ANALYSIS OF NATURAL SPEECH FOR THE ASSESSMENT OF MOOD DISORDERS

by

Katerina Dikaios

Submitted in partial fulfilment of the requirements
for the degree of Masters of Science

at

Dalhousie University
Halifax, Nova Scotia
August 2021

Dalhousie University is located in Mi'kma'ki,
the ancestral and unceded territory of the Mi'kmaq.
We are all Treaty people.

Table of Contents

List of Tables

List of Figures

Abstract

Psychiatric evaluation relies on subjective assessment. Biomarkers are objective indicators of illness that can aid in diagnostic classification and help guide treatment for individuals with psychiatric illnesses. Speech has been identified as an informative biomarker that is objective and easy to collect. Speech analysis has been shown to be effective in diagnostic classification, assessment of severity and prognosis, and early onset prediction of psychiatric illness. We aimed to synthesize results of published work and validate speech analysis methods for clinical application. We completed a systematic review to explore the state of the field and identify areas for further investigation. We present the collection and analysis methods for a speech study aimed at creating a corpus of high-quality speech data. We analyzed a preliminary sample using content speech features to differentiate bipolar from unipolar depression. Results from these preliminary analyses demonstrate the efficacy of our data collection procedure and the utility of content variables for tackling important classification problems in psychiatry.

Acknowledgements

First, I want to thank Dr. Uher. Rudolf, your interesting, critical perspectives on psychopathology have taught me invaluable lessons that I will take forward into my career.

I want to thank everyone at FORBOW for their tireless efforts collecting and coding speech samples. I would not have been able to do this without your help.

I need to thank our participants. These individuals show up and lend their experiences wholeheartedly to our research. Not only would this all not be possible without them, but it is a pleasure to experience their selflessness and willingness to participate if it can contribute to helping others.

I would like to thank our colleagues in the Computer Science Department at Dalhousie. Sri Harsha Dumpala and Sageev Oore, thank you for bringing new perspectives into our speech project, and bringing forth the methods that will increase the impact of this work. Also, thank you for your smiles, positive attitudes, and last-minute assistance.

I want to say a huge thank you to Sheri Rempel. Without you, our speech projects wouldn't exist, and I may not have my sanity nearly as intact! Thank you for being an incredible friend and teacher of all things speech!

I must thank Emily Howes-Vallis. Em, your guidance through the process of thesis-ing kept me on track and gave me important tools for the future of my research career. The fact that you did it all with kindness was just a bonus!

I have to thank my incredible partner Julia. Amor, me mantuviste sobre la tierra cuando me estresaba, y siempre sabias como hacerme sonreir. Gracias por todo.

Lastly, I want to thank my family. You guys instilled in me the importance of education, and have constantly supported my academic pursuits, even when it made me into a snobby researcher. I love you.

Chapter 1: Introduction - Applications of Speech Analysis in Psychiatry

Contributions:

KD and RU developed the concept for this manuscript. KD systematically searched the databases and screened all search results. KD extracted and organized the data, synthesized results, and wrote the initial draft. KD and RU completed the writing. RU, SR, SO and MK reviewed the draft and providing edits. KD kept the database search up to date and finalized the writing.

Applications of Speech Analysis in Psychiatry

Katerina Dikaios, B.Sc.,[1,2], Sheri Rempel, M.Sc.[2], Sageev Oore, PhD.[3,4], Michael Kiefte, PhD[5], Rudolf Uher, M.D.[1,2]

[1]Department of Psychiatry, Dalhousie University, Halifax, NS
[2]Nova Scotia Health, Halifax, NS
[3]Faculty of Computer Science, Dalhousie University, Halifax, NS
[4]Vector Institute for Artificial Intelligence, Toronto, ON
[5]School of Communication Sciences and Disorders, Dalhousie University, Halifax, NS

Abstract

Assessment and diagnosis of mental disorders rely on clinical evaluation and self-report measures, both of which are prone to bias. The need for objective measurement has stimulated interest in alternative indicators of illness presence and severity. Speech may offer a source of information that bridges the subjective and objective in the assessment of mental disorders. We reviewed the literature to understand the utility of speech analysis for psychiatric applications. Acoustic features including speech rate, pausing, and pitch variability reflect reduced psychomotor tempo and restricted range of affect in depression and poverty of speech and blunting of affect in schizophrenia. Semantic features including connectedness and coherence reflect thought disorder and disorganization in psychosis. Content features including positive and negative word use and personal pronouns reveal attributional style and self-reference in mood and psychotic disorders. Models using multiple speech features show high levels of accuracy for diagnostic classification and assessment of psychiatric disorders. Patterns of speech characteristics can help differentiate multiple psychiatric diagnoses, estimate severity, and predict onset of mental disorders, prognosis and response to treatments. Automated analysis methods may perpetuate bias of human rating, unless sources of bias are adequately addressed. This review offers perspectives on potential harms associated with applications of speech analysis and their mitigation. Convergent progress in speech and computer sciences is opening avenues for implementing speech analysis to add objectivity to the assessment and prognostics in the clinical practice.

Introduction

Mental health clinicians rely on a combination of clinical judgement, patient recall, and assessment of present symptoms when attempting to determine a patient's diagnosis, illness severity, prognosis, or best treatment choice (Pavlova & Uher, 2020). The practice of psychiatry has been scrutinized for lacking objective measures for patient assessment; agreement between psychiatrists on the same diagnosis is variable (Matuszak & Piasecki, 2012; Pies, 2007). Research has shown that biomarkers, physiological signals of illness, may aid clinicians in making important decisions about a patient's illness or treatment (Fernandes et al., 2020; Huang et al., 2020; Poletti et al., 2021; Vai et al., 2020; Zheng et al., 2019). One such biomarker is speech.

Core features of mental illness are communicated through speech and language. A person's speech reflects their internal state in the content, tone, rhythm and volume (Newman & Mather, 1938; Straw & Callison-Burch, 2020). Clinical assessment and diagnosis in psychiatry often rely on speech features, as clinicians use patients' speaking behaviors to infer information about symptoms, diagnosis, and prognosis (Alhanai et al., 2018). Because many components of speech are habitual and spontaneous, speech analysis has a potential to provide an objective measure of illness.

Psychiatrists assess the spontaneity, rate, content, cohesion and affective modulation of speech when completing the mental state examination, a clinical protocol that is used in most patient encounters and contributes key diagnostic information. Yet, the reliability of this exam is rarely tested and available information suggests that the agreement between psychiatrists is limited (Barzilay et al., 2019; Rozenzweig et al., 1961; Serby, 2003). Poor agreement between clinicians, coupled with the risk of unconscious biases and

stereotyping, highlight the need for objective measures in psychiatric assessment (Pavlova & Uher, 2020a). Because audio recordings of speech can be easily and inexpensively collected, speech analysis presents the potential of objective measures that are broadly applicable in clinical practice (de Boer et al., 2018; Erol et al., 1995). Spurred by the development of natural language processing and speech recognition technology in the last two decades, evidence has started to accumulate suggesting the utility of speech characteristics in diagnosis, prognosis, symptom monitoring, illness onset prediction and treatment outcome monitoring (Corcoran et al., 2020; de Boer et al., 2018; Hanakawa, 2004a). Various fields contribute to the vast body of existing literature on the topic, creating a rich but heterogeneous picture of the utility of speech analysis in psychiatry.

We completed a systematic review for the purpose of synthesizing research results from multiple disciplines, with the aim of understanding potential psychiatric applications of speech analysis. The broad scope of this review is intended to present a high-level synthesis of the available research on this topic and convey clinically relevant considerations for mental health professionals. We systematically searched the existing and emerging literature on speech analysis across mental disorders, identifying relevant features of speech, highlighting promising approaches, and evaluating the potential for clinical applications.

Methods

Literature Search. We searched PubMed and PsycInfo for articles that combined topics from the domains of speech, psychiatric disorders, computation and diagnosis or prediction, indexed from the database conception date to January 13, 2021. Additionally, we searched the bibliographies of included articles. See Section A of Appendix A the for details of the literature search.

Eligibility Criteria. We included studies of participants living with a mental disorder. Articles were included if they focused on an aspect of speech as it related to one or more psychiatric disorders, their symptom severity, or prediction of illness onset, prognosis, or treatment outcomes.

We included articles that analyzed transcriptions of participants' speech. We excluded papers that focused on written text, such as participants' writing, social media postings, clinician's notes, suicide notes, hospital charts, or discharge reports. Written text was excluded because it may not always reflect unmetered syntactical and semantic mental processes, as spoken word does. We excluded case studies. We also excluded articles focusing on questionnaire measures without psychiatric diagnoses (i.e. depressive *symptoms* or schizotypal *traits*). We excluded studies of individuals with neurological illness, brain injury, intellectual disability or dementia.

We have identified 128 eligible reports, of which 86 dealt with diagnostic classification, 28 investigated severity assessment of mental disorders, 10 looked at the prediction of illness onset in high-risk populations, and 9 examined prognosis or treatment outcome prediction (Section B, Appendix A). Papers were classified under more than one category if they included analyses on multiple outcomes.

Speech features

Perspectives from psychiatry, psychology, computer analytics, linguistics, and speech and language pathology reveal aspects of speech and language that may reflect the speaker's mental state. Names and definitions of speech characteristics are partly discipline-specific. We provide a consensus glossary of relevant speech features in Table 1.

The key distinction agreed upon across disciplines is between *acoustic* features and *content* features. *Acoustic* features represent the sound of speech. *Content* features represent the meaning and grammar of spoken language. *Semantic* features refer to the meaning of speech and are included under content, but are also frequently used as a separate term. Table 1 provides information on how semantic features contribute to the greater content category.

Classifying diagnoses using speech analysis

The most common application of speech analysis has been the comparison between individuals with and without a psychiatric disorder for diagnostic classification. Of the 128 studies identified, 86 (67%) compared speech features between individuals with psychiatric disorders and controls.

Depression and Schizophrenia

The majority of studies published to date have focused on major depressive disorder and schizophrenia (Figure 1; Table 2) During a depressive episode, lack of energy and psychomotor retardation affect the production of speech, resulting in audible acoustic changes (Flint et al., 1993; France et al., 2000; Quatieri & Malyska, 2012; Scherer et al., 2015; M. Smith et al., 2020; Wang et al., 2019). Slower and decreased movement of muscles results in speech that has lower pitch, lower fundamental frequency and narrower pitch range than speech of people without depression (Alghowinem, 2015; Breznitz, 1992; Nilsonne et al., 1988). Individuals with depression tend to pause more often when speaking, and their pause duration is more variable than in individuals without depression, resulting in a loss of natural speech rhythm (Alghowinem, 2015; Breznitz, 1992). Similar speech features characterize depression in adults and children (McGinnis et al., 2019). These acoustic speech features can be objectively measured using acoustic analysis (Jiang et al., 2018).

Acoustic features can distinguish people with and without depression with an average accuracy of 83% across studies (Table 2), and may be robust to manipulation. In one study, participants were asked to conceal any speech behavior they thought might be indicative of their depression. Although the participants were attempting to disguise their

depression when speaking, the accuracy of depression classification remained high at 81%, only slightly lower than the 88% accuracy without concealment (Solomon et al., 2015). In another study, the volume and frequency spectra components of speech classified depression consistently across 12 speech scenarios, including neutral, positive and negative contexts (Wang et al., 2019)**.**

Speech content in depression reflects a bias towards negative experiences. The negative attributional style, characteristic to depression, leads to emphasis on negative words and attribution of positive experiences and accomplishments to outside influences rather than personal efforts (Sweeney et al., 1986). Individuals with depression tend to dwell on negative aspects of experiences, and ruminate about past errors (Breznitz, 1992). They use more negative words than controls in various types of speaking tasks (Breznitz, 1992; Himmelstein et al., 2018; Mete, Schnurr, et al., 1993). They also use more first person pronouns than controls when recalling negative memories, but less when recalling positive memories, reflecting a stronger personal identification with negative experiences than positive ones (Himmelstein et al., 2018). While confirming the expectations about negative bias in depressive speech, content features may be more easily manipulated by the speaker, and less objective than acoustics. To date, it is unclear if content features improve the classification of depression over and above acoustic features alone.

Speech analysis in schizophrenia and psychotic illness highlights semantic features, such as connectedness and coherence. Individuals with schizophrenia often speak in a way that is less connected and predictable than individuals without psychotic disorders (Elvevåg et al., 2007). Reduced coherence relates to both negative and positive symptoms of

schizophrenia, and can be objectively measured (Maher et al., 2005; Moe et al., 2016; Willits et al., 2018).

In speech, positive symptoms of schizophrenia manifest as increased normative associations, where words are followed by other words that are semantically connected, but unrelated to the logical line of the narrative (Maher et al., 2005; Manschreck et al., 2012). Negative symptoms manifest as a slower rate of speech with shorter utterances, more pauses, reduced variation in frequency, and decreased density of ideas (Ayer et al., 2016; Moe et al., 2016; Willits et al., 2018). The reduced coherence, increased normative associations and reduced richness of content reflect the disorganization domain of schizophrenia symptoms, which includes formal thought disorder (Ayer et al., 2016; Elvevåg et al., 2007; Pauselli et al., 2018). Coherence scores of participants with schizophrenia correlate negatively with clinical measures of formal thought disorder, supporting this speech construct as an accurate measure of disorganization (Elvevåg et al., 2007). Automated analysis of these semantic features distinguishes individuals with schizophrenia from healthy controls with an accuracy ranging from 70% to 83% (Elvevåg et al., 2007, 2010; Willits et al., 2018). The detection of psychosis using semantic features may be partly task-dependent (Cohen et al., 2016).

Other Disorders

Common speech features found in both depression and schizophrenia include low speech rate, frequent pauses, reduced frequency variation, and increased use of first-person pronouns. Other mental disorders have been less studied (Figure 1; Table 2). The one study of bipolar disorder found that speech of individuals in a manic phase is marked by faster rate and higher frequency variation, as well as semantic features similar to those

characterizing the positive symptoms in schizophrenia (J. Zhang et al., 2018). Two studies found that individuals with anxiety disorders switch topics more often than controls, speak with more hesitation, and make more breathy sounds (Kotsopoulos & Mellor, 1986; Toazza et al., 2016). Published studies suggest that automated analysis of speech can accurately distinguish individuals' mental disorders from healthy controls with high accuracy, but more work is needed to extend the knowledge beyond depression and schizophrenia.

Differentiating Disorders and Understanding Illness Characteristics

While differentiating between mental disorders may be harder than separating between patients and healthy controls, both acoustic and content features may help in this effort (Mota et al., 2012; Sonnenschein et al., 2018). For example, shorter utterances in schizophrenia and grammatical deviance in bipolar disorder contribute to differentiating these two disorders (R. E. Hoffman, 1986). In some cases, discrimination between diagnoses can be made with word categories. Dictionaries have been created for use with content analysis methods, categorizing words by different semantic meaning (Maher et al., 2005; Mete, Doganer, et al., 1993; Rosenberg et al., 1990). Word categories have been able to differentiate between schizophrenia, depression, bipolar disorder and controls with 48-85% accuracy (Erol et al., 1995, 1995; Mete, Schnurr, et al., 1993; Oxman et al., 1988; Rosenberg et al., 1990).

Some studies found that specific word categories are used more often by individuals with schizophrenia (Minor et al., 2015; Novack, 2003; Rosenberg, 1979). Categories such as "pleasure" and "distress" (Novack, 2003), and social-, sadness-, anger-, and work-related words (Minor et al., 2015) helped discriminate individuals with schizophrenia from those

with other diagnoses. Similarly, expressive word use (such as “sad” or “anxiety” related words) has been shown to accurately differentiate individuals with anxiety from those with depression; however, the frequent comorbidity of depression and anxiety blurs this distinction (Sonnenschein et al., 2018).

Multi-faceted semantic features are able to differentiate between psychotic illnesses. Notably, semantic graphs, used to measure the connectedness and coherence of speech, distinguished individuals with bipolar disorder from those with schizophrenia with 93% accuracy (Mota et al., 2012). Individuals with mania display higher levels of speech connectivity, have more flighty discourse, and speak more quickly than those with psychotic illness.

As some illness features are shared between psychiatric diagnoses, modeling multiple speech features differentiates diagnoses better than any single feature on its own (Fraser et al., 1986; Stassen & Bomben, 1991). Depression and schizophrenia both present flat affect and reduced speech production. Schizophrenia and bipolar disorder share reduced coherence and increased normative associations, but differ in the rate and volume of speech production. Depression and anxiety share preferential use of negative words, but frequent switching of topics and breath sounds may be relatively specific to anxiety. Distinction between multiple diagnoses may require modeling of many speech features to leverage both common and specific speech features across diagnoses (Cohen et al., 2012; Novack, 2003; Perlini et al., 2012). While acoustic features perform well in the assessment of flat affect and psychomotor arousal (Novack, 2003; Stassen & Bomben, 1991), semantic features can help differentiate between psychotic and non-psychotic

types of mental illness (Mota et al., 2012). More work is required to gage the potential of combining feature categories for multi-diagnosis classifications.

Assessing illness severity using speech analysis

Repeated measurement of symptom severity improves treatment outcomes by contributing updated information towards treatment planning (Guo et al., 2015), but severity measurement is underused in practice because of the time required to complete measurement scales (Aboraya et al., 2018). Additionally, the subjective measurement with questionnaires and interviews may be prone to bias (Pavlova & Uher, 2020a). Automated analysis of speech may offer a more efficient and objective tool for the evaluation of treatment efficacy and measurement-based care.

Assessing Depression Severity

Nineteen published studies have examined speech analysis in the measurement of severity of depression and psychosis (Figure 1). Depression severity can be measured by quantifying psychomotor retardation, which can be detected in the speech of affected individuals through the analysis of acoustic features. As depression worsens, psychomotor retardation manifests as progressively slower speech rate, quieter speech sounds, less energy variability, and more and longer pauses (Alpert et al., 2001; Cummins, Epps, & Ambikairajah, 2013; Cummins, Epps, Sethu, et al., 2013; Mundt et al., 2007; Quatieri & Malyska, 2012; Szabadi et al., 1976). Fewer studies examine agitated depression, and the fact that both extremes of psychomotor tempo may indicate worsening illness (Alpert et al., 2001; Novack, 2003). Retarded and agitated forms of depression share reduced affective modulation of speech, which is reflected in less variable pitch (Novack, 2003).

Content of speech may inform the measurement of depressive symptoms, especially when it is analyzed in the context of positive and negative emotion. Depression severity

correlates positively with word count and the frequency of first-person pronouns when recalling negative memories, but correlates negatively with word count and frequency of first-person pronouns when recalling positive memories (Himmelstein et al., 2018). We identified no studies that used both acoustic and content features in the same model, therefore it remains to be established whether the inclusion of content features improves the measurement of depression severity over acoustic features alone.

Assessing Severity of Psychotic Disorders

Acoustic, semantic and content features have shown promising results in measuring the severity of symptoms in psychotic disorders. Acoustic features have been used to assess the severity of symptoms, such as flat affect. Individuals with schizophrenia who score high on clinical measures of flat affect are less fluent, make less inflections, speak more slowly and pause more than individuals with schizophrenia with preserved affective reactivity (Alpert et al., 2000; Cohen et al., 2008). Greater severity of schizophrenia is also indexed by reduced coherence (Buck et al., 2015; Holshausen et al., 2014). While reduced coherence is specifically associated with positive symptoms, reduced variation of pitch is associated with negative symptoms (Hanakawa, 2004b). Other studies have reported that reduced connectedness in the speech of individuals with schizophrenia is associated with both positive and negative symptoms (Holshausen et al., 2014; Pauselli et al., 2018).

