., Smith, K., Eyben, F., Jiang, B., Bilakhia, S., Schnieder, S., Cowie, R., Pantic, M., 2013. AVEC 2013: The continuous audio/visual emotion and depression recognition challenge. In: Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge. Barcelona, Spain, pp. 3–10.
Valstar, M., Schuller, B., Smith, K., Almaev, T., Eyben, F., Krajewski, J., Cowie, R., Pantic, M., 2014. AVEC 2014: 3D dimensional affect and depression recognition challenge. Proceedings of the 4th ACM International Workshop on Audio/Visual Emotion Challenge (AVEC '14). ACM, Orlando, Florida, USA, pp. 3–10.
Vanger, P., Summerfield, A.B., Rosen, B.K., Watson, J.P., 1992. Effects of communication content on speech behavior of depressives. Compr. Psych. 33, 39–41.
Vang, F.J., Ryding, E., Träskman-Bendz, L., van Westen, D., Lindström, M.B., 2010. Size of basal ganglia in suicide attempters, and its association with temperament and serotonin transporter density. Psych. Res. 183, 177–179.
Van Orden, K.A., Witte, T.K., Cukrowicz, K.C., Braithwaite, S.R., Selby, E.A., Joiner Jr., T.E., 2010. The interpersonal theory of suicide. Psychol. Rev. 117, 575–600.
Verona, E., Sachs-Ericsson, N., Joiner Jr., T.E., 2004. Suicide attempts associated with externalizing psychopathology in an epidemiological sample. Am. J. 161, 444–451.
Ververidis, D., Kotropoulos, C., 2006. Emotional speech recognition: resources, features, and methods. Speech Commun. 48, 1162–1181.
Vogel, A.P., Fletcher, J., Maruff, P., 2010. Acoustic analysis of the effects of sustained wakefulness on speech. J. Acoust. Soc. Am. 128, 3747–3756.
Walker, J., Murphy, P., 2007. A review of glottal waveform analysis. In: Stylianou, Y., Faundez-Zanuy, M., Esposito, A. (Eds.), Progress in Nonlinear Speech Processing SE – 1, . In: Lecture Notes in Computer Science. Springer, Berlin Heidelberg, pp. 1–21.
Watson, D., 2005. Rethinking the mood and anxiety disorders: a quantitative hierarchical model for DSM-V. J. Abnorm. Psychol. 114, 522–536.
Weeks, J.W., Lee, C.-Y., Reilly, A.R., Howell, A.N., France, C., Kowalsky, J.M., Bush, A., 2012. "The Sound of Fear": assessing vocal fundamental frequency as a physiological indicator of social anxiety disorder. J. Anxiety Disord. 26, 811–822.
Weiland-Fiedler, P., Erickson, K., Waldeck, T., Luckenbaugh, D.A., Pike, D., Bonne, O., Charney, D.S., Neumeister, A., 2004. Evidence for continuing neuropsychological impairments in depression. J. Affect. Disord. 82, 253–258.
Williamson, J.R., Quatieri, T.F., Helfer, B.S., Horwitz, R., Yu, B., Mehta, D.D., 2013. Vocal biomarkers of depression based on motor incoordination. In: Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge. Barcelona, Spain, pp. 41–48.
Williamson, J., Quatieri, T., Helfer, B., Ciccarelli, G., Mehta, D.D., 2014. Vocal and facial biomarkers of depression based on motor incoordination and timing. Proceedings of the 4th ACM International Workshop on Audio/Visual Emotion Challenge (AVEC '14). ACM, Orlando, Florida, USA, pp. 65–72.
World Health Organisation, 2014. Preventing Suicide: A Global Imperative. Geneva.
Xia, R., Liu, Y., 2012. Using i-vector space model for emotion recognition. In: 13th Annual Conference of the International Speech Communication Association Interspeech 2012. Portland, USA, pp. 2230–33.
Yang, Y., Fairbairn, C., Cohn, J., 2012. Detecting depression severity from vocal prosody. IEEE Trans. Affect. Comput. 4, 142–150.
Yap, T.F., Ambikairajah, E., Epps, J., Choi, E.H.C., 2010. Cognitive load classification using formant features. In: 10th International Conference on Information Sciences Signal Processing and Their Applications, 2010 (ISSPA '10), pp. 221–224.
Yingthawornsuk, T., Shiavi, R.G., 2008. Distinguishing depression and suicidal risk in men using GMM based frequency contents of affective vocal tract response. In: International Conference on Control, Automation and Systems, 2008, ICCAS 2008, pp. 901–904.
Yingthawornsuk, T., Keskinpala, H.K., France, D., Wilkes, D.M., Shiavi, R.G., Salomon, R.M., 2006. Objective estimation of suicidal risk using vocal output characteristics. Proceedings of Interspeech. ISCA, Pittsburgh, USA, pp. 649–652.
Yingthawornsuk, T., Keskinpala, H.K., Wilkes, D.M., Shiavi, R.G., Salomon, R.M., 2007. Direct acoustic feature using iterative EM algorithm and spectral energy for classifying suicidal speech. Proceedings of Interspeech. ISCA, Antwerp, Belgium, pp. 749–752.
You, C.H., Lee, K.A., Li, H., 2010. GMM-SVM kernel with a Bhattacharyya-based distance for speaker recognition. IEEE Trans. Audio, Speech, Lang. Proc. 18, 1300–1312.
Zhou, G., Hansen, J.H.L., Kaiser, J.F., 2001. Nonlinear feature based classification of speech under stress. IEEE Trans. Speech Audio Process. 9, 201–216.
Zhou, Y., Scherer, S., Devault, D., Gratch, J., Stratou, G., Morency, L.P., Cassell, J., 2013. Multimodal prediction of psychological disorders: learning verbal and nonverbal commonalities in adjacency pairs. In: Proceedings of Workshop Series on the Semantics and Pragmatics of Dialogue, pp. 160–169.
Zue, V., Seneff, S., Glass, J., 1990. Speech database development at MIT: TIMIT and beyond. Speech Commun. 9, 351–356.