Automated analysis of acoustic features has helped identify confounds in the clinical assessment of symptoms (Alpert et al., 2000; Cohen et al., 2008). Alpert and colleagues found that independent clinical raters tended to include impressions of alogia, or poverty of speech, into flat affect ratings, increasing the likelihood that they would rate

participants with schizophrenia as having flat affect (1995, 2000). Similarly, Cohen found that clinician's ratings of vocal inflection were associated with automated measures of speech rate, a variable that measures speed of speech (2008). While clinician's ratings of various negative symptoms in schizophrenia displayed overlapping associations, computer-based measurements of the same phenomena were relatively unrelated to each other, suggesting greater specificity of automated measurement (Cohen et al., 2008).

Meta-Analysis: Clinical Validity of Speech Features

We analyzed the correlations between speech features and clinical severity measures in 14 studies with adequate information and found that individual speech features were weakly but consistently correlated with severity measures (Figure 2). The associations of speech features with severity measures of depression and psychosis were of similar magnitude (mean correlations of 0.23 [95%CI 0.13 to 0.32] and 0.20 [95%CI 0.09 to 0.30] respectively). Acoustic features showed, on average, slightly stronger correlations with severity measures than content features (Figure 2). While individual speech features showed weak correlations with clinical severity measures, algorithms combining multiple, primarily acoustic speech features led to estimates with correlations up to 0.80 with established severity measures (Braun et al., 2016; Mundt et al., 2007, 2012; Stassen et al., 1998). These results suggest that automated analysis of multiple speech features may allow more objective measurement of symptom severity with a meaningful accuracy. The applicability of such measurement across settings and populations remains to be established.

Predicting the onset of psychiatric illness in high-risk populations

Antecedents for mental illness can be detected early in childhood and adolescence. Some of these antecedents are predictive of illness onset in later years of life (Uher et al., 2014). Predicting illness onset in high risk populations is an important step towards targeted prevention and early treatment, which can lead to better lifetime illness outcomes (McGorry & Nelson, 2016).

Psychosis Onset Prediction

Automated analysis of speech may identify early subtle indications of risk. Psychosis onset can be predicted using semantic variables (Corcoran et al., 2018; Hanakawa, 2004a). Lack of intelligibility and poverty of speech are the characteristics most regularly identified in the speech of participants at high risk of developing a psychotic illness (Bearden et al., 2000). Notably, models that incorporate coherence and complexity variables predict the onset of psychosis within 2.5 years with 79% to 100% accuracy (Bedi et al., 2015; Corcoran et al., 2018; Elvevåg et al., 2010; Rezaii et al., 2019). In one such study, a classifier model that incorporated coherence measured by latent semantic analysis (LSA), frequency of determiners (“what”, “which”, “that”, etc.), and maximum number of words per phase to measure poverty of speech perfectly predicted which youth developed a psychotic disorder over a 2.5-year follow-up in a small sample (Bedi et al., 2015).

The high accuracy of psychotic illness prediction from semantic features supports that development of psychosis may manifest objectively in speech before reaching a point where a clinical diagnosis is possible. Semantic speech variables can also differentiate healthy relatives of people with psychotic illness from controls with up to 90% accuracy

(Elvevåg et al., 2010; Manschreck et al., 2012), likely reflecting a subclinical manifestation of familial risk (Corcoran et al., 2018). Given that family history is the strongest known risk factor for psychosis, the findings in relatives strengthen the validity of semantic speech features in predicting psychosis.

Depression Onset Prediction

The only study that examined the utility of speech analysis for predicting depression onset (Figure 1) found that acoustic features, especially glottal features, predicted which high-risk adolescents would develop depression over a 2-year follow-up with 73% accuracy (Ooi et al., 2013). The glottis is a physical structure in the throat that consists of the vocal cords and the space between them, and is affected by symptoms of depression such as psychomotor retardation. Glottal features have also been implicated in depression classification and severity assessment (Moore et al., 2008, 2004). Acoustic features extracted from glottal processes tap the physiology of speech production that may overlap with the early pathogenesis of depression. Future research may examine acoustic variables and other speech features to replicate and extend the promising results.

In summary, while the literature on onset prediction is relatively limited, it suggests that speech analysis may accurately predict onsets of both psychosis and depression. The strong early results suggest that speech sampling should be implemented as part of prospective studies aiming at early identification of risk for mental illness.

Predicting illness prognosis and treatment outcomes

Choosing a treatment is one of the most important and complex clinical decisions in psychiatry. Clinical and biological markers may not provide reliable signal for a patient's likelihood to respond, making the outcome of many treatments unknown before they begin. The ability to predict treatment outcome and illness progression can allow clinicians to personalize treatment for their patients, ideally leading to better long-term outcomes. Speech features are easy to obtain, making them a desirable component for clinical decision support.

The literature on this topic has focused on depression (Figure 1). Content features have been used to predict later illness outcomes and treatment response. One study found that individuals who used more first-person singular pronouns had poorer prognoses and higher depression severity at eight-month follow-up (Zimmermann et al., 2017). This result echoes severity assessment literature on the role of self-focused attention in depression. Three articles looked at emotional word use to predict treatment outcomes in depression and reported conflicting results. Miller found that people who used more emotional words were more likely to be responders to antidepressants (1996). Carrillo and colleagues found that people who used fewer emotional words were more likely to respond to psilocybin (2018). In a third study, participants who used more positive words and fewer past-focused and negative words responded better to psychological therapy (Huston et al., 2019). It remains to be established whether the opposing results may be due to the difference in depressed samples (treatment-resistant (Carrillo et al., 2018) vs. treatment-responsive (Miller, 1996)), the choice of treatment, or other reasons. Given the

limited scope of the literature on this topic, more research is needed to assess whether speech content features will be useful in the prediction of treatment outcomes.

One study examined patients with recent-onset psychosis to determine their likelihood to receive a schizophrenia diagnosis. Speech disorganization, as detected by semantic graphs of memory reports, predicted a diagnosis of schizophrenia 6 months later, but did not predict a diagnosis of bipolar disorder (Mota et al., 2017). These results lend further support to speech coherence and disorganization as relatively specific markers of psychosis.

Although acoustic features show promising evidence when used to classify or assess the severity of psychiatric disorders, we found no studies that used acoustic measures at baseline as an indicator of treatment outcome or prognosis. While the results published thus far show potential for the role of speech analysis in predicting outcomes, more research is needed to find out which speech features best predict treatment outcomes and how accurate prediction can be achieved.

Detection of suicide risk in speech

Acoustic speech features have been used to assess suicide risk in individuals with depression (France et al., 2000; Pestian et al., 2016; Scherer et al., 2015). Levels of vocal jitter increase with suicide risk; individuals at a higher risk for suicide have more erratic fundamental frequencies than those at lower or no risk (France et al., 2000; Ozdas et al., 2004). Glottal flow spectrum and jitter are measures of disruption in fundamental frequency, and are products of lack of vocal cord control or coordination (Ozdas et al., 2004; Teixeira et al., 2013). A machine learning classifier that incorporated both measures was able to discriminate low-risk depressed patients from high-risk suicide patients with 75% accuracy (Ozdas et al., 2004). The acoustic representation of suicide risk appears distinct from that of depression and merits a separate focus in speech analysis research.

Considerations for the Interpretation and Integration of Automated Analysis Results

Although the need for unbiased assessment has been one of the key drivers for developing automated speech analysis, recent work has highlighted the propagation of biases through machine learning models (Chouldechova & Roth, 2018). As human rating is used to train speech and language models, the sources of bias already present in healthcare become embedded into the development of automated measurement (Straw & Callison-Burch, 2020). This is pertinent to mental health, where biases relating to gender, race, and culture have been consistently identified (Chaplin, 2015; Fàbrega, 2001; Lee, 2002). Therefore, sources of bias have to be identified, assessed and addressed to ensure fairness before any automated model is considered for deployment in health care (Bird et al., 2019; Thieme et al., 2020).

The broad distribution of fields contributing to psychiatric speech analysis research presents the problem of divided research efforts (Straw & Callison-Burch, 2020). The advent of open-access psychiatric speech datasets provides the opportunity for a wider reach of inquiry, but may allow for interpretation of results that are detached from their origin. For instance, computer science researchers creating machine learning models to handle tasks of psychiatric assessment are not always fully informed about the background of their data. Features such as sampling and collection procedures and clinical and demographic characteristics of a sample can introduce unconscious biases that will subsequently be propagated in machine learning models (Straw & Callison-Burch, 2020). In the current review, few machine learning studies acknowledged possible

sources of bias, and fewer still attempted to control for them. Although promising, the results presented here must be interpreted with an understanding of the likely sources of bias present within them, and the possibility for their influence to perpetuate these biases in potentially harmful ways. Sun et al. provide recommendations for dealing with bias in speech-specific machine learning models (2019).

Along with bias, additional considerations for machine learning methods stem from the desire to objectify assessment in psychiatry. Integration of automated assessment tools such as speech analysis in clinical settings has the potential to dehumanize the clinical process. Although clinical judgment can carry biases (Snowden, 2003), replacing it with automated processes carries the potential for other problems. Trained clinicians' interpretation of psychiatric symptoms may detect more subtle nuances of patient illness that could be essential to providing timely and appropriate treatment (Pies, 2007). Recommendations for incorporating automated speech analysis tools into clinical environments should take care not to overestimate the ability of automation to replace clinical judgement. Instead, the addition of automated tools can add objectivity and efficiency to traditional clinical encounters in psychiatry, which may increase a clinician's ability to optimize assessment and care for patients (Carroll & Rounsaville, 2010).

Discussion

One of the key challenges for psychiatry in the 21st century is to reconcile the inherently subjective character of mental illness with the need for objective measurement. The growing accessibility of computational analysis tools coupled with the ease of obtaining audio samples make the automated analysis of speech a natural choice for bridging the subjective and objective in psychiatry. Evidence shows that speech analysis can classify depression and psychosis with high accuracy, measure their current severity, and predict their onset among individuals at risk. The concurrent validity with established clinical measures qualifies speech analysis as a research tool that can add objectivity and specificity to the measurement of psychopathology. The high accuracy of many reported predictions suggest that automated speech analysis could also be implemented in clinical settings.

The existing literature is disproportionately focused on differences between groups of individuals with depression or schizophrenia and healthy controls (Figure 1). A smaller but consistent body of literature suggests that automated analysis of speech can also be used as a measure of depression severity. It may provide an attractive alternative to clinical rating scales with a potential to facilitate implementation of measurement-based care (Lewis et al., 2018). The studies using speech to predict onsets or outcomes of mental illness are few in number, but may have the highest implementation potential. Speech analysis could enable more accurate prediction of the development of schizophrenia (Bedi et al., 2015) or depression (Ooi et al., 2013), as it appears that indicators may manifest in speech years before onset. High prediction accuracies mandate the application of speech sampling in large prospective studies to probe the

reproducibility and generalizability of what appears to be one of the most promising predictive biomarkers in psychiatry.

Rapidly growing interest in speech analysis has inevitably brought about variability in the focus of the literature, often dependent on discipline. Additionally, time period, language, demographics, medications, sex and gender all play roles in the variance of speech behaviors that a person may exhibit, and in the outcomes of studies throughout time and across disciplines. Analyzing demographic and cultural factors of all records included was beyond the scope of this review, however there are important implications that these factors may have for integrating these methods into clinical practice. Understanding speech features as indicators of mental illness requires the consideration and control of these and many other variables. Although existing results are promising, the work is characterized by methodological heterogeneity. Future work should endeavor to gain a more in-depth understanding of whether speech features can endure as indicators of illness after controlling for the many potential confounding influences. Straw et al. provide an extensive exploration of these and other potential sources of bias in speech and language analysis in psychiatry (2020).

Recently, some research has aimed to quantify differences in speech analysis results that are due to methodological heterogeneity. Cohen (2016) and Wang (2019) questioned previous conclusions by combining existing samples to tease out important findings, and were able to clarify previous results with more reliability and power. Open-source speech datasets will play a large role in advancing speech analysis past the research domain, as they allow for tests of reproducibility, verification and validation of previous findings.

With maturation of methodology, future work will substantiate existing results and decrease the quantity of contradictory results as the field advances.

The articles reviewed provided evidence for the both transdiagnostic and disorder-specific speech features. Studies focused on a single diagnosis have demonstrated the ability for speech features to discriminate between participants and controls, but few have accounted for the potential transdiagnostic and comorbidity confounds. Although clinical measurements have been used to validate the results that speech features present, it is unclear what proportion of these validations included samples of individuals with psychiatric comorbidities. As comorbid diagnoses are commonplace in clinical environments, clinical implementation of speech analysis requires a foundation of research that includes participants with various diagnoses and accompanying comorbidities.

Most studies to date have examined multiple variables from the same speech feature category, i.e. either acoustic or content/semantic features (Cummins et al., 2011; Jiang et al., 2018; McGinnis et al., 2019; Moore et al., 2008; Ooi et al., 2013). Evidence suggests that a predictive model should incorporate all of the speech features that best fit the illness characteristics of the disorder being studied. Machine learning methods allow combination of features in generalizable multivariate models (Tasnim & Stroulia, 2019). With increasing size of available samples, multivariate methods using both acoustic and content features of speech may be optimized and implemented for clinical settings.

Speech analysis for the diagnosis and assessment of psychiatric disorders has been attracting attention for eight decades (Newman & Mather, 1938). Advances in clinical, speech and computer sciences are now making its application a plausible option. The

existing literature is rich with potential, and is beginning to address biases and innovate towards practical and fair methods for clinical application.

Tables and Figures

Table 1

Glossary of acoustic and content features

ACOUSTICS – Range of measurable sound components of voice	
Pitch Frequency/Fundamental frequency (F_0) Tone	Features associated with the rise and fall of vocal sound
Volume Loudness Intensity Energy	Features associated with the strength of a speech sound
Prosody	*Inflection/Intonation* Variation in pitch/frequency
	Emphasis/Syllable stress Variation in loudness
	Pace/Rhythm/Rate Speed of verbal communication
	Fluency Frequency and length of utterances and pauses

CONTENT – Presence and usage of linguistic components of speech

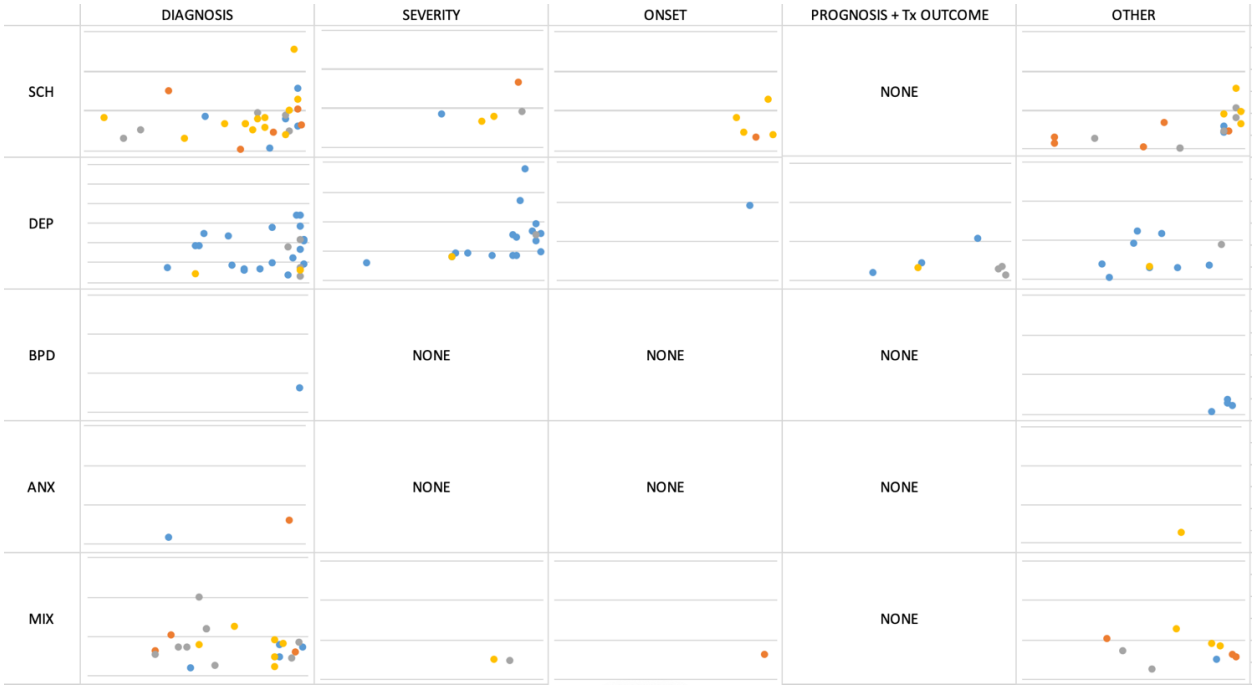
Semantic	*Coherence/ Connectedness/Disorganization* Relatedness of words, sentences and topics in a discourse
	Cohesion Grammatical accuracy within a sentence
	Intelligibility Ease of understanding (of speech)
	Syntactic Error Grammatical and phonological flaws or inaccuracies
Word categories Pronoun use	Words belonging to differently valanced categories (“sad”, “distressed”, “excited”), sentiment of words, word usage, and the distinguishing use of personal pronouns in discourse.

Table 2

Number of studies classifying psychiatric disorders by diagnosis, including average accuracies and sample sizes

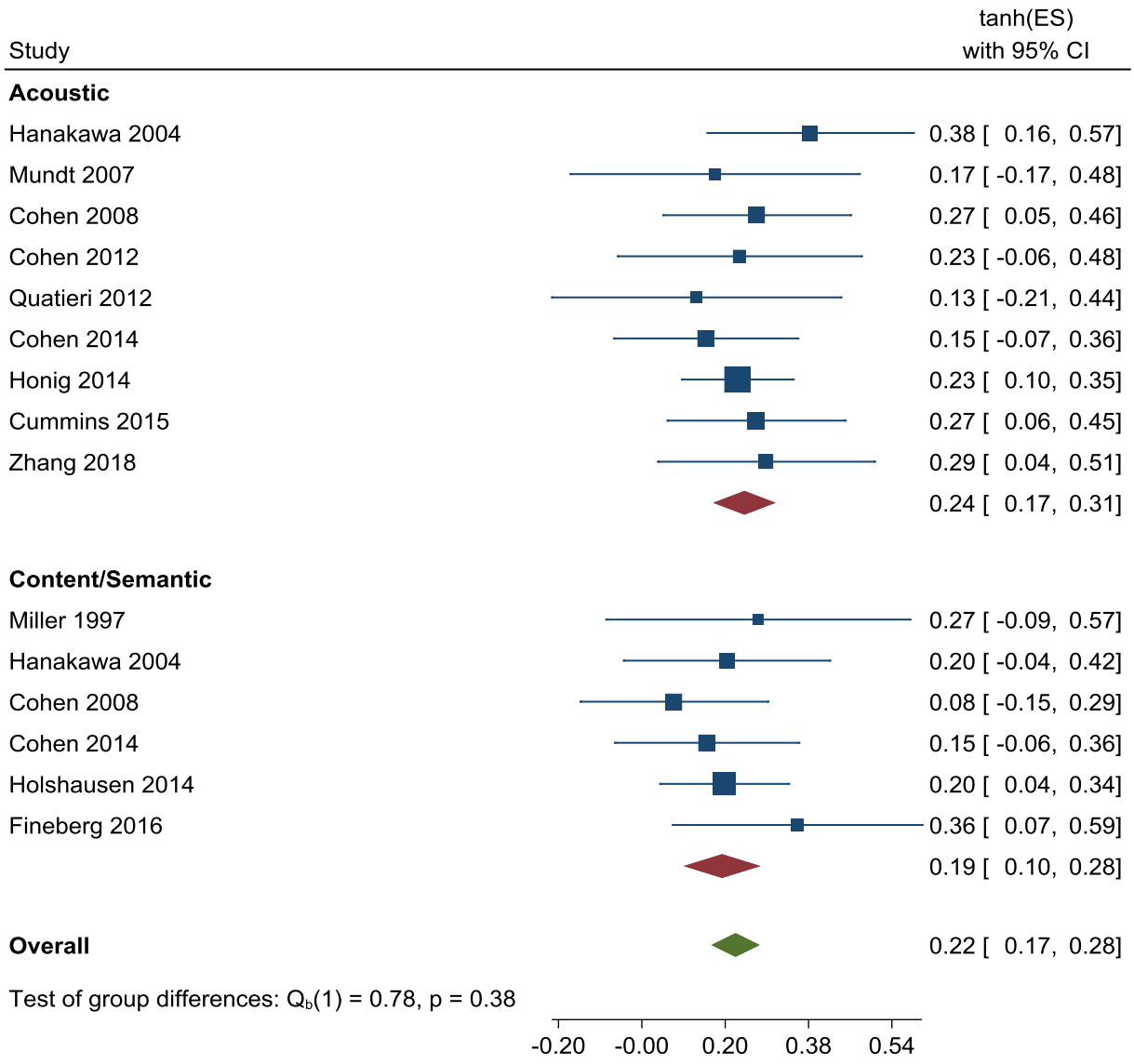
Diagnosis	Number of studies	Classification accuracy (mean)	Number of participants per study (mean)
Schizophrenia	29	81.8%	89
Depression	28	83.0%	78
Bipolar Disorder	1	N/A	60
Anxiety	2	N/A	39
Other	2	N/A	45
Multiple diagnoses	24	78.1%	105

Figure 1: *Disorders and topics addressed in published studies on speech analysis*.



Note. In each scatterplot in the matrix, the x-axes represent publication year (1965-2020), and the y-axes represent sample size (0-300). Colors represent speech feature category (acoustic=blue, semantic=orange, content=gray, mixed categories=yellow)

Figure 2: *Correlations between speech features and traditional severity measures.*



Note. The forest plot summarizes 379 correlations between acoustic and content speech features and severity measures of depressive and psychotic symptoms reported in 14 studies.

Chapter 2: Informed Study Design

Our intention in reviewing published speech analysis literature was to create a wide knowledge base on which to design an original study. Important contributions were made to the large body of existing research by experts from three main areas: psychiatry and psychology, speech language pathology and speech science, and computer science and artificial intelligence/machine learning. Studies from psychiatry and psychology were characterized by well-assessed samples from clinical populations, and they prioritized the evaluations of speech features in a manner that was relevant to traditional clinical judgement. Primarily, this information allowed us to understand how to meaningfully index psychopathology using speech features relevant to illness, and the importance of precisely measuring participants on various facets of their illness for the purpose of validating speech behavior in our analyses.

Studies from speech fields provided an in-depth look at the physiology and motor components of speech production. These results gave context to findings from psychiatry and psychology by providing a connection to the pathophysiological origins of disorder symptoms. For example, studies positioned glottal and spectral features as the bridge between the aroused or flat speech of depression and the psychomotor agitation or retardation that causes it (Moore et al., 2008, 2004; Ooi et al., 2013; Ozdas et al., 2004; Teixeira et al., 2013). In terms of acoustic analysis, these findings provided a deeper understanding of the importance of certain features for disorder-specific assessment.

Investigations from computer science experts put a spotlight on automated analysis. Machine learning models are being integrated into mental health research widely (Thieme et al., 2020). Including these studies in our review was critical, as their findings

represent the future for the creation of automated clinical tools in healthcare. We learned about the importance of multiple feature modelling for more accurate analysis of psychopathology in speech.

We have put together a multidisciplinary research team with experts representing the fields of psychiatry, speech language pathology, speech science, and computer science. Understanding the various perspectives on speech analysis for the purpose of assessing psychopathology equipped our team to develop a study that addressed prominent gaps in the literature. Incorporating approaches from all faculties optimizes our ability to ensure best-practice sampling and analysis techniques.

The development of the study methods presented in Chapter 3 took place over a period of 2 years, with influence from team members of all backgrounds. We collected a dataset of speech samples from participants enrolled in ongoing projects and performed preliminary analyses to validate and solidify our methods. We continuously engaged team members in collaborative discussions to improve the efficacy of our methods and usability of our data. We obtained ethics approval in August of 2020 from the Nova Scotia Health Authority Research Ethics Board for The Vocal Mind Project, whose protocol is described in the next chapter.

Chapter 3: Sampling of Natural Speech for the Assessment of Psychopathology: Data Collection Procedure and Interrater Reliability

Contributions:

KD and RU outlined the concept for this manuscript. KD and SHD analyzed the data. KD wrote the manuscript with contributions from SR and RU. RU reviewed the draft and provided edits. KD finalized the manuscript.

Sampling of natural speech for the assessment of psychopathology: data collection procedure and inter-rater reliability

Dikaios, K.[1,2], Rempel, S.[2], Dumpala SH.[3], Oore S.[3,4], Alda M.[1,2], Kiefte, M.[5], Matwin, S.[3], Uher, R.[1,2]

[1]Department of Psychiatry, Dalhousie University, Halifax, NS

[2]Nova Scotia Health, Halifax, NS

[3]Faculty of Computer Science, Dalhousie University, Halifax, NS

[4]Vector Institute for Artificial Intelligence, Toronto, ON

[5]School of Communication Sciences and Disorders, Dalhousie University, Halifax, NS

Abstract

Speech is being studied extensively as a biomarker of psychiatric illness. Methods for speech analysis differ depending on research discipline (speech science, psychiatry, computer science). Consistent gaps have been identified in the current literature that limit the applicability of speech analysis methods into clinical environments. We present the sampling procedure and coding methods for a novel speech analysis study assessing psychopathology. The study described was developed to complement existing work by incorporating features to address gaps in the field. Speech variables were created to index important features of psychiatric illness to aid in objective assessment. The speech collection procedure successfully elicits continuous, natural narratives from participants while minimizing and standardizing experimenter speech, providing audios that are ideal for both manual and automated analysis. Multi-level coding practices allow detailed analysis of all parts of participant speech, which can be compared against thorough psychopathological assessment by experimenters. Interrater reliability for all speech variables is presented and discussed.

Introduction

Assessment of mental disorders through interviews and questionnaires is prone to bias (Pavlova & Uher, 2020b). Speech may be an accessible and rich source of information for a more objective measurement of psychopathology. Both speech content and speech sounds contain information about a person's mood and mental state (Newman & Mather, 1938). Measuring speech is becoming increasingly more common and simple – natural speech is easy to record, and automated analysis methods continue to become more accessible (Ratana et al., 2019; Tasnim & Stroulia, 2019). Acoustic, semantic, and content features have been used to classify diagnoses of mental disorders, assess their severity, and predict illness onset and treatment outcomes (Dikaios et al., submitted). However, gaps in literature include the trans-diagnostic vs. disorder-specific nature of speech features (Fraser et al., 1986; Stassen & Bomben, 1991), study of disorders other than psychosis and depression (Dikaios et al., submitted), multi-diagnostic classification (Sonnenschein et al., 2018), and application of speech analysis in clinical decision-making to improve prognosis or treatment outcomes.

Most prior studies focused on either the content or the acoustics of speech (Cohen & Elvevåg, 2014; Tasnim & Stroulia, 2019), leaving the relative contributions and added value of joint analysis of content and acoustic features unexplored. Advanced analytical models may address these remaining questions and bring the speech analysis tools closer to clinical application. These methods will require large representative samples of individuals with a variety of diagnoses and speech sampling with standardized procedures.

We report on the development of a speech collection and analysis protocol aimed to facilitate large scale sample collection and support the next steps in the development of speech analysis for psychiatric applications. We further report the development of an integrated semantic and content feature coding system informed by multidisciplinary expertise, with the aim of indexing speech features relevant to symptom profiles across diagnoses. Finally, we present initial data on inter-rater reliability of speech coding.

Method

Participants

Participants include adults with a variety mental disorders and adults without any mental disorder according to the Diagnostic and Statistical Manual, version 5 (American Psychiatric Association, 2013). Participants for the present investigation were enrolled through a participatory registry of patients at the Nova Scotia Health (Uher, 2016), clinical referrals, and from the community of the Halifax metropolitan area in Nova Scotia, Canada. We use broad inclusion criteria focused on the ability to provide a valid speech sample in English: age 16 or older, able to give informed consent, and native English speaker or English acquisition and fluency before 11 years of age. English language proficiency is an essential criterion. Speech produced by someone for whom English is a second or third language may introduce artifacts that appear as language deficits or psychopathology. Exclusion criteria are a neurological illness affecting speech or a language disorder.

The project plans to enroll 600 participants to create a corpus of data including at least 120 hours of audio recording. The present report describes the development of the data collection, processing and coding procedures and evaluates these procedures in the first 200 participants.

Assessment of Psychopathology

On the day of speech sampling, each participant undergoes a diagnostic interview using the Structured Clinical Interview for DSM-5 (SCID-5; First et al., 2015) that identifies current and lifetime diagnoses of mental disorders, and additional interviews and questionnaires to assess symptom severity and provide demographic and additional clinical

information. The diagnostic interviews are completed by a clinician who is not involved in obtaining the speech sample. The same clinician also completes the Mongomery and Åsberg Depression Rating Scale (MADRS), an extensively validated interview-based measure of depressive symptom severity (MADRS; Montomergy & Asberg, 1979). Depression symptoms are further assessed using the self-report Quick Inventory for Depression Symptomatology (QIDS;(Rush et al., 2003)). Manic symptoms are assessed using the Young Mania Rating Scale (YMRS;(Young et al., 1978)). Presence of mixed episode symptoms is assessed using the experimenter rated Koukopoulos Mixed Depression Rating Scale (Sani et al., 2018) and hypomanic symptoms are assessed with the self-report Hypomania Checklist (HCL-32; Angst et al., 2005). Each participant also undergoes a language assessment using the Expressive Vocabulary Test (EVT-3; Williams, 2018) to identify the presence of a language disorder or significant language impairment.

Speech Collection

Speech collection methodology has evolved over time. Before advanced transcription and analysis tools became available, structured speech collection methods were used to decrease variability in speech. These included reading a passage (Flint et al., 1992; Martínez-Sánchez et al., 2015; Moore et al., 2008) and reciting a set of numbers (Stassen & Bomben, 1991; Szabadi et al., 1976). These methods decreased the variability of speech content, but often removed nuanced individual differences. This resulted in speech correlates of illness with limited transferability to more natural settings (Cohen et al., 2016; Wang et al., 2019). With the progress of speech analysis technology, “free” or natural speech collection methods have been preferred for their similarity to clinical interactions, which increases their generalizability (Cohen & Elvevåg, 2014).

There are many tested approaches to elicit a natural speech narrative. Previous studies have used specific prompting techniques (Breznitz, 1992; Carrillo et al., 2018; Miller, 1996; Mota et al., 2017), whereas others analyzed recordings of unprompted clinical patient-psychiatrist interactions (Gideon et al., 2016; O'Dell & Winder, 1975; Rosenberg, 1979; Zimmermann et al., 2017). There is no established superiority of natural speech elicitation/collection methods, however some research has shown that speech features may vary under certain emotional conditions (Himmelstein et al., 2018). Emotionally-valanced speech prompting is used to prime participants with a certain emotion that may influence their speech patterns and behavior. This is often accomplished by prompting participants to speak about an emotionally salient time or event in their lives (Brockmeyer et al., 2015; Hong et al., 2015; Novack, 2003). These methods may be some of the most valid, as they closely mimic everyday interactions while a retaining consistent structure constant across participants.

For the present study, we require a speech collection procedure that elicits a sufficient amount of natural speech and allows manifestation of speech features relevant to a variety of mental disorders. To elicit content relevant to mood and anxiety disorders, we include prompts to elicit positive and negative autobiographic content. To facilitate the study of illness development, prognosis and treatment, we require the speech collection procedure to be repeatable with the same participant.

To meet the above requirements, we developed a procedure that uses three prompts to elicit autobiographical narratives in neutral, positive and negative emotional contexts.

Procedure

We recorded prompted speech using TASCAM DR-05X recorders. These recorders were selected due to the high sound quality from bi-directional microphones that allow the device to capture high quality sound at both high and low decibel spectrums. Experimenters calibrate each recorder to the noise specifications of the room by adjusting input level volumes before participants are assessed (see Section B of Appendix B). During the recording of the sample, participants sit approximately 2 feet away from the microphones. They are asked to speak for three minutes, three times, in response to three different prompts that ask about a neutral, positive, and negative experience in the recent past.

We developed a prompting script with the aims of maximizing natural participant speech and standardizing and minimizing experimenter speech (Section C, Appendix B). After a brief description of the procedure and giving an opportunity to ask questions, the experimenter prompts for neutral, positive and negative narratives:

Neutral prompt:

"First, I would like to hear about how your last couple of weeks have been, and how you've been spending your time. Tell me how you've been feeling and what you've been up to lately."

Positive prompt:

"Next, I want you to think about a time in the past few weeks when things went well for you. Think about when you had a positive experience or when something good may have happened to you. Take your time to think about it, and you can go ahead whenever you're ready."

Negative prompt:

"Lastly, I now want you to think about a time in the past few weeks when things didn't go well for you. Think about when you had a negative experience or when something bad may have happened to you. Take your time to think about it, and you can go ahead whenever you're ready."

Each emotionally valanced prompt has a standardized set of short, secondary prompts that are available as aids to the experimenter if initial elicitation does not produce a narrative of 3 minutes, or if a participant asks questions about the subject matter or timing of the procedure. These secondary prompts were developed over the preliminary data collection phase to address the most common problems and questions encountered with participants, keeping the variation in experimenter speech to a minimum. The full description of the speech collection procedure, including the script, is provided in Section B of Appendix B.

This procedure enables repeated assessments to capture any changes in mood or other psychopathology over time and detect related changes in speech behavior. Speech sampling was repeated after approximately 16 weeks, with variance allowed in the time interval between samples capture change in mental state or treatment effects (Figure 1). Further details of the speech collection procedures are available in Section B of Appendix B.

Evaluating success of the speech collection procedure

We assessed the success of our novel script for eliciting natural narratives from participants. We measured success as audio lengths over 10 minutes, which includes the minimum time for prompt responses (9 minutes) with the addition of experimenter prompting (1 minute). We calculated audio lengths and the proportions of participant and

interviewer speech. Additionally, we looked at the proportion of participants who were able to provide us with sufficient speech for automated and manual analysis.

Audio sample processing and coding

Information relevant for psychopathology can be extracted from audio samples of speech through coding by trained human raters or through automated computerized procedures. The two approaches have complementary advantages. Computerized procedures may be more reproducible, less prone to bias and more scalable to large number of samples. However, current computational analysis typically requires a very large number of examples and works well with segments of several seconds. Human coding may be more directly informed by clinical knowledge and better capture the context across minutes of speech. Importantly, computerized procedures require external labels to effectively learn to extract relevant information from audio recordings. To leverage the potential of human coding and computerized analysis, we have designed a procedure where speech samples are first separated into segments lasting several seconds each, and then each segment is coded for a number of relevant characteristics by trained human raters (Figure 1B). Manually rated samples can be used directly to predict clinical diagnoses and outcomes, but can also serve as segment-level labels for computerized speech analysis.

Each audio sample is electronically transcribed using Kaldi, an offline transcription software (*Kaldi: Kaldi*, 2009) and the transcription is edited for accuracy by experimenters. The transcriptions are then segmented by trained staff members. The transcriptions are divided into sentence-length segments based on changes in speaker, sentiment polarity, reference, and topic.

Based on literature review and multidisciplinary expertise, we selected features that are discernible in speech and relevant for psychopathology. Some clinical features are made up of multiple speech characteristics. For example, anxiety may manifest as fast speech, breathy sounds, inappropriate laughter or repetitive statements or dwelling on a topic (Drost et al., 2014; Kotsopoulos & Mellor, 1986; Spinhoven et al., 2018; Toazza et al., 2016). Since the relative validity of specific elementary features and composite speech characteristics is unknown, we chose a selection of simple and composite features for coding (Table 1; See Section D of Appendix B for detailed descriptions of each variable). The selected variables are manually coded at segment level by trained human raters with expertise in speech and language pathology, psychology, and psychiatry. Additional variables are coded as global impression after listening to the entire speech sample. Two thirds of the speech samples are rated by multiple raters independently to establish inter-rater reliability. We quantified interrater reliability with intraclass correlation coefficient (ICC) estimated in mixed effects models using maximum likelihood.

Coding

Once the samples are segmented, they are coded by trained raters. The majority of speech variables developed to reflect psychopathological features are coded per segment, which allows all speech parts to be described using multiple variables. Some variables are coded only once per sample. Sample level ratings are akin to global impressions that a patient may transmit to a clinician; segment-level ratings allow this impression to be further quantified. Some segment level variables also have global (sample-level) counterparts. By employing a two-leveled system, we can begin to understand the nuance behind clinical phenomenology from speech, which will contribute towards integrating objective

measures into psychiatric practice. The coding of all variables is described in Section D of Appendix B.

Dataset structure

The above described procedures result in a corpus of speech audio samples with independent external labels at two levels. (1) Diagnosis and symptom severity are established at the level of individual assessment, by clinicians who do not hear the speech sample. (2) Psychopathology-relevant features of speech are rated at the segment level based on the speech audio only, by independent raters who are blind to the clinical diagnosis and severity rating.

Results

Participants

The first 200 participants included 157 females (70%) and 43 males (30%), and their average age was 42 years old (range 21-81). The most common diagnoses were major depressive disorder (MDD, 49.5%), anxiety disorders (45.5%), and bipolar disorder (BD, 20.5 %). Just over one fourth of participants were controls (26.5%) with no mental disorder. The majority of participants with a mental disorder had more than one diagnosis (Table 2).

Speech Elicitation

The first 200 participants provided 239 speech samples. We evaluated the success of speech elicitation by measuring the cumulative length of participant speech in the audio recordings. The audios were on average 11.6 minutes long (*M (seconds)* = 699.85, *SD* = 129.69). The average proportion of participant speech was 85% (in seconds, *M* = 596, *SD* = 120.66, *Min.* = 46, *Max.* = 1312), or approximately 10 minutes, with an average of 15% experimenter speech (consisting of the script introduction and prompts). 96% of samples contained over 5 minutes of participant speech.

Speech samples were divided into three emotionally valanced segments using specific prompts. On average, responses to the neutral (first) prompt were just over 3 minutes long (in seconds, *M* = 202.11, *SD* = 55.6, *Min.* = 41, *Max.* = 688), as were responses to the second (positive) prompt (in seconds, *M* = 191.09, *SD* = 49.26, *Min.* = 0, *Max.* = 343) and the third (negative) prompt (in seconds, *M* = 202.81, *SD* = 59.04, *Min.* = 0, *Max.* = 525). A one way ANOVA revealed that responses to the positive prompt were

significantly shorter than responses to the neutral or negative prompt ($F = 5.36$, $p = 0.0001$). This difference was unrelated to diagnosis ($t = -1.04$, $p = 0.297$).

Inter-Rater Reliability

Table 2 shows interclass correlation coefficient (kappa) values for manually coded speech samples from the first 200 participants. Over two-thirds (69%) of the samples were coded by two or more raters. Raters accomplished high inter-rater reliability for sentiment, affect, reference, and emotions (ICC 0.60 to 0.79). Rater agreement on variables that were developed to detect more nuanced and sophisticated features of psychiatric illness was lower (ICC 0.24 - 0.47).

Discussion

The current paper describes a new data collection procedure and study design using speech analysis for patient assessment in psychiatric setting. The procedure was developed to elicit natural narratives from participants that would allow multi-level human coding and automated analysis of content and audio features in the assessment of psychopathology. The protocol aims to generate data that will support the development of alternative measurement of psychiatric symptoms, adding objectivity to clinical practice. The development of variables and coding practices was informed by literature review (Dikaios et al., submitted) and expertise from a multidisciplinary team. The speech elicitation procedure obtained adequate audio samples from almost all eligible participants. Audio processing methods including segmenting allowed for extensive manual and automated analysis. Human coders achieved high levels of agreement on sentiment, basic emotions and person-relatedness aspects of the natural speech at segment level.

Most participants sampled from the clinical population are patients of a mood disorders clinic. Further enrollment efforts will focus more on recruiting from the community and psychosis clinics. Participants enrolled to date represent a wide range of mood and anxiety disorders. Participant samples include individuals with comorbid diagnoses, a feature of this study that will increase the generalizability of the results to clinical environments. As the literature has established natural, unstructured speech as the most versatile (Cohen & Elvevåg, 2014), the current study uses natural speech elicited with emotionally valanced prompts to produce unmetered speech that simulates everyday conversation. The success of speech elicitation is demonstrated by substantial audio

recording lengths within emotionally polarized conditions. Minimal experimenter interruption indicates the script achieves the goal of evoking a self-sustaining narrative, which is ideal for both computerized and hand-coded analyses. As the script does not specifically prompt for dialogue about illness or symptoms, it creates conditions more closely mimicking an everyday conversation, which is less likely to be adjusted or metered by the speaker. In this way, we are able to treat the speech as an objective reflection of internal psychological processes, and therefore speech may be considered as an accessible biomarker for potential clinical implementation.

Traditional speech feature variables were amended to reflect psychiatric symptoms traditionally assessed using clinical measures or clinician judgement. In an effort to objectify illness features, new variables were developed to index symptom characteristics that are often assessed in clinical environments, but have yet to be explored using speech analysis. With the development of a novel coding system, tests of interrater reliability were performed frequently on coded speech samples to ensure agreement and mitigate drift. Many of our speech variables demonstrated moderate-strong reliability between multiple raters (McHugh, 2012).

Variables developed to detect more nuanced, abstract clinical features such as attributional style (self-criticism) and cognitive capacity (coherence, richness) yielded lower agreement between raters. There are a few potential explanations for this. First, these lower kappa values may reflect the phenomenological nature of these illness features, mirroring the difficulty clinicians may have in understanding these symptoms (Cohen et al., 2013, 2014; Monferrer et al., 2021). Alternatively, it is possible that,

although all raters received standard, comprehensive training for manual coding, different expertise and clinical backgrounds may have contributed to rater drift.

Preliminary analysis of study data has provided us with considerations for moving forward. Further recruitment efforts will focus on enrolling more men with the aim of balancing the sex distribution of the sample, as well as recruiting from clinical populations in order to sample more participants with schizophrenia spectrum disorders (Table 2). We will increase the frequency of reliability testing for all speech variables to achieve higher agreement between raters on more complex variables. The success of the speech elicitation methods, as well as high interrater reliability on foundational speech variables, highlight the strength and effectiveness of the current protocol to sample and test speech as an objective biomarker of psychiatric illness.

Tables and Figures

Table 1
Speech variable descriptions and rationale

Variable	Definition	Rationale
Sentiment Negative, Positive, Neutral	Used to identify polarity of segment.	Measuring levels of negative sentiment can help assess depression (Breznitz, 1992; Himmelstein et al., 2018).
Richness	Refers to lexical and semantic complexity of speech.	Speech richness has been used to assess depression and psychosis (Cohen et al., 2014; Corcoran & Cecchi, 2020).
Reference	Indicates whether the speaker is referencing self, other, or a relationship.	More frequent self-reference can be indicative of depression or schizophrenia (Himmelstein et al., 2018; Hong et al., 2015; Zimmermann et al., 2017).
Emotions Neutral, Sadness, Joy, Fear, Anger	Presence of emotion as indicated by speech content and acoustic tone.	Speech emotion has been linked to severity and illness improvement in depression and schizophrenia (Harati et al., 2018; Minor et al., 2015; Novack, 2003).
Coherence	Quantifies a speaker's ability to follow a logical narrative	Coherence is used to index speech disorganization in psychosis (Ayer et al., 2016; Elvevåg et al., 2007; Pauselli et al., 2018) and attention deficit or distractibility (Engelhardt et al., 2010; Mota et al., 2012; Raucher-Chéné et al., 2017).
Variable	**Definition**	**Rationale**

Worry	Speaker repeatedly returns to future-oriented narrative with anticipation of a negative outcome.	Higher levels of worry can be a predictor of negative mental health outcomes (Drost et al., 2014; Spinhoven et al., 2018; Struijs et al., 2018).
Anxiousness	Determined by acoustic features such as rate of speech, stop/restarts (stuttering), inappropriate laughter, or unusually high pitch.	Limited literature on the manifestation of anxiety in speech points to more hesitation, more variable speech rate, and more breathy sounds (Kotsopoulos & Mellor, 1986; Toazza et al., 2016).
Criticism	Including critical or demeaning statements about anything or anyone other than oneself.	Research shows higher levels of outward-directed criticism linked to depression (Breznitz, 1992).
Self-criticism	Statements containing self-critical or self-deprecating comments.	Higher levels of self-criticism linked to worse outcomes in many disorders (Löw et al., 2020).

Table 2

Demographic and diagnostic information of the first 200 participants

		N	**Percent**
N		200	-
Sex (female)		157	70.09%
Diagnosis			
	Controls*	53	26.50%
	MDD	99	49.50%
	Bipolar	41	20.50%
	Schizophrenia Spectrum	3	1.50%
	Anxiety**	91	45.50%
	OCD	11	5.50%
	ADHD	31	15.50%
		Mean	**SD**
Age		42.48	11.90
MADRS score		10.36	9.41

Note. Some individuals have more than one diagnosis, accounting for overlap

*Controls are defined as participants who do not have any of the diagnoses listed

**Anxiety includes diagnoses of generalized anxiety disorder (GAD), social anxiety, panic disorder, agoraphobia, and specific phobia

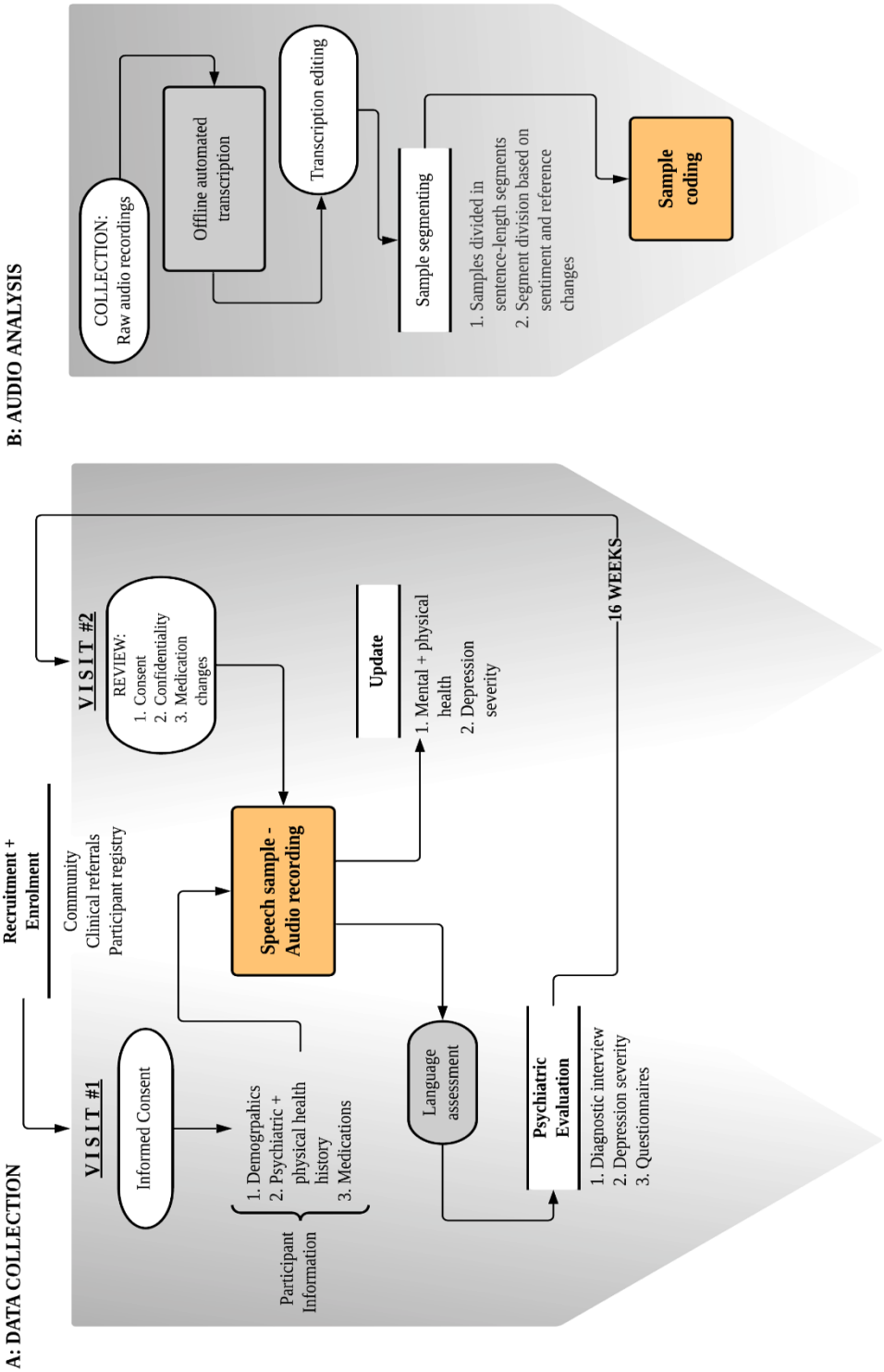
Table 3

Inter-rater reliability for speech variables

Speech variable	ICC (kappa)
Sentiment	0.79
Affect	0.68
Reference	0.61
Emotions Joy, Fear, Sadness, Anger, Neutral	0.60
Criticism	0.47
Coherence	0.37
Worry	0.32
Self-criticism	0.27
Anxiousness	0.25
Richness	0.24

Figure 1

Flow of study procedure and sample analysis



Chapter 4: Facilitating Original Data Analysis

Speech data for preliminary analysis and method validation were collected from two ongoing research projects within the Nova Scotia Health Authority. Speech analysis protocols were added to the procedures of these studies to increase biomarker assessment and facilitate study development for The Vocal Mind Project (VMP). The Canadian Depression Research and Intervention Network (CDRIN) Registry is an online registry of patients, research participants, and community members with mood and anxiety disorders. Participants of this registry can use research-validated self-rated measures to monitor mood and related symptoms. The level of assessment of these participants varies depending on their involvement in clinical care or research – those enrolled in research studies undergo diagnostic assessment using the SCID-5 (First et al., 2015) and depression severity assessment using the MADRS (Montomergy & Asberg, 1979). Families Overcoming Risks and Building Opportunities for Well-Being (FORBOW) is a longitudinal study of offspring and their parents aimed at early detection and monitoring of mental disorders. The parent sample is enriched for participants with severe mental illness (SMI; MDD, BD, and schizophrenia spectrum disorders). Adult participants of the FORBOW study undergo diagnostic assessment using the SCID-5 (First et al., 2015) and depression severity assessment (MADRS; Montomergy & Asberg, 1979).

The VMP protocol (described in Chapter 3) includes a wide range of assessments appropriate for evaluating both speech and psychopathology (specifically mood and anxiety disorders). Over the next three years, this project is set to recruit 400 participants from both clinical and community populations. Restrictions relating to the COVID-19 pandemic impacted our ability to recruit the proposed amount of VMP participants within

the first year of the study. The analyses in the following chapter were performed on a blended sample of participants from the CDRIN Registry, the FORBOW study, and the VMP. As CDRIN and FORBOW participants do not undergo formal language assessment (EVT-3; Williams, 2018) or extensive assessment of mood disorder symptoms (Quick Inventory for Depression Symptomatology (QIDS;(Rush et al., 2003)), Young Mania Rating Scale (YMRS;(Young et al., 1978)), Koukopoulos Mixed Depression Rating Scale (Sani et al., 2018), Hypomania Checklist (HCL-32; Angst et al., 2005)), we were unable to incorporate all participant assessment measures described in Chapter 3 in the following analyses. Although not all participants underwent formal language assessment, all samples were screened by a speech language pathologist for abnormalities in expressive and receptive language abilities.

As recruitment and enrollment continues for the VMP, we will have the ability to prioritize more complete investigation of how mood symptoms and language ability affect speech behavior in relation to psychopathology. The analyses in the following chapter are preliminary, and results are meant to demonstrate the utility of our methods and serve as guidance for future research.

Chapter 5: Differentiating Bipolar and Unipolar Depression Using Speech Analysis

Contributions:

KD and RU developed the concept for this manuscript. SHD, MA and MK provided guidance on methods and statistical analyses. KD analyzed the data, with statistical assistance from RU. KD wrote the initial draft of the manuscript. RU, MA, and MK reviewed the draft and provided edits. KD finalized the manuscript.

Differentiating bipolar and unipolar depression using speech analysis

Dikaios, K.[1,2], Rempel, S.[2], Dumpala SH.[3], Alda M.[1,2], Kiefte, M.[4], Uher, R.[1,2]

[1]Department of Psychiatry, Dalhousie University, Halifax, NS

[2]Nova Scotia Health, Halifax, NS

[3]Faculty of Computer Science, Dalhousie University, Halifax, NS

[4]School of Communication Sciences and Disorders, Dalhousie University, Halifax, NS

Abstract

Misdiagnosis of bipolar disorder is common, and subsequent treatment with unopposed antidepressants can increase the likelihood of negative outcomes long-term. Speech has been studied as a tool for objective assessment and shows potential as a biomarker of psychiatric illness. There are few speech analysis studies focused on bipolar disorder. The current study aimed to differentiate bipolar and unipolar depression using content features of speech in a preliminary sample. Content variables were chosen to represent features of depression, and each speech sample was manually rated by trained research staff members. Three participant groups were included in the analysis: 47 participants with bipolar disorder, 47 participants with major depressive disorder, and 47 controls. Primary logistic regression model using content speech variables showed no significant differences in speech features between unipolar and bipolar depression. In classification accuracy analysis, the same speech features were able to discriminate between bipolar and unipolar depressed participants with moderate accuracy (64.41%). Follow-up analyses showed that speech content features were more strongly associated with depression severity in the bipolar disorder group than the major depressive disorder group. Limited power will be addressed by larger sample sizes as the investigation continues. The potential utility of the analysis of psychiatric content variables is demonstrated by moderate accuracy in discriminating bipolar and unipolar depression. These results highlight the validity of speech as a biomarker of illness, and provide direction for future work in the classification of affective disorders.

Introduction

Bipolar disorder (BD) is an affective illness characterized by episodes of depression and mania or hypomania, interjected by periods of full or partial remission. Prevalence rates of bipolar disorders are estimated to be above 5%, and illness is most often accompanied by significant life impairment and increased risk of suicide (Judd & Akiskal, 2003; Merikangas et al., 2011; Moreno & Andrade, 2005). BD symptoms vary in presentation and severity, making the disorder heterogeneous and sometimes difficult to diagnose (Filakovi, 2011; Hantouche & Akiskal, 2005; Hirschfeld, 2014). The delay of accurate diagnosis, estimated to be 5-10 years (Diler et al., 2017), can prolong the course of BD, allowing for more negative long-term outcomes (Bowden, 2005). Currently, the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) necessitates the history or presence of at least one manic or hypomanic episode in a patient's life to grant a diagnosis of bipolar disorder (American Psychiatric Association, 2013), however, it's been estimated that over 50% of BD cases begin with a depressive episode (Chang, 2009). Bipolar depressive episodes lack consistent symptomatic differences from unipolar depressive episodes (Cuellar et al., 2005); 25-69% of patients with BD are initially misdiagnosed with major depressive disorder (MDD), and subsequently treated for unipolar depression (Birmaher, 2013; Bowden, 2005; Ghaemi, 2000; Gosek et al., 2019). Treatment with antidepressant medications for individuals with BD can trigger manic episodes, initiate rapid cycling, and may increase treatment resistance to subsequent medication trials (Altshuler et al., 1995; Bowden, 2001; Fava, 2020; Post, 2005; Wehr et al., 1988).

Research is exploring the early detection of BD using clinical, behavioral, and genetic antecedents. As the presentation of depression can be heterogeneous, the assessment of depressive symptoms alone is often unreliable for detecting BD (Brockington et al., 1982; Goldberg, 2011). Objective biomarkers are being studied extensively to aid in diagnostic accuracy (Fernandes et al., 2020; Huang et al., 2020; Poletti et al., 2021; Vai et al., 2020; Zheng et al., 2019). One of these biomarkers is speech. Natural speech can act as an objective marker of psychiatric illness and is easy to collect and analyze. Acoustic, semantic, and content features of speech have been shown to classify psychiatric diagnoses with high accuracy (Elvevåg et al., 2010; Minor et al., 2015; Oxman et al., 1988; Solomon et al., 2015; Wang et al., 2019; Willits et al., 2018).

Speech analysis literature in psychiatric research has focused on the study of depression and schizophrenia (Dikaios et al., submitted). Studies on BD are limited, and are mostly focused on assessing mood state and detecting mood changes (Faurholt-Jepsen et al., 2016; Gideon et al., 2016; Pan et al., 2018; Vanello et al., 2012; J. Zhang et al., 2018). These studies address issues of early intervention within bipolar illness for the purposes of minimizing negative outcomes of manic and depressive episodes, however they do not address the clinical task of initial, accurate diagnosis. Samples of participants with BD have been included in studies of semantic speech features for differentiating between psychotic disorders, but these have focused only on manic states of bipolar illness (R. E. Hoffman, 1986; Mota et al., 2012). The task of differentiating unipolar and bipolar depression has only recently emerged in the literature; only one study was identified. Su and colleagues used acoustic features and machine learning methods to differentiate between affective disorders and achieved moderate-high classification accuracy (2020). It

has been shown that results obtained using acoustic features can vary widely depending on both task (Cohen et al., 2016) and technology (Gideon et al., 2016), highlighting the potential for the unique utility of content analysis in psychiatric settings. Differentiation of unipolar and bipolar depression has not been attempted using semantic or content features, so it is unknown if these types of features may offer better classification, or if the addition of these features could increase diagnostic classification accuracy in a combined model with acoustics.

As results differentiating affective illnesses are limited, it may be possible to make inferences to speech from studies of clinical correlates of BD. Research has identified depressive symptoms that, considered in combination, have the ability to increase the accuracy of initial BD diagnoses (Benazzi et al., 2002; Bowden, 2005; Diler et al., 2017; Gosek et al., 2019). Although some results pertaining to which depressive symptoms might be differentiating are conflicting, studying the potential for diagnostic classification may advance the understanding of the bipolar spectrum (Angst et al., 2018; Benazzi et al., 2002; Mondimore, 2005). Many of the same depressive symptoms explored in these studies have been described in speech literature, used to study unipolar depression. For instance, acoustic features index affective blunting, or flat affect in depression (Moore et al., 2008, 2004; Novack, 2003; Quatieri & Malyska, 2012; Stassen & Bomben, 1991). Certain aspects of speech content can illustrate cognitive biases in depression such as strong personal association with negative life events (Breznitz, 1992; Mete, Schnurr, et al., 1993) or self-focused thinking (Himmelstein et al., 2018). Characterizing these common symptoms using speech variables can allow more objective

analysis of depressive features and may provide a more accurate assessment of both unipolar and bipolar depressive illness.

Using content features to characterize depressive illness may increase our understanding of many symptoms that are traditionally assessed phenomenologically. The nature of speech content is such that its analysis can reveal important insights into an individual's internal state, above and beyond the words that are said (Pennebaker et al., 2003; Tausczik & Pennebaker, 2010). In the context of BD, this understanding may be of much greater importance due to the potential for negative outcomes in the case of misdiagnosis. While clinical self-report and interview measures to assess bipolarity exist (Feng et al., 2017; Hantouche & Akiskal, 2005; X. Zhang et al., 2021), they are underused due to time constraints and perceived burden on ill patients, and rely heavily on retrospective recall from patients that may lack insight or capacity at the time of assessment (Pavlova & Uher, 2020a). Speech analysis could provide a faster, more objective method of assessing for bipolarity. Specifically, content features may have unique power to represent psychopathological features in a nuanced way.

The current study aims to assess whether a model of combined speech content variables can differentiate between unipolar and bipolar depression. The variables were developed specifically to index depressive symptoms.

Methods

Novel speech methods were developed with the aim of eliciting a natural narrative from participants with minimal experimenter prompting or interruption. Methods for the speech collection procedure are explained in further detail in Chapter 3.

Participants

Participants were included if they were 18+ years old and spoke fluent English (see Chapter 3 for fluency criteria). Participants were excluded if they were currently experiencing a manic or hypomanic episode, had a diagnosed neurological illness or a history of severe concussion or traumatic brain injury with symptoms lasting more than a few weeks. Participants were excluded if they had a previously documented language disorder, significant receptive or expressive language difficulties (assessed by a Certified Speech Language Pathologist), or if English was not their first language. All participants were assessed for psychopathology using the SCID-5 (First et al., 2015). Participants with a diagnosis of a schizophrenia spectrum disorder were excluded from the sample. Participants were divided into three groups: MDD, BD, and controls. The BD group included participants with diagnoses of either bipolar disorder I (BDI) or bipolar disorder II (BDII). It has been shown that individuals with BD II also experience significant impairment (Moreno & Andrade, 2005). Additionally, research suggests that individuals with BDII are at a higher risk of being misdiagnosed with MDD than those with BDI, as the presence of hypomania can be much more difficult to detect than mania, and the depressive features more similar to MDD than BDI (Hantouche & Akiskal, 2005). We included individuals with BDII in our sample to increase the generalizability of results. The control group consisted of participants with no major mood disorder.

Speech Collection

We collected 10–15-minute speech samples from each participant. Speech was elicited using an emotionally valanced prompting script (Section C of Appendix B). The script was developed with the aim of effectively eliciting an uninterrupted, natural narrative about the participants' lives while minimizing and standardizing experimenter speech within the recording. Participants were asked to speak for three minutes in response to three questions; the first prompted for a neutral autobiographical narrative, the second for a negative narrative, and the third for a positive narrative.

Assessment

All interviews were conducted by trained research staff and graduate students, under the supervision of a clinical psychologist and a psychiatrist. Demographic information and physical and mental health history were collected from all participants at the beginning of each assessment. The presence or absence of a psychiatric disorder was determined using the Structured Clinical Interview for the DMS-5 (SCID-5; First et al., 2015). All diagnoses were confirmed in a consensus meeting with a psychiatrist or psychologist. Depression symptoms were assessed using the experimenter-rated Montgomery and Asberg Depression Rating Scale (MADRS; Montomergy & Asberg, 1979).

Variable Selection

The methodology of the current study was developed to address gaps in published speech analysis research. We selected speech features based on clinical relevance, published evidence (Dikaios et al., submitted) and adequate interrater reliability (McHugh, 2012). Variables with an intraclass correlation coefficient less than 0.20, representing no

agreement between raters, were excluded from the analysis (See Chapter 3 for more detailed information on speech variables and reliability).

As speech analysis literature for BD is limited, speech features were selected based on a synthesis of results from speech analysis research and clinical research on differentiating between unipolar and bipolar depression. We completed a brief literature review to understand the differences between clinical presentations of depression in MDD and BD (Table 1). Speech variables were selected to optimize the representation of the symptoms most often specified in the literature. Rationale for the selection of each speech variable for the current study is presented below, organized by the symptoms they were selected to represent. See Section E of Appendix B for full descriptions of each variable. Speech feature variables were used to code segmented transcripts of audio samples from each participant. Further information on speech sample processing and feature coding is outlined in Chapter 3.

The following symptoms are categorized by whether the literature has indicated their differential presentation between MDD and BD (directional) or not (exploratory) (Table 1). Variables used to index directional symptoms formulate the primary hypothesis. Exploratory analyses will be undertaken to investigate non-directional symptoms with the respective variables indicated.

Directional:

Attention Deficit

Literature suggests that attention deficits are more prevalent in BD (Table 1). Attentional problems have often been used as prodromal antecedent markers for the development of the illness (Meyer et al., 2004). Measures of verbal fluency have been used to test

attention deficit in BD (Raucher-Chéné et al., 2017). In the current study, attention deficit and distractibility were explored using *coherence* and *richness*.

Racing thoughts/flight of ideas

One hallmark symptom of bipolar mania is racing thoughts or flight of ideas. It has been suggested that these features of mania may derive from the same neuropsychological deficit as the attention, fluency, and executive functioning deficiencies in bipolar depression (Gruber et al., 2007; Malhi et al., 2007). We posited that the same variables used to index attention deficit may work to characterize the underlying neuropsychological effects that cause various types of cognitive deficits in BD (*coherence* and *richness*).

Negative emotions

An increase in negative emotions has been consistently shown to be more prevalent in unipolar depressive illness than bipolar (Table 1). We included the emotion variables *anger* and *sadness*, as well as the *sentiment* variable into the primary analysis to index this illness feature.

Guilt

Guilt, or self-directed blame, is a hallmark symptom of depression, and may increase the likelihood of suicidal ideation (Kealy et al., 2021; Kim et al., 2011). Existing literature suggests this symptom is more commonly exhibited in unipolar depression (Forty et al., 2008; Gosek et al., 2019; Hantouche & Akiskal, 2005). We modelled guilt with the variable *self-criticism*.

Exploratory:

Agitation

The presence of agitation in depression has been implicated in both unipolar and bipolar diagnoses (Table 1). For this reason, it has been included in the exploratory analyses for the current study. Swann and colleagues (2013) proposed different operationalization of agitation for MDD and BD. They suggested that agitation in MDD presents as "painful inner tension", whereas agitation in BD presents as "goal directed activity" (Swann, 2013). Definitions of psychomotor agitation have fluctuated across time and literature, but generally implicate both physical and mental processes (Day, 1999). In speech, agitation has been explored using primarily acoustic variables (Alpert et al., 2001; Novack, 2003). In the current study, agitation for the differentiation of bipolar and unipolar depression was modeled using *coherence*, *richness*, and *anxiousness*.

Psychomotor Retardation/Affect

Psychomotor retardation is more often referenced as a feature characteristic of bipolar depression; however, some results are incongruent (Table 1). Psychomotor retardation is characterized by physical and mental slowing and changes in affective modulation (Flint et al., 1993; G. M. Hoffman et al., 1985). Affective blunting, or the presence of a flat affect, often occurs as a product of psychomotor retardation (Cummins, Epps, & Ambikairajah, 2013). Semantic speech correlates of affective flattening are exhibited as monotonous speech with a lack of vocal inflection. In the current study, the presence of flattened affect was explored using the *affect* variable.

Anxiety

The presence of anxiety in depression is common. It is unclear whether anxiety better characterizes unipolar or bipolar depression (Table 1). We measured the presence of anxiety in speech using the variables *worry* and *anxiousness*.

Irritability

Table 1 shows that irritability has been found to be a feature of both bipolar and unipolar depressive disorders. Studies have defined irritability as persistent anger, sometimes with outbursts and blaming or criticism of others (Benazzi & Akiskal, 2005; Perlis et al., 2004). To characterize irritability, we used *anger* and *criticism* variables.

Statistical Analyses:

Hypothesis

Symptoms listed above under the directional category compose the primary hypotheses: Select speech features modelled together will significantly discriminate between bipolar and unipolar depression.

a. Coherence
b. Anger
c. Sadness
d. Sentiment
e. Self-criticism

Exploratory Analyses:

The ability of speech features to index the symptoms in the previous section identified as "exploratory" will be tested using the respective variables indicated.

Data analysis:

All analyses were performed using the statistical software STATA 16.1. Descriptive statistics were used to explore participant demographics and any differences between groups were assessed using Pearson's chi squared test for homogeneity or one-way ANOVAs. A logistic regression model was used to test the primary hypothesis,

differentiating unipolar and bipolar depression using selected speech variables. Classification accuracy of the primary analysis was tested using post-estimation outcomes. Statistical significance for the primary hypothesis was defined as a two-tailed p-value<0.05.

Logistic regressions were used for exploratory analyses. Linear regression models were used as follow-up analyses to test associations between depression severity and select speech features in BD and MDD groups. Statistical significance for exploratory and follow-up analyses was adjusted to p-value<0.001 to account for multiple testing, determined by taking a conservative estimate of the traditional p<0.05 divided by the number of additional tests.

Sensitivity analyses were completed to control for the effects of psychiatric medications and comorbid anxiety, obsessive compulsive disorder (OCD), and attention deficit/hyperactivity disorder (ADHD) on the primary model. We included these variables of interest individually as covariates in the primary model, to see if they impacted the results. Age and sex were included as fixed effects covariates in all analyses.

Results

Participants

One participant was excluded because of history of a traumatic brain injury with symptoms lasting more than a few weeks. Two participants were excluded due to significant receptive and expressive language difficulties. Two participants were excluded due to having English as a second language without sufficient proficiency. No participants presented with significant mixed or hypomanic symptoms.

The final sample consisted of 141 participants: 47 participants in the BD group, 47 participants in the MDD group, and 47 participants in the control group. Participants from the control group were hand-matched to those in the BD based on age and sex. MDD participants were matched to BD participants according to age, sex, and depression severity level.

Participants were grouped based on depression severity as scored on the MADRS: MADRS<7 indicates the absence of depression; MADRS 7-19 indicates mild depression; MADRS 20-35 indicates moderate depression; MADRS>35 indicates severe depression (Muller et al., 2003). Table 2 shows the number of participants in each depression severity category per diagnosis group. Participant groups did not differ in terms of age or sex (Table 3). Participants in the MDD group had significantly more anxiety than those in the control group (χ^2=8.57, p=0.014; Table 3). The groups did not differ significantly on rates of ADHD or OCD. The MDD and BD groups did not differ significantly with respect to depression severity (F=0.70, p=0.8421). Participants were mostly female (65%) and had an average age of 42 (SD=12.370).

Groups differed significantly on rates of some medications (Table 3). Significant group differences on medications were expected based on diagnosis. Participant groups (including controls) did not differ significantly on any psychiatric medication use (χ^2=1.713, p=0.425).

Selected speech characteristics differentiating bipolar and unipolar depression

A binomial logistic regression model was used to test the ability of selected speech features to differentiate between bipolar and unipolar depression (Table 4). The model was not significant (pseudo R^2=0.084, p=0.4407); the selected speech features did not account for a significant amount of the variance between bipolar and unipolar depression. Testing the classification accuracy of the regression model showed that the model correctly classified 64.41% of depressed participants as either having a BD or MDD diagnosis; model sensitivity was 60%, specificity was 68.97%.

Another binomial logistic regression was used to test whether the speech variables could differentiate between affective diagnoses regardless of the presence of a major depressive episode (MDD and BD; Table 5). This model included participants with BD and MDD diagnoses that fell under the "no depression" group as defined by MADRS score. That model was not significant (pseudo R^2=0.022, p=0.8953); the selected speech features did not account for a significant amount of the variance between BD and MDD. The model showed low accuracy, correctly classified 56.38% of participants into their diagnostic groups with 57.45% sensitivity and 55.32% specificity.

We performed a logistic regression to see if the same speech features could differentiate between participants with a diagnosis of BD or MDD and controls (Table 6). That model was not significant (pseudo R^2=0.059, p=0.1538). The selected speech features did not

significantly differentiate between unipolar and bipolar depression, or between affective diagnoses and controls.

Exploring additional speech characteristics in differentiating bipolar disorder from depression

Agitation

We tested differences in agitation between unipolar and bipolar depression using binomial logistic regression (Table 7). *Coherence*, *richness*, and *anxiousness* were included in the model, along with age and sex as fixed effects covariates. The model was not significant (pseudo R^2=0.076, p=0.2874).

Affect

The role of affect for differentiating between unipolar and bipolar depression was tested using logistic regression. The model was not significant (R^2=0.024, p=0.5742; Table 8). *Affect* did not account for a significant amount of the variance in speech parameters between bipolar and unipolar depression.

Irritability

Irritability was modeled using *anger* and *criticism* in a logistic regression (Table 9). The model was not significant (pseudo R^2=0.027, p=0.6917).

Anxiety

Differences in anxiety symptoms between bipolar and unipolar depression were tested using logistic regression (Table 10). *Worry* and *anxiousness* were included in the model. The model was not significant, pseudo R^2=0.1166, p=0.0491.

Associations between selected speech variables and depression severity

We used multiple linear regression tests to examine associations between depression severity scores (MADRS) and the speech variables that were included in the primary analysis within each affective diagnosis group (BD and MDD; Table 11). Each model included age and sex as covariates.

Speech *coherence* did not account for a significant amount of the variance in depression severity in either the BD group ($R^2 = 0.0603$, $F(3, 44) = 0.88$, $p = 0.4606$) or the MDD group ($R^2 = 0.056$, $F(3, 44) = 0.86$, $p = 0.4680$). *Anger* did not account for a significant amount of the variance in depression severity in the BD group ($R^2 = 0.128$, $F(3, 44) = 2.01$, $p = 0.1281$) or the MDD group ($R^2 = 0.052$, $F(3, 44) = 0.81$, $p = 0.4951$). Notably, speech *anger* was correlated with depression severity more strongly in the BD group ($\beta = 2.129$, $t(44) = 1.79$, $p = 0.081$) than in the MDD group ($\beta = 0.247$, $t(44) = 0.21$, $p = 0.831$), although the test was not significant.

Sadness in speech accounted for a marginally significant amount of the variance in depression severity in the BD group ($R^2 = 0.308$, $F(3, 44) = 6.09$, $p = 0.0016$) at the p-value<0.001 level. *Sadness* was more strongly correlated with depression severity in the BD group ($\beta = 3.786$, $t(44) = 3.83$, $p < 0.001$) than in the MDD group ($\beta = 2.973$, $t(44) = 3.06$, $p = 0.004$).

Negative sentiment accounted for a significant amount of the variance in depression severity in both the BD ($R^2 = 0.207$, $F(3, 44) = 3.57$, $p = 0.0220$) and MDD groups ($R^2 = 0.203$, $F(3, 44) = 3.74$, $p = 0.0178$) at the p-value<0.05 level, however significance did not survive p-value corrections. Strength of correlation between *negative sentiment* and depression severity did not differ between BD ($\beta = 3.42$, $t(44) = 2.76$, $p = 0.009$) and MDD groups ($\beta = 3.026$, $t(44) = 2.89$, $p = 0.006$).

Self-criticism in speech did not account for a significant amount of the variance in depression severity in either the BD ($R^2 = 0.123$, $F(3, 44) = 1.91$, $p = 0.1424$) or MDD groups ($R^2 = 0.053$, $F(3, 44) = 0.82$, $p = 0.4883$). Although not significant, the correlation between *self-criticism* and depression severity was stronger in the BD group ($\beta = 1.773$, $t(44) = 1.71$, $p = 094$) than in the MDD group ($\beta = 0.324$, $t(44) = 0.29$, $p = 0.776$). We performed linear regressions including all speech variables to understand their combined contribution to the variance in depression severity in each group (Table 12). Speech features from the primary hypothesis model accounted for 39% of the variance in depression severity in the BD group ($R^2 = 0.393$, $F(7, 40) = 3.43$, $p = 0.0064$), which trended towards significance at the p-value<0.001 level. Speech features explained 27% of the variance in depression severity in the MDD group ($R^2 = 0.272$, $F(7, 40) = 2.14$, $p = 0.0613$), which was non-significant at the p-value<0.001 level.

Sensitivity analysis

Sensitivity analyses were performed to control for the effects of medications and comorbid diagnoses in the primary and significant models. It has been suggested that comorbid diagnoses may affect speech parameters for certain disorders (Sonnenschein et al., 2018), and few studies included participants with multiple diagnoses (Dikaios et al., submitted). Medications were included to account for possible sedative or arousing effects that could influence speech production.

Psychiatric comorbidities and medications were included as covariates in the primary logistic regression model and explored individually. Inclusion of these effects did not change the results of the model.

We further explored the significant associations between depression severity and speech variables by controlling for effects of medications and comorbid diagnoses. Medications and comorbid diagnoses had no effect on the significant association between depression severity and *sadness* in speech.

Statistical Power

We completed a post-hoc power calculation to understand our ability to have detected results in the proposed sample (100 participants per group), contrasted with our ability to detect results in the current sample (47 participants per group). We will use this information going forward as we continue to recruit and enroll participants for this and related speech analysis investigations.

The power analysis performed on the primary logistic regression model showed that, with 47 participants per group, we were underpowered to detect significant differences between unipolar and bipolar depression (Table 13). Based on the accuracy of prediction identified in our primary model, a sample of 100 participants per group will be required to provide sufficient power to determine whether the selected speech features differentiate bipolar from unipolar depression at alpha level of 5% (Table 12).

Discussion

The current study investigated speech differences between bipolar and unipolar depression using speech content features. The differentiation between bipolar and unipolar depression has not been previously addressed. We aimed to look at the potential for the analysis of psychopathology content variables to provide objectivity in the differentiation of bipolar and unipolar depression in our preliminary dataset. Speech variables for the current study were developed to index important symptoms of depression that have been identified in the literature as being potentially discriminatory between BD and MDD. We hypothesized that speech features selected based on published findings discriminating BD from MDD would be able to differentiate between unipolar and bipolar depression.

Our primary model did not significantly differentiate bipolar from unipolar depression in our sample. Classification accuracy was 64.41%. This negative result may be a combination of a relatively weak signal and less than adequate statistical power. Statistical power is an important consideration, especially in the context of heterogeneity of the disorders under study and the current sample in terms of depression severity, comorbid diagnoses, and psychiatric medication use. A post-hoc power analysis suggested that we were underpowered to detect significant results in our primary model. Due to restrictions brought on by the COVID-19 pandemic, we were unable to enroll the number of participants originally proposed for this preliminary investigation.

Samples from identified speech analysis studies using content features have ranged from 12-100+ participants (Dikaios et al., submitted). Most of these studies have used simple, and in some cases, automated, content analysis methods, potentially explaining their

ability to detect significant results with lower numbers of participants. Future analyses will benefit from continuing recruitment from clinical and population samples to increase our power to detect these differences.

Results from analyses with depression severity scores allowed us to understand the findings of the power analysis. Speech features from the primary model accounted for 39% of the variance in depression severity scores in the BD group. This finding suggests potential for our content analysis methods to detect important differences between affective disorders.

Testing for correlations of individual speech features with depression severity showed revealed notable findings. Although the significance of some results did not survive p-value corrections, they highlight potential differences that could be better evaluated with higher power. *Anger* correlated with depression severity more strongly in the BD group than the MDD group, although these results were non-significant. These results may be indicative of effects that could be better understood with higher power, however current interpretation must be understood to be speculative. Literature on clinical correlates of depressive disorders suggests that negative emotions are more characteristic of unipolar depression (Batmaz et al., 2013; Fierro et al., 2016). However, studies that have characterized negative emotion as a feature of MDD have defined it as sadness (Fierro et al., 2016) and frequency of negative cognitions (Batmaz et al., 2013). In the present study, *anger* was coded to represent a wider range of negative emotions with arousal, including irritation and hostility (Section E, Appendix B). Some studies have demonstrated that symptoms such as agitation and irritability may better characterize BD (Benazzi et al., 2002; Diler et al., 2017; Perlis et al., 2004). This finding suggests that

it may be more difficult than originally thought to parse out the differences between emotional anger and anger as a feature of agitation in clinical populations (Swann, 2013). More thorough assessment of agitation and irritability in all participants may allow us to better explain this difference in the future.

Sadness and *self-criticism* correlated with depression severity in the BD group more strongly, although these differences were not significant. It is possible that higher power would allow these differences to come through more strongly, however these trends may not be a true representation of the differences present between BD and MD. Published evidence often characterizes the presence of sadness and self-criticism (guilt) as characteristic of unipolar depression (Table 1); it is possible that indexing these symptoms with content speech variables is revealing differences between unipolar and bipolar depression not yet encountered using traditional measures. Some evidence suggests that speech analysis may have the ability to uniquely define phenomena in psychiatric illness that may be muddied by traditional symptom measures and subjective assessment by clinicians (Alpert et al., 1995, 2000; Cohen et al., 2008). Higher power would allow us to interpret results with more certainty.

Our models were fit onto a small preliminary dataset of participants. As we used a traditional statistical approach to identify speech features that discriminated between groups, the potential of overfitting of models must be considered. Speech data is complex and effects of gender, race, age, education, and other demographic variables can influence speech production, potentially adding noise to the dataset. The size of the dataset did not allow for the incorporation of training and validating analyses. Without

being able to assess the relative generalizability of our results in a larger dataset, we treat the present results as preliminary.

We used participant group matching to control for the possibly confounding effects of age, sex, and depression severity on our results, in an attempt to detect only speech effects related to psychopathology. Although group matching may mitigate some noise in our dataset, there are implications to consider. Matching the number of participants in the control group to that of the experimental groups limits the range of speech feature variability that defines "unaffected" participants. This may increase the likelihood of committing Type I error. Increasing the number of participants overall will allow us to structure our dataset in a manner more representative of the sampling population.

The present study benefits from thorough assessment of psychopathology for all participants, verified by mental health clinicians specializing in diagnosis and treatment of affective disorders. Study methods were developed with reference to literature from the fields of psychiatry, speech science, and computer science and machine learning (Dikaios et al., submitted), allowing us to incorporate established speech analysis norms and address gaps present in the work published thus far. Preliminary results show promise for the utility of novel speech content variables in the assessment of bipolar depression. The capability of these variables to differentiate unipolar and bipolar depression will be better assessed in a larger sample. The investigation is ongoing, and further recruitment and enrollment will increase statistical power and provide us with a clearer picture of important differences between unipolar and bipolar depression in the future.

Tables and Figures

Table 1

Synthesized results of studies investigating clinical correlates of bipolar and unipolar depression

Symptom	Bipolar depression	Unipolar depression	No difference/mixed
Agitation	(Benazzi, 2007; Benazzi et al., 2002, 2004; Diler et al., 2017; Hantouche & Akiskal, 2005; Hirschfeld, 2014)	(Bowden, 2005; Cuellar et al., 2005; Filakovi, 2011; Galvão et al., 2013; Gosek et al., 2019)	(Swann, 2013)
Psychomotor Retardation/Affect	(Bowden, 2005; Fierro et al., 2016; Filakovi, 2011; Mondimore, 2005; Motovsky & Pecenak, 2013)	(Hantouche & Akiskal, 2005)	(Cuellar et al., 2005; Galvão et al., 2013)
Anxiety	(Diler et al., 2017; Galvão et al., 2013; Hirschfeld, 2014)	(Bowden, 2005; Cuellar et al., 2005; Gosek et al., 2019; Nuñez et al., 2018; Scott et al., 2013)	(Galvão et al., 2013; Hantouche & Akiskal, 2005)
Attention deficit	(Diler et al., 2017; Gosek et al., 2019)	-	-
Negative emotions (sadness, anger)	-	(Batmaz et al., 2013; Bowden, 2005; Diler et al., 2017; Fierro et al., 2016)	(Cuellar et al., 2005)
Racing thoughts, flight of ideas	(Diler et al., 2017; Hirschfeld, 2014; Scott et al., 2013)	-	-

Symptom	Bipolar depression	Unipolar depression	No difference/mixed
Guilt	(Motovsky & Pecenak, 2013)	(Forty et al., 2008; Gosek et al., 2019; Hantouche & Akiskal, 2005)	(Cuellar et al., 2005)
Self-reference	-	(Himmelstein et al., 2018; Zimmermann et al., 2017)	-
Irritability	(Benazzi & Akiskal, 2005; Perlis et al., 2004)	(Diler et al., 2017; Perlis et al., 2009)	-

Table 2

Depression severity stratification by participant group

Severity	Controls	MDD	BD	χ^2	*p*-value
No depression	39	18	17	25.4596	<0.001
Mild depression	8	21	24	10.6824	0.005
Moderate depression	-	8	6	9.2842	0.010
Severe depression	-	-	-	-	-

Note. No depression = MADRS<7; mild depression = MADRS 7-19; moderate depression = MADRS 20-35; severe depression = MADRS 35+

Table 3

Participant demographics

		Controls	Unipolar depression (MDD)	Bipolar depression (BD)	F	*p*-value
n		47	47	47	-	-
Age (mean (SD))		41.43 (10.98)	43.04 (13.09)	42.53 (13.08)	0.58	0.9751
					χ^2	*p*-value
Females (%)		29 (61.70%)	32 (68.09%)	28 (59.57%)	0.46	0.796
Other diagnoses (n (%))	Anxiety[1]	11 (23.40%)	25 (53.19%)	21 (44.68%)	8.57	0.014
	ADHD	5 (10.64%)	8 (17.02%)	4 (8.51%)	1.51	0.470
	OCD	1 (2.13%)	3 (6.38%)	5 (10.64%)	1.15	0.563
Psychiatric medications (n (%))	Antidepressants	2 (4.26%)	13 (28.26%)	6 (13.64%)	10.15	0.006
	Anticonvulsants	1 (2.13%)	3 (6.52%)	10 (22.73%)	13.82	0.001
	Antipsychotics	1 (2.13%)	1 (2.17%)	13 (29.55%)	24.86	<0.001
	Lithium	0	0	19 (43.18%)	52.69	<0.001
	Benzodiazepines	3 (6.38%)	3 (6.52%)	13 (29.55%)	11.06	0.004
	Stimulants	0 (0%)	2 (4.35%)	0 (0%)	4.14	0.126

[1] Anxiety encompasses diagnoses of social anxiety, generalized anxiety disorder, specific phobia, agoraphobia, and panic disorder

Table 4

Binomial logistic regression results using speech features to differentiate between unipolar and bipolar depression

Full model:
Pseudo R^2=0.0842, p=0.4407

Speech variables*	Odds Ratio	Standard Error	z	p-value	95% Confidence Interval	
					LL	*UL*
Coherence	1.065	0.357	0.19	0.851	0.552	2.056
Anger	1.519	0.458	1.39	0.165	0.842	2.743
Sadness	1.121	0.388	0.33	0.743	0.568	2.210
Negative sentiment	0.480	0.225	-1.57	0.117	0.192	1.202
Self-criticism	1.548	0.457	1.48	0.139	0.868	2.762
Age	0.997	0.020	-0.18	0.861	0.958	1.036
Sex	0.699	0.449	-0.56	0.577	0.199	2.460

Note. All speech variables (*) have been standardized

Table 5

Binomial logistic regression results using speech features to classify between MDD and BD

Full model:
Pseudo R^2=0.0221, p=0.8953

Speech variables*	Odds Ratio	Standard Error	*z*	*p*-value	95% Confidence Interval	
					LL	*UL*
Coherence	1.158	0.256	0.67	0.506	0.751	1.786
Anger	1.015	0.244	0.06	0.951	0.633	1.626
Sadness	0.847	0.233	-0.60	0.547	0.494	1.454
Negative sentiment	0.865	0.275	-0.46	0.648	0.463	1.614
Self-criticism	1.097	0.232	0.44	0.662	0.725	1.659
Age	0.994	0.016	-0.38	0.702	0.962	1.026
Sex	0.761	0.348	-0.60	0.551	0.311	1.866

Note. All speech variables (*) have been standardized

Table 6

Multinomial logistic regression results using speech features to classify experimental group participants (BD and MDD) from controls

Full model:
Pseudo R^2=0.0594, p=0.1538

Speech variables*	Odds Ratio	Standard Error	*z*	*p*-value	95% Confidence Interval	
					LL	*UL*
Coherence	0.930	0.179	-0.37	0.708	0.638	1.356
Anger	0.691	0.155	-1.64	0.101	0.445	1.074
Sadness	1.227	0.551	0.65	0.513	0.665	2.264
Negative sentiment	1.786	0.551	1.88	0.060	0.975	3.270
Self-criticism	1.035	0.234	0.15	0.878	0.665	1.612
Age	1.011	0.016	0.65	0.513	0.979	1.043
Sex	0.886	0.357	-0.30	0.764	0.402	1.951

Note. All speech variables (*) have been standardized

Table 7

Binomial logistic regression results using speech features indexing agitation to differentiate between bipolar and unipolar depression

Full model:
Pseudo R^2=0.0758, p=0.2874

Speech variables*	Odds Ratio	Standard Error	z	p-value	95% Confidence Interval	
					LL	*UL*
Coherence	1.168	0.375	0.48	0.628	0.623	2.190
Richness	1.467	0.488	1.15	0.248	0.765	2.811
Anxiousness	0.656	0.220	-1.26	0.207	0.340	1.264
Age	0.988	0.019	-0.62	0.538	0.951	1.026
Sex	0.454	0.278	-1.29	0.197	0.136	1.510

Note. All speech variables (*) have been standardized

Table 8

Binomial logistic regression results using affect to differentiate between bipolar and unipolar depression

Full model:
Pseudo R^2=0.0244, p=0.5742

Speech variables*	Odds Ratio	Standard Error	z	p-value	95% Confidence Interval	
					LL	*UL*
Affect	0.887	0.241	-0.44	0.660	0.521	1.510
Age	0.993	0.019	-0.39	0.694	0.957	1.030
Sex	0.509	0.297	-1.16	0.247	0.163	1.595

Note. All speech variables (*) have been standardized

Table 9

Binomial logistic regression results using variables indexing irritability to differentiate between bipolar and unipolar depression

Full model:
Pseudo R^2=0.0274, p=0.6917

Speech variables*	Odds Ratio	Standard Error	*z*	*p*-value	95% Confidence Interval	
					LL	*UL*
Criticism	0.783	0.393	-0.49	0.625	0.293	2.093
Anger	1.374	0.682	0.64	0.522	0.520	3.633
Age	0.989	0.020	-0.56	0.573	0.951	1.028
Sex	0.482	0.281	-1.25	0.211	0.154	1.511

Note. All speech variables (*) have been standardized

Table 10

Binomial logistic regression results using variables indexing anxiety to differentiate between bipolar and unipolar depression

Full model:
Pseudo R^2=0.1166, p=0.0491

Speech variables*	Odds Ratio	Standard Error	z	p-value	95% Confidence Interval	
					LL	*UL*
Worry	1.953	0.629	2.08	0.038	1.039	3.670
Anxiousness	0.569	0.205	-1.57	0.117	0.281	1.152
Age	0.985	0.020	-0.74	0.461	0.946	1.025
Sex	0.424	0.259	-1.40	0.160	0.128	1.403

Note. All speech variables (*) have been standardized

Table 11

Linear regression results showing selected speech features predicting depression severity in BD and MDD groups

Speech variables*	Prediction in BD		Prediction in MDD	
	β	95% CI	*β*	95% CI
Coherence	0.122	-2.250, 2.495	-0.618	-3.452, 2.216
Anger	2.129	-0.277, 4.535	0.247	-2.078, 2.574
Sadness	3.786	1.791, 5.781	2.973	1.013, 4.932
Negative sentiment	3.424	0.917, 5.931	3.026	0.918, 5.134
Self-criticism	1.773	-0.318, 3.864	0.324	-1.955, 2.903
Full model	R^2	*p*-value	R^2	*p*-value
	0.393	0.0064	0.272	0.0613

Note. All speech variables (*) have been standardized

Table 12

Power analysis for primary logistic regression

Power	N
0.60	46
0.65	51
0.70	57
0.75	63
0.80	70
0.85	79
0.90	91

Chapter 6: Discussion

This thesis represents a novel investigation into the utility of natural speech analysis for the assessment of psychopathology. We systematically reviewed the literature to obtain a comprehensive understanding of the state of the field from the perspectives of various disciplines. This knowledge provided a foundation upon which to base the development of our original study. We identified areas of opportunity for innovation through variable development and novel testing methods. We created a protocol aimed at optimizing naturalistic speech and assessing variables developed to index key symptoms of psychiatric disorders. We then applied these methods to an important psychiatric classification problem: differentiating bipolar from unipolar depression. Results support the validity and utility of our methods and provide direction for future investigation.

The breadth of the systematic review facilitated a high-level synthesis of results across fields and over time. It was beyond our scope to include in-depth evaluations of all aspects of speech analysis research; other studies have looked into comparing different collection or analysis methods (Gideon et al., 2016; Low et al., 2020), and breaking down the effects of possibly confounding factors (Cohen et al., 2016; Solomon et al., 2015; Wang et al., 2019). As the field continues to expand, we thought it most important to highlight predominant findings echoed throughout the literature, identify consistent gaps, and make recommendations for future directions. Potential applications of speech analysis were treated as the foundation for the synthesis of published data, in an effort to bring the results closer to clinical implementation.

Results of published literature revealed the unique utility of speech analysis for assessing psychopathology. Disorder classification using speech may be highly applicable as an aid

in clinical settings if similar high accuracies can be demonstrated for diagnoses other than depression and schizophrenia (Low et al., 2020). Limited results predicting illness prognosis and treatment outcomes using speech analysis leave this application farther behind in terms of practical applicability. Possibly the most promising function for speech analysis is detecting the onset of psychotic disorders and depression (Bearden et al., 2000; Bedi et al., 2015; Corcoran et al., 2018; Elvevåg et al., 2010; Hanakawa, 2004a; Ooi et al., 2013; Rezaii et al., 2019). Research suggests that speech patterns may reflect underlying, pre-clinical pathophysiological indicators of illness that could be used to identify individuals at risk (Bedi et al., 2015; Ooi et al., 2013). Other means of predicting psychiatric illness onset rely on subjective measures of symptoms (Purper-Ouakil et al., 2002; Schweizer et al., 2021), cognition (Schweizer et al., 2020), or family history (Correll et al., 2007; Paruk et al., 2017), which may be subject to biases in participant recall, differences in educational and socio-economic statuses, and the knowledge/presence of illness history, respectively. Speech analysis could play a unique and important role in early detection and subsequently better management of psychiatric illnesses. More research is needed to expand the scope of these findings across diagnoses. Heterogeneity of sample demographics and methodology throughout the field have limited the generalizability and applicability of speech analysis in clinical settings; divided research efforts from different fields have produced findings that are not easily compared. Commentary regarding the utility and limitations of speech analysis have been echoed in previous reviews (Low et al., 2020; Pennebaker et al., 2003; Robin et al., 2020). We focused the development of our study methods on addressing the gaps we identified in the literature to bring speech analysis closer to clinical application. We

considered three main areas of design: speech collection methods, speech feature variables, and analysis.

Most published research has focused on either content or acoustic analysis of speech (Cohen & Elvevåg, 2014; Tasnim & Stroulia, 2019). Generally, manual (human-coded) feature analysis methods have been used with content features (Carrillo et al., 2018; Himmelstein et al., 2018; Huston et al., 2019; Miller, 1996; Minor et al., 2015), whereas automated machine learning analysis methods have been applied almost exclusively to acoustic variables (Low et al., 2020; McGinnis et al., 2019; Ozdas et al., 2004; Pan et al., 2018). The potential for models incorporating both types of features, analyzed using both automated and manual methods, has yet to be determined. For this reason, we aimed to develop a speech collection protocol that would provide us with unadulterated, natural speech with emotional valance, suitable for various types of analysis.

In the study of mental disorders, speech features have been categorized based on their relative utility for assessing diagnoses; acoustic and content speech features have been positioned as uniquely indexing certain illness symptoms. Acoustic features may be more valid across demographic groups (Alghowinem et al., 2016), but lack the ability to provide insight into how individuals process experiences. Content variables provide that insight, but different sources define and measure these features in different ways, making it difficult to fully understand their capacity to assess psychopathology (Pennebaker et al., 2003). Human-coded variables included in our study were developed to assess content and semantic speech features. Variables indexed features of psychopathology commonly assessed in clinical environments as behaviors reflected in speech of affected individuals. We used tests of interrater reliability to measure the reliability of our coding methods

among multiple raters. We found no records of previous studies testing the reliability of content speech variables. As levels of agreement between psychiatrists have been shown to be inconsistent (Matuszak & Piasecki, 2012; Rozenzweig et al., 1961), the potential for speech analysis to provide objectivity to assessment may hinge on ensuring high levels of reliability for human-coded variables to demonstrate their validity.

We attempted to mitigate bias in the collection and rating of speech and related data, but it is important to discuss potential sources where it may have been introduced. Participants were assessed for presence and severity of psychopathology by trained research staff members. Evidence suggests the accuracy of assessments can be affected by biases in the rater and various demographic factors of the participants (Aboraya, 2007; Aggarwal, 2017; Matuszak & Piasecki, 2012). To mitigate this, and obtain more certainty of the diagnosis confirmation, diagnostic assessments were verified by a psychiatrist or psychologist. Speech samples coding introduced further variability into our analyses, as demonstrated by low interrater reliability for some variables (Chapter 3). Different backgrounds of raters contributed to different understandings of the psychiatric constructs embedded in our rating system. We attempted to mitigate rater drift by meeting consistently with our research team members to address coding discrepancies and engage in dialogue about the operationalizations of our novel variables. It is important to consider that manually rated datasets will be inherently biased based on many factors including expertise and attitudes of the rater(s). Blinding raters to participant information can help, and incorporating the role of independent raters who are unfamiliar with the research question and the population characterizes a more ideal scenario for manual coding.

In studying biomarkers, we are attempting to gain further objectivity, basing our conclusions on the "gold standard" of psychiatric diagnosing. Evidence suggests that the reliability of psychiatric diagnoses is variable, and can be poor (Matuszak & Piasecki, 2012). In our study, diagnostic certainty was approached by employing structured clinical interviews to assess psychopathology, and diagnostic consensus obtained by trained clinicians. For the purpose of increasing the validity of biomarker findings, current recommendations suggest the integration of transcultural psychiatry practices, more extensive blinding or raters and clinicians to participants' race, gender, and other demographic factors, and a better understanding of the utility of a given dataset for a certain research question (Straw & Callison-Burch, 2020).

Speech analysis methods have yet to be employed for the differentiation of unipolar and bipolar depression. This classification problem does not yet have a standard clinical solution; accurate diagnosis often relies on a variety of contributing factors (Benazzi & Akiskal, 2008; Bowden, 2001; Cuellar et al., 2005; Gruber et al., 2007; Hirschfeld, 2014). Machine learning classifiers have demonstrated the ability to differentiate BD from MDD with moderate-high accuracy using acoustic features (Huang et al., 2020; Su et al., 2020). These studies make use of an open source dataset, but do not include ratings of psychopathology or details of participant assessment in their methods or analysis, likely due to the studies' origins in computer science. Results demonstrate acoustic features' ability to aid in diagnostic differentiation, but are hard to interpret in terms of the psychiatric utility of characterizing BD and MDD.

Acoustic features have been shown to be effective in characterizing symptoms of mood disorders, such as psychomotor retardation (Wang et al., 2019) and arousal (J. Zhang et

al., 2018). Their ability to differentiate individuals with unipolar and bipolar depression has yet to be demonstrated. We chose to analyze speech variables representing content and semantic features as they can characterize clinical symptoms used to differentiate unipolar and bipolar depression, as described in Chapter 5. This type of variable selection has not been recorded in the past, however these variables may provide more effective disorder classification within depression than acoustic features, which demonstrate a more transdiagnostic symptom characterization. This has been demonstrated in studies using acoustic features to index depression with both BD and MDD participants (Arevian et al., 2020; Cohen et al., 2012).

Results of our analysis in Chapter 5 did not support our hypothesis. Follow-up analyses exploring the ability of speech features to predict depression severity showed some group differences. *Sadness* in speech was more strongly associated with depression severity in BD than in MDD. Literature suggests that negative emotions such as sadness may be more characteristic of unipolar depression (Table 1, Chapter 5); however, we identified no studies that attempted to assess depression severity between affective diagnoses using negative emotions (*Assessing Depression Severity*, Chapter 1). Selected speech features in a combined model accounted for more variance in BD (39%) than MDD (27%). The BD model was marginally significant after corrections, possibly indicating differences in speech behavior patterns between unipolar and bipolar depression. There may be potential for content speech variables to differentiate between unipolar and bipolar depression; however, we were underpowered to detect significant differences in our primary model, and likely also in follow-up testing. Understanding our results as preliminary, moderate classification accuracy indicates the potential of these features to

provide relevant diagnostic information for clinicians. More research on this topic is needed to expand the understanding of differential speech behaviors.

Moving forward, we aim to continue validating our methods and exploring further applications for speech analysis in psychiatry. Firstly, continued recruitment will improve the power of future results from this dataset. Next, we will begin to test the validity of speech variables by correlating them against clinical measures of psychiatric symptom severity. This will allow us to better understand and further operationalize the manifestation of illness in speech, and support the validity of our speech analysis methods for clinical application. For bipolar and unipolar depression differentiation, we would like to test acoustic analysis of participants; speech samples to compare acoustic classification ability against that of content variables. We will also explore the potential of acoustic analysis to add value to classification in a combined feature model with content features. We are beginning to explore the potential for automated analysis of content variables using machine learning methods, with the objective of creating an automated tool for clinical assistance in assessment. Given the demonstrated potential of speech analysis to assess various types of psychopathology, we hope our preliminary results can inspire future research efforts to explore the utility of speech analysis in BD and other, less-studied disorders.

References

Aboraya, A. (2007). Clinicians' Opinions on the Reliability of Psychiatric Diagnoses in Clinical Settings. *Psychiatry (Edgmont)*, *4*(11), 31–33.

Aboraya, A., Nasrallah, H. A., Elswick, D. E., Ahmed, E., Estephan, N., Aboraya, D., Berzingi, S., Chumbers, J., Berzingi, S., Justice, J., Zafar, J., & Dohar, S. (2018). Measurement-based Care in Psychiatry—Past, Present, and Future. *Innovations in Clinical Neuroscience*, *15*(11–12), 13–26.

Aggarwal, N. K. (2017). Culture, Communication, and DSM-5 Diagnostic Reliability. *Journal of the National Medical Association*, *109*(3), 150–152. https://doi.org/10.1016/j.jnma.2017.03.005

Alghowinem, S. (2015). *Multimodal Analysis of Verbal and Nonverbal Behaviour on the Example of Clinical Depression* [PhD]. The Australian National University.

Alghowinem, S., Goecke, R., Epps, J., Wagner, M., & Cohn, J. (2016). *Cross-Cultural Depression Recognition from Vocal Biomarkers*. 1943–1947. https://doi.org/10.21437/Interspeech.2016-1339

Alhanai, T., Ghassemi, M., & Glass, J. (2018). Detecting Depression with Audio/Text Sequence Modeling of Interviews. *Interspeech 2018*, 1716–1720. https://doi.org/10.21437/Interspeech.2018-2522

Alpert, M., Pouget, E. R., & Silva, R. R. (1995). Cues to the Assessment of Affects and Moods: Speech Fluency and Pausing. *Psychopharmacological Bulletin*, *31*(2), 421–424.

Alpert, M., Pouget, E. R., & Silva, R. R. (2001). Reflections of depression in acoustic measures of the patient's speech. *Journal of Affective Disorders*, *66*(1), 59–69. https://doi.org/10.1016/S0165-0327(00)00335-9

Alpert, M., Rosenberg, S. D., Pouget, E. R., & Shaw, R. J. (2000). Prosody and lexical accuracy in flat affect schizophrenia. *Psychiatry Research*, *97*(2–3), 107–118. https://doi.org/10.1016/S0165-1781(00)00231-6

Altshuler, L. L., Post, R. M., Leverich, G. S., Mikalauskas, K., Rosoff, A., & Ackerman, L. (1995). Antidepressant-induced mania and cycle acceleration: A controversy revisited. *American Journal of Psychiatry*, *152*(8), 1130–1138.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual for Mental Disorders* (5th ed.). American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425596

Angst, J., Adolfsson, R., Benazzi, F., Gamma, A., Hantouche, E., Meyer, T., Skeppar, P., Vieta, E., & Scott, J. (2005). The HCL-32: Towards a self-assessment tool for hypomanic symptoms in outpatients. *Journal of Affective Disorders*, *88*(2), 217–233. https://doi.org/10.1016/j.jad.2005.05.011

Angst, J., Merikangas, K. R., Cui, L., Van Meter, A., Ajdacic-Gross, V., & Rössler, W. (2018). Bipolar spectrum in major depressive disorders. *European Archives of Psychiatry and Clinical Neuroscience*, *268*(8), 741–748. https://doi.org/10.1007/s00406-018-0927-x

Arevian, A. C., Bone, D., Malandrakis, N., Martinez, V. R., Wells, K. B., Miklowitz, D. J., & Narayanan, S. (2020). Clinical state tracking in serious mental illness through computational analysis of speech. *PLoS ONE*, *15*(1). https://doi.org/10.1371/journal.pone.0225695

Ayer, A., Yalınçetin, B., Aydınlı, E., Sevilmiş, Ş., Ulaş, H., Binbay, T., Akdede, B. B., & Alptekin, K. (2016). Formal thought disorder in first-episode psychosis. *Comprehensive Psychiatry*, *70*, 209–215. https://doi.org/10.1016/j.comppsych.2016.08.005

Barzilay, R., Israel, N., Krivoy, A., Sagy, R., Kamhi-Nesher, S., Loebstein, O., Wolf, L., & Shoval, G. (2019). Predicting Affect Classification in Mental Status Examination Using Machine Learning Face Action Recognition System: A Pilot Study in Schizophrenia Patients. *Frontiers in Psychiatry*, *10*, 288. https://doi.org/10.3389/fpsyt.2019.00288

Batmaz, S., Kaymak, S. U., Soygur, A. H., Ozalp, E., & Turkcapar, M. H. (2013). The distinction between unipolar and bipolar depression: A cognitive theory perspective. *Comprehensive Psychiatry*, *54*(7), 740–749. https://doi.org/10.1016/j.comppsych.2013.02.004

Bearden, C. E., Rosso, I. M., Hollister, J. M., Sanchez, L. E., Hadley, T., & Cannon, T. D. (2000). A Prospective Cohort Study of Childhood Behavioral Deviance and Language Abnormalities as Predictors of Adult Schizophrenia. *Schizophrenia Bulletin*, *26*(2), 395–410. https://doi.org/10.1093/oxfordjournals.schbul.a033461

Bedi, G., Carrillo, F., Cecchi, G. A., Slezak, D. F., Sigman, M., Mota, N. B., Ribeiro, S., Javitt, D. C., Copelli, M., & Corcoran, C. M. (2015). Automated analysis of free speech predicts psychosis onset in high-risk youths. *Npj Schizophrenia*, *1*(1). https://doi.org/10.1038/npjschz.2015.30

Benazzi, F. (2007). Is There a Continuity between Bipolar and Depressive Disorders? *Psychotherapy and Psychosomatics*, *76*(2), 70–76. https://doi.org/10.1159/000097965

Benazzi, F., & Akiskal, H. (2005). Irritable-hostile depression: Further validation as a bipolar depressive mixed state. *Journal of Affective Disorders*, *84*(2–3), 197–207. https://doi.org/10.1016/j.jad.2004.07.006

Benazzi, F., & Akiskal, H. S. (2008). How best to identify a bipolar-related subtype among major depressive patients without spontaneous hypomania: Superiority of age at onset criterion over recurrence and polarity? *Journal of Affective Disorders*, *107*(1–3), 77–88. https://doi.org/10.1016/j.jad.2007.07.032

Benazzi, F., Helmi, S., & Bland, L. (2002). Agitated Depression: Unipolar? Bipolar? or Both? *Annals of Clinical Psychiatry*, *14*(2), 97–104. https://doi.org/10.3109/10401230209149096

Benazzi, F., Koukopoulos, A., & Akiskal, H. S. (2004). Toward a validation of a new definition of agitated depression as a bipolar mixed state (mixed depression). *European Psychiatry*, *19*(2), 85–90. https://doi.org/10.1016/j.eurpsy.2003.09.008

Bird, S., Hutchinson, B., Kenthapadi, K., Kiciman, E., & Mitchell, M. (2019). Fairness-Aware Machine Learning: Practical Challenges and Lessons Learned. *WWW'19: Companion Proceedings of The 2019 World Wide Web Conference*, 1297–1298.

Birmaher, B. (2013). Bipolar disorder in children and adolescents. *Child and Adolescent Mental Health*, *18*(3). https://doi.org/10.1111/camh.12021

Bowden, C. L. (2001). Strategies to Reduce Misdiagnosis of Bipolar Depression. *Psychiatric Services*, *52*(1), 51–55. https://doi.org/10.1176/appi.ps.52.1.51

Bowden, C. L. (2005). A different depression: Clinical distinctions between bipolar and unipolar depression. *Journal of Affective Disorders*, *84*(2–3), 117–125. https://doi.org/10.1016/S0165-0327(03)00194-0

Braun, S., Annovazzi, C., Botella, C., Bridler, R., Camussi, E., Delfino, J. P., Mohr, C., Moragrega, I., Papagno, C., Pisoni, A., Soler, C., Seifritz, E., & Stassen, H. H. (2016). Assessing Chronic Stress, Coping Skills, and Mood Disorders through Speech Analysis: A Self-Assessment "Voice App" for laptops, Tablets and Smartphones. *Psychopathology*, *49*(6), 406–419. https://doi.org/10.1159/000450959

Breznitz, Z. (1992). Verbal Indicators of Depression. *The Journal of General Psychology*, *119*(4), 351–363.

Brockington, I. F., Altman, E., Hillier, V., Meltzer, H. Y., & Nand, S. (1982). The Clinical Picture of Bipolar Affective Disorder in its Depressed Phase: A Report from London and Chicago. *British Journal of Psychiatry*, *141*(6), 558–562. https://doi.org/10.1192/bjp.141.6.558

Brockmeyer, T., Zimmermann, J., Kulessa, D., Hautzinger, M., Bents, H., Friederich, H.-C., Herzog, W., & Backenstrass, M. (2015). Me, myself, and I: Self-referent word use as an indicator of self-focused attention in relation to depression and anxiety. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01564

Buck, B., Minor, K. S., & Lysaker, P. H. (2015). Differential lexical correlates of social cognition and metacognition in schizophrenia; a study of spontaneously-generated life narratives. *Comprehensive Psychiatry*, *58*, 138–145. https://doi.org/10.1016/j.comppsych.2014.12.015

Carrillo, F., Sigman, M., Fernández Slezak, D., Ashton, P., Fitzgerald, L., Stroud, J., Nutt, D. J., & Carhart-Harris, R. L. (2018). Natural speech algorithm applied to baseline interview data can predict which patients will respond to psilocybin for treatment-resistant depression. *Journal of Affective Disorders*, *230*, 84–86. https://doi.org/10.1016/j.jad.2018.01.006

Carroll, K. M., & Rounsaville, B. J. (2010). Computer-assisted Therapy in Psychiatry: Be Brave—It's a New World. *Current Psychiatry Reports*, *12*(5), 426–432. https://doi.org/10.1007/s11920-010-0146-2

Chang, K. (2009). Challenges in the diagnosis and treatment of pediatric bipolar depression. *Dialogues in Clinical Neuroscience*, *11*(1), 73–80.

Chaplin, T. M. (2015). Gender and Emotion Expression: A Developmental Contextual Perspective. *Emotion Review*, *7*(1), 14–21. https://doi.org/10.1177/1754073914544408

Chouldechova, A., & Roth, A. (2018). The Frontiers of Fairness in Machine Learning. *ArXiv:1810.08810 [Cs, Stat]*. http://arxiv.org/abs/1810.08810

Cohen, A. S., Alpert, M., Nienow, T. M., Dinzeo, T. J., & Docherty, N. M. (2008). Computerized measurement of negative symptoms in schizophrenia. *Journal of Psychiatric Research*, *42*(10), 827–836. https://doi.org/10.1016/j.jpsychires.2007.08.008

Cohen, A. S., & Elvevåg, B. (2014). Automated Computerized Analysis of Speechin Psychiatric Disorders. *Current Opinion in Psychiatry*, *27*(3), 203–209. https://doi.org/10.1097/YCO.0000000000000056

Cohen, A. S., Kim, Y., & Najolia, G. M. (2013). Psychiatric symptom versus neurocognitive correlates of diminished expressivity in schizophrenia and mood disorders. *Schizophrenia Research*, *146*(1–3), 249–253. https://doi.org/10.1016/j.schres.2013.02.002

Cohen, A. S., McGovern, J. E., Dinzeo, T. J., & Covington, M. A. (2014). Speech deficits in serious mental illness: A cognitive resource issue? *Schizophrenia Research*, *160*(1–3), 173–179. https://doi.org/10.1016/j.schres.2014.10.032

Cohen, A. S., Mitchell, K. R., Docherty, N. M., & Horan, W. P. (2016). Vocal expression in schizophrenia: Less than meets the ear. *Journal of Abnormal Psychology*, *125*(2), 299–309. https://doi.org/10.1037/abn0000136

Cohen, A. S., Najolia, G. M., Kim, Y., & Dinzeo, T. J. (2012). On the boundaries of blunt affect/alogia across severe mental illness: Implications for Research Domain Criteria. *Schizophrenia Research*, *140*(1–3), 41–45. https://doi.org/10.1016/j.schres.2012.07.001

Corcoran, C. M., Carrillo, F., Fernández-Slezak, D., Bedi, G., Klim, C., Javitt, D. C., Bearden, C. E., & Cecchi, G. A. (2018). Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World Psychiatry*, *17*(1), 67–75. https://doi.org/10.1002/wps.20491

Corcoran, C. M., & Cecchi, G. A. (2020). Using Language Processing and Speech Analysis for the Identification of Psychosis and Other Disorders. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *5*(8), 770–779. https://doi.org/10.1016/j.bpsc.2020.06.004

Corcoran, C. M., Mittal, V. A., Bearden, C. E., E. Gur, R., Hitczenko, K., Bilgrami, Z., Savic, A., Cecchi, G. A., & Wolff, P. (2020). Language as a biomarker for psychosis: A natural language processing approach. *Schizophrenia Research*, *226*, 158–166. https://doi.org/10.1016/j.schres.2020.04.032

Correll, C. U., Penzner, J. B., Lencz, T., Auther, A., Smith, C. W., Malhotra, A. K., Kane, J. M., & Cornblatt, B. A. (2007). Early identification and high-risk strategies for bipolar disorder. *Bipolar Disorders*, *9*(4), 324–338. https://doi.org/10.1111/j.1399-5618.2007.00487.x

Cuellar, A. K., Johnson, S. L., & Winters, R. (2005). Distinctions between bipolar and unipolar depression. *Clinical Psychology Review*, *25*(3), 307–339. https://doi.org/10.1016/j.cpr.2004.12.002

Cummins, N., Epps, J., & Ambikairajah, E. (2013). Spectro-temporal analysis of speech affected by depression and psychomotor retardation. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 7542–7546. https://doi.org/10.1109/ICASSP.2013.6639129

Cummins, N., Epps, J., Breakspear, M., & Goecke, R. (2011). An Investigation of Depressed Speech Detection: Features and Normalization. *Interspeech 2011*, 2997–3000.

Cummins, N., Epps, J., Sethu, V., Breakspear, M., & Goecke, R. (2013). Modeling Spectral Variability for the Classification of Depressed Speech. *Interspeech 2013*, 857–861.

Day, R. K. (1999). Psychomotor agitation: Poorly defined and badly measured. *Journal of Affective Disorders*, 10.

de Boer, J. N., Voppel, A. E., Begemann, M. J. H., Schnack, H. G., Wijnen, F., & Sommer, I. E. C. (2018). Clinical use of semantic space models in psychiatry and neurology: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *93*, 85–92. https://doi.org/10.1016/j.neubiorev.2018.06.008

Diler, R. S., Goldstein, T. R., Hafeman, D., Merranko, J., Liao, F., Goldstein, B. I., Hower, H., Gill, M. K., Hunt, J., Yen, S., Keller, M. B., Axelson, D., Strober, M., Iyengar, S., Ryan, N. D., & Birmaher, B. (2017). Distinguishing Bipolar Depression from Unipolar Depression in Youth: Preliminary Findings. *Journal of Child and Adolescent Psychopharmacology*, *27*(4), 310–319. https://doi.org/10.1089/cap.2016.0154

Drost, J., van der Does, W., van Hemert, A. M., Penninx, B. W. J. H., & Spinhoven, P. (2014). Repetitive negative thinking as a transdiagnostic factor in depression and anxiety: A conceptual replication. *Behaviour Research and Therapy*, *63*, 177–183. https://doi.org/10.1016/j.brat.2014.06.004

Elvevåg, B., Foltz, P. W., Rosenstein, M., & DeLisi, L. E. (2010). An automated method to analyze language use in patients with schizophrenia and their first-degree relatives. *Journal of Neurolinguistics*, *23*(3), 270–284. https://doi.org/10.1016/j.jneuroling.2009.05.002

Elvevåg, B., Foltz, P. W., Weinberger, D. R., & Goldberg, T. E. (2007). Quantifying incoherence in speech: An automated methodology and novel application to schizophrenia. *Schizophrenia Research*, *93*(1–3), 304–316. https://doi.org/10.1016/j.schres.2007.03.001

Engelhardt, P. E., Corley, M., Nigg, J. T., & Ferreira, F. (2010). The role of inhibition in the production of disfluencies. *Memory & Cognition*, *38*(5), 617–628. https://doi.org/10.3758/MC.38.5.617

Erol, A., Mete, L., Doganer, I., & Sorias, S. (1995). The Use of Computerized Content Analysis Method for Psychiatric Diagnosis. *Turkish Journal of Psychiatry*, *6*(4), 263–268.

Fàbrega, H. (2001). CULTURE AND HISTORY IN PSYCHIATRIC DIAGNOSIS AND PRACTICE. *Psychiatric Clinics of North America*, *24*(3), 391–405. https://doi.org/10.1016/S0193-953X(05)70236-7

Faurholt-Jepsen, M., Busk, J., Frost, M., Vinberg, M., Christensen, E. M., Winther, O., Bardram, J. E., & Kessing, L. V. (2016). Voice analysis as an objective state marker in bipolar disorder. *Translational Psychiatry*, *6*(7), e856–e856. https://doi.org/10.1038/tp.2016.123

Fava, G. A. (2020). May antidepressant drugs worsen the conditions they are supposed to treat? The clinical foundations of the oppositional model of tolerance. *Therapeutic Advances in Psychopharmacology*, *10*. https://doi.org/10.1177/2045125320970325

Feng, Y., Wang, Y.-Y., Huang, W., Ungvari, G. S., Ng, C. H., Wang, G., Yuan, Z., & Xiang, Y.-T. (2017). Comparison of the 32-item Hypomania Checklist, the 33-item Hypomania Checklist, and the Mood Disorders Questionnaire for bipolar disorder. *Psychiatry and Clinical Neurosciences*, *71*(6), 403–408. https://doi.org/10.1111/pcn.12506

Fernandes, B. S., Karmakar, C., Tamouza, R., Tran, T., Yearwood, J., Hamdani, N., Laouamri, H., Richard, J.-R., Yolken, R., Berk, M., Venkatesh, S., & Leboyer, M. (2020). Precision psychiatry with immunological and cognitive biomarkers: A multi-domain prediction for the diagnosis of bipolar disorder or schizophrenia using machine learning. *Translational Psychiatry*, *10*. https://doi.org/10.1038/s41398-020-0836-4

Fierro, M., Bustos, A., & Molina, C. (2016). Diferencias en la experiencia subjetiva entre depresión unipolar y bipolar. *Revista Colombiana de Psiquiatría*, *45*(3), 162–169. https://doi.org/10.1016/j.rcp.2015.09.006

Filakovi, P. (2011). New Strategies in the Treatment of Bipolar Disorder. *Psychiatria Danubina*, *23*(3), 293–299.

Flint, A. J., Black, S. E., Campbell-Taylor, I., Gailey, G. F., & Levinton, C. (1992). Acoustic analysis in the differentiation of Parkinson's disease and major depression. *Journal of Psycholinguistic Research*, *21*(5), 383–399. https://doi.org/10.1007/BF01067922

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. *Journal of Psychiatric Research*, *27*(3), 309–319.

Forty, L., Smith, D., Jones, L., Jones, I., Caesar, S., Cooper, C., Fraser, C., Gordon-Smith, K., Hyde, S., Farmer, A., McGuffin, P., & Craddock, N. (2008). Clinical differences between bipolar and unipolar depression. *British Journal of Psychiatry*, *192*(5), 388–389. https://doi.org/10.1192/bjp.bp.107.045294

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 Transactions on Biomedical Engineering*, *47*(7), 829–837. https://doi.org/10.1109/10.846676

Fraser, W. I., King, K. M., Thomas, P., & Kendell, R. E. (1986). The Diagnosis of Schizophrenia by Language Analysis. *British Journal of Psychiatry*, *148*(3), 275–278. https://doi.org/10.1192/bjp.148.3.275

Galvão, F., Sportiche, S., Lambert, J., Amiez, M., Musa, C., Nieto, I., Dubertret, C., & Lepine, J. P. (2013). Clinical differences between unipolar and bipolar depression: Interest of BDRS (Bipolar Depression Rating Scale). *Comprehensive Psychiatry*, *54*(6), 605–610. https://doi.org/10.1016/j.comppsych.2012.12.023

Ghaemi, S. N. (2000). Diagnosing Bipolar Disorder and the Effect of Antidepressants: A Naturalistic Study. *J Clin Psychiatry*, 8.

Gideon, J., Provost, E. M., & McInnis, M. (2016). Mood state prediction from speech of varying acoustic quality for individuals with bipolar disorder. *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2359–2363. https://doi.org/10.1109/ICASSP.2016.7472099

Goldberg, D. (2011). The heterogeneity of "major depression." *World Psychiatry*, *10*(3), 226–228.

Gosek, P., Heitzman, J., Stefanowski, B., Antosik-Wójcińska, A., & Parnowski, T. (2019). Symptomatic differences and symptoms stability in unipolar and bipolar depression. Medical charts review in 99 inpatients. *Psychiatria Polska*, *53*(3), 655–672. https://doi.org/10.12740/PP/102656

Gruber, S., Rathgeber, K., Bräunig, P., & Gauggel, S. (2007). Stability and course of neuropsychological deficits in manic and depressed bipolar patients compared to patients with Major Depression. *Journal of Affective Disorders*, *104*(1–3), 61–71. https://doi.org/10.1016/j.jad.2007.02.011

Guo, T., Xiang, Y.-T., Xiao, L., Hu, C.-Q., Chiu, H. F. K., Ungvari, G. S., Correll, C. U., Lai, K. Y. C., Feng, L., Geng, Y., Feng, Y., & Wang, G. (2015). Measurement-Based Care Versus Standard Care for Major Depression: A Randomized Controlled Trial With Blind Raters. *American Journal of Psychiatry*, *172*(10), 1004–1013. https://doi.org/10.1176/appi.ajp.2015.14050652

Hanakawa, Y. (2004a). *Comparison of Linguistic and Acoustic Analyses of Schizophrenic Language* [PhD]. Adelphi University.

Hanakawa, Y. (2004b). Patterns of schizophrenic language, analyzed by computerized content analysis. *Journal of the American Psychoanalytic Association*, *52*(2), 475–476.

Hantouche, E. G., & Akiskal, H. S. (2005). Bipolar II vs. unipolar depression: Psychopathologic differentiation by dimensional measures. *Journal of Affective Disorders*, *84*(2–3), 127–132. https://doi.org/10.1016/j.jad.2004.01.017

Harati, S., Crowell, A., Mayberg, H., & Nemati, S. (2018). Depression Severity Classification from Speech Emotion. *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 5763–5766. https://doi.org/10.1109/EMBC.2018.8513610

Himmelstein, P., Barb, S., Finlayson, M. A., & Young, K. D. (2018). Linguistic analysis of the autobiographical memories of individuals with major depressive disorder. *PLOS ONE*, *13*(11), e0207814. https://doi.org/10.1371/journal.pone.0207814

Hirschfeld, R. M. (2014). Differential diagnosis of bipolar disorder and major depressive disorder. *Journal of Affective Disorders*, *169*, S12–S16. https://doi.org/10.1016/S0165-0327(14)70004-7

Hoffman, G. M., Gonze, J. C., & Mendlewicz, J. (1985). Speech Pause Time as a Method for the Evaluation of Psychomotor Retardation in Depressive Illness. *British Journal of Psychiatry*, *146*(5), 535–538. https://doi.org/10.1192/bjp.146.5.535

Hoffman, R. E. (1986). A Comparative Study of Manic vs Schizophrenic Speech Disorganization. *Archives of General Psychiatry*, *43*(9), 831–838. https://doi.org/10.1001/archpsyc.1986.01800090017003

Holshausen, K., Harvey, P. D., Elvevåg, B., Foltz, P. W., & Bowie, C. R. (2014). Latent semantic variables are associated with formal thought disorder and adaptive behavior in older inpatients with schizophrenia. *Cortex*, *55*, 88–96. https://doi.org/10.1016/j.cortex.2013.02.006

Hong, K., Nenkova, A., March, M. E., Parker, A. P., Verma, R., & Kohler, C. G. (2015). Lexical use in emotional autobiographical narratives of persons with schizophrenia and healthy controls. *Psychiatry Research*, *225*(1–2), 40–49. https://doi.org/10.1016/j.psychres.2014.10.002

Huang, K.-Y., Wu, C.-H., Su, M.-H., & Kuo, Y.-T. (2020a). Detecting Unipolar and Bipolar Depressive Disorders from Elicited Speech Responses Using Latent Affective Structure Model. *IEEE Transactions on Affective Computing*, *11*(3), 393–404. https://doi.org/10.1109/TAFFC.2018.2803178

Huang, K.-Y., Wu, C.-H., Su, M.-H., & Kuo, Y.-T. (2020b). Detecting Unipolar and Bipolar Depressive Disorders from Elicited Speech Responses Using Latent Affective Structure Model. *IEEE Transactions on Affective Computing*, *11*(3), 393–404. https://doi.org/10.1109/TAFFC.2018.2803178

Huston, J., Meier, S., Faith, M., & Reynolds, A. (2019). Exploratory study of automated linguistic analysis for progress monitoring and outcome assessment. *Counselling and Psychotherapy Research*, *19*(3), 321–328. https://doi.org/10.1002/capr.12219

Jiang, H., Hu, B., Liu, Z., Wang, G., Zhang, L., Li, X., & Kang, H. (2018). Detecting Depression Using an Ensemble Logistic Regression Model Based on Multiple Speech Features. *Computational and Mathematical Methods in Medicine*, *2018*. https://doi.org/10.1155/2018/6508319

Judd, L. L., & Akiskal, H. S. (2003). The prevalence and disability of bipolar spectrum disorders in the US population: Re-analysis of the ECA database taking into account subthreshold cases. *Journal of Affective Disorders*, *73*(1–2), 123–131. https://doi.org/10.1016/S0165-0327(02)00332-4

Kaldi: Kaldi. (2009). http://kaldi-asr.org/doc/

Kealy, D., Treeby, M. S., & Rice, S. M. (2021). Shame, guilt, and suicidal thoughts: The interaction matters. *British Journal of Clinical Psychology*, *n/a*(n/a). https://doi.org/10.1111/bjc.12291

Kim, S., Thibodeau, R., & Jorgensen, R. S. (2011). Shame, Guilt, and Depressive Symptoms: A Meta-Analytic Review. *Psychology Bulletin*, *137*(1), 68–96.

Kotsopoulos, S., & Mellor, C. (1986). Extralinguistic Speech Characteristics of Children with Conduct and Anxiety Disorders. *Journal of Child Psychology and Psychiatry*, *27*(1), 99–108. https://doi.org/10.1111/j.1469-7610.1986.tb00625.x

Lee, S. (2002). Socio-Cultural and Global Health Perspectives for the Development of Future Psychiatric Diagnostic Systems. *Psychopathology*, *35*(2–3), 152–157. https://doi.org/10.1159/000065136

Lewis, C. C., Puspitasari, A., Boyd, M. R., Scott, K., Marriott, B. R., Hoffman, M., Navarro, E., & Kassab, H. (2018). Implementing measurement based care in community mental health: A description of tailored and standardized methods. *BMC Research Notes*, *11*. https://doi.org/10.1186/s13104-018-3193-0

Löw, C. A., Schauenburg, H., & Dinger, U. (2020). Self-criticism and psychotherapy outcome: A systematic review and meta-analysis. *Clinical Psychology Review*, *75*, 101808. https://doi.org/10.1016/j.cpr.2019.101808

Low, D. M., Bentley, K. H., & Ghosh, S. S. (2020). Automated assessment of psychiatric disorders using speech: A systematic review. *Laryngoscope Investigative Otolaryngology*, *5*(1), 96–116. https://doi.org/10.1002/lio2.354

Machado-Nascimento, N., Melo e Kümmer, A., & Lemos, S. M. A. (2016). Alterações Fonoaudiológicas no Transtorno de Déficit de Atenção e Hiperatividade: Revisão sistemática de literatura. *CoDAS*, *28*(6), 833–842. https://doi.org/10.1590/2317-1782/20162015270

Magaña, A. B., Goldstein, M. J., Karno, M., Miklowitz, D. J., Jenkins, J., & Falloon, I. R. H. (1986). A brief method for assessing expressed emotion in relatives of psychiatric patients. *Psychiatry Research*, *17*(3), 203–212. https://doi.org/10.1016/0165-1781(86)90049-1

Maher, B. A., Manschreck, T. C., Linnet, J., & Candela, S. (2005). Quantitative assessment of the frequency of normal associations in the utterances of schizophrenia patients and healthy controls. *Schizophrenia Research*, *78*(2–3), 219–224. https://doi.org/10.1016/j.schres.2005.05.017

Malhi, G. S., Ivanovski, B., Hadzi- Pavlovic, D., Mitchell, P. B., Vieta, E., & Sachdev, P. (2007). Neuropsychological deficits and functional impairment in bipolar depression, hypomania and euthymia. *Bipolar Disorders*, *9*(1–2), 114–125. https://doi.org/10.1111/j.1399-5618.2007.00324.x

Manschreck, T. C., Merrill, A. M., Jabbar, G., Chun, J., & DeLisi, L. E. (2012). Frequency of normative word associations in the speech of individuals at familial high-risk for schizophrenia. *Schizophrenia Research*, *140*(1–3), 99–103. https://doi.org/10.1016/j.schres.2012.06.034

Martínez-Sánchez, F., Muela-Martínez, J. A., Cortés-Soto, P., García Meilán, J. J., Vera Ferrándiz, J. A., Egea Caparrós, A., & Pujante Valverde, I. M. (2015). Can the Acoustic Analysis of Expressive Prosody Discriminate Schizophrenia? *The Spanish Journal of Psychology*, *18*. https://doi.org/10.1017/sjp.2015.85

Matuszak, J., & Piasecki, M. (2012). Inter-Rater Reliability in Psychiatric Diagnosis. *Psychiatric Times*, *29*(10). https://www.psychiatrictimes.com/view/inter-rater-reliability-psychiatric-diagnosis

McGinnis, E. W., Anderau, S. P., Hruschak, J., Gurchiek, R. D., Lopez-Duran, N. L., Fitzgerald, K., Rosenblum, K. L., Muzik, M., & McGinnis, R. S. (2019). Giving Voice to Vulnerable Children: Machine Learning Analysis of Speech Detects Anxiety and Depression in Early Childhood. *IEEE Journal of Biomedical and Health Informatics*, *23*(6), 2294–2301. https://doi.org/10.1109/JBHI.2019.2913590

McGorry, P., & Nelson, B. (2016). Why We Need a Transdiagnostic Staging Approach to Emerging Psychopathology, Early Diagnosis, and Treatment. *JAMA Psychiatry*, *73*(3), 191. https://doi.org/10.1001/jamapsychiatry.2015.2868

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 276–282. https://doi.org/10.11613/BM.2012.031

Merikangas, K. R., Jin, R., He, J.-P., Kessler, R. C., Lee, S., Sampson, N. A., Viana, M. C., Andrade, L. H., Hu, C., Karam, E. G., Ladea, M., Mora, M. E. M., Browne, M. O., Ono, Y., Posada-Villa, J., Sagar, R., & Zarkov, Z. (2011). Prevalence and Correlates of Bipolar Spectrum Disorder in the World Mental Health Survey Initiative. *Archives of General Psychiatry*, *68*(3), 241–251. https://doi.org/10.1001/archgenpsychiatry.2011.12

Mete, L., Doganer, I., Erol, A., Sorias, S., & Ertas, M. (1993). Contents of Speech in Schizophrenia, Mania and Depression. *Turkish Journal of Psychiatry*, *4*(3).

Mete, L., Schnurr, P. P., Rosenberg, S. D., Oxman, T. E., Doganer, I., & Sorias, S. (1993). Language content and schizophrenia in acute phase Turkish patients. *Social Psychiatry and Psychiatric Epidemiology*, *28*(6), 275–280. https://doi.org/10.1007/BF00795907

Meyer, S. E., Carlson, G. A., Wiggs, E. A., Martinez, P. E., Ronsaville, D. S., Klimes–Dougan, B., Gold, P. W., & Radke–Yarrow, M. (2004). A prospective study of the association among impaired executive functioning, childhood attentional problems, and the development of bipolar disorder. *Development and Psychopathology*, *16*(02). https://doi.org/10.1017/S095457940404461X

Miller, N. L. (1996). *The Effect of Antidepressants on Language and Symptoms* [PhD]. Adelphi University.

Minor, K. S., Bonfils, K. A., Luther, L., Firmin, R. L., Kukla, M., MacLain, V. R., Buck, B., Lysaker, P. H., & Salyers, M. P. (2015). Lexical analysis in schizophrenia: How emotion and social word use informs our understanding of clinical presentation. *Journal of Psychiatric Research*, *64*, 74–78. https://doi.org/10.1016/j.jpsychires.2015.02.024

Moe, A. M., Breitborde, N. J. K., Shakeel, M. K., Gallagher, C. J., & Docherty, N. M. (2016). Idea density in the life-stories of people with schizophrenia: Associations with narrative qualities and psychiatric symptoms. *Schizophrenia Research*, *172*(1–3), 201–205. https://doi.org/10.1016/j.schres.2016.02.016

Mondimore, F. M. (2005). Unipolar depression/bipolar depression: Connections and controversies. *International Review of Psychiatry*, *17*(1), 39–47. https://doi.org/10.1080/09540260500068216

Monferrer, M., Ricarte, J. J., Montes, M. J., Fernández-Caballero, A., & Fernández-Sotos, P. (2021). Psychosocial remediation in depressive disorders: A systematic review. *Journal of Affective Disorders*, *290*, 40–51. https://doi.org/10.1016/j.jad.2021.04.052

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 Transactions on Biomedical Engineering*, *55*(1), 96–107. https://doi.org/10.1109/TBME.2007.900562

Moore, E., Clements, M., Peifer, J., & Weisser, L. (2004). Comparing objective feature statistics of speech for classifying clinical depression. *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, *23*, 17–20. https://doi.org/10.1109/IEMBS.2004.1403079

Moreno, D. H., & Andrade, L. H. (2005). The lifetime prevalence, health services utilization and risk of suicide of bipolar spectrum subjects, including subthreshold categories in the São Paulo ECA study. *Journal of Affective Disorders*, *87*(2–3), 231–241. https://doi.org/10.1016/j.jad.2005.04.010

Mota, N. B., Copelli, M., & Ribeiro, S. (2017). Thought disorder measured as random speech structure classifies negative symptoms and schizophrenia diagnosis 6 months in advance. *Npj Schizophrenia*, *3*(1). https://doi.org/10.1038/s41537-017-0019-3

Mota, N. B., Vasconcelos, N. A. P., Lemos, N., Pieretti, A. C., Kinouchi, O., Cecchi, G. A., Copelli, M., & Ribeiro, S. (2012). Speech Graphs Provide a Quantitative Measure of Thought Disorder in Psychosis. *PLoS ONE*, *7*(4), e34928. https://doi.org/10.1371/journal.pone.0034928

Motovsky, B., & Pecenak, J. (2013). *PSYCHOPATHOLOGICAL CHARACTERISTICS OF BIPOLAR AND UNIPOLAR DEPRESSION - POTENTIAL INDICATORS OF BIPOLARITY*. *25*(1), 6.

Muller, M. J., Himmerich, H., Kienzle, B., & Szegedi, A. (2003). Differentiating moderate and severe depression using the Montgomery–A˚ sberg depression rating scale (MADRS). *Journal of Affective Disorders*, 6.

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. *Journal of Neurolinguistics*, *20*(1), 50–64. https://doi.org/10.1016/j.jneuroling.2006.04.001

Mundt, J. C., Vogel, A. P., Feltner, D. E., & Lenderking, W. R. (2012). Vocal Acoustic Biomarkers of Depression Severity and Treatment Response. *Biological Psychiatry*, *72*(7), 580–587. https://doi.org/10.1016/j.biopsych.2012.03.015

Newman, S., & Mather, V. G. (1938). Analysis of Spoken Language of Patients with Affective Disorders. *American Journal of Insanity*, *94*(4), 913–942. https://doi-org.ezproxy.library.dal.ca/10.1176/ajp.94.4.913

Nilsonne, Å., Sundberg, J., Ternström, S., & Askenfelt, A. (1988). Measuring the rate of change of voice fundamental frequency in fluent speech during mental depression. *The Journal of the Acoustical Society of America*, *83*(2), 716–728. https://doi.org/10.1121/1.396114

Novack, D. L. (2003). *Aspects of Emotional Experssion in Flat Affect Schizophrenia and Depression* [PhD]. New York University.

Nuñez, N. A., Comai, S., Dumitrescu, E., Ghabrash, M. F., Tabaka, J., Saint-Laurent, M., Vida, S., Kolivakis, T., Fielding, A., Low, N., Cervantes, P., Booij, L., & Gobbi, G. (2018). Psychopathological and sociodemographic features in treatment-resistant unipolar depression versus bipolar depression: A comparative study. *BMC Psychiatry*, *18*(1), 68. https://doi.org/10.1186/s12888-018-1641-y

O'Dell, J. W., & Winder, P. (1975). Evaluation of a content-analysis system for therapeutic interviews. *Journal of Clinical Psychology*, *31*(4), 737–744.

Ooi, K. E. B., Lech, M., & Allen, N. B. (2013). Multichannel Weighted Speech Classification System for Prediction of Major Depression in Adolescents. *IEEE Transactions on Biomedical Engineering*, *60*(2), 497–506. https://doi.org/10.1109/TBME.2012.2228646

Oxman, T. E., Rosenberg, S. D., Schnurr, P. P., & Tucker, G. J. (1988). Diagnostic classification through content analysis of patients' speech. *American Journal of Psychiatry*, *145*(4), 464–468. https://doi.org/10.1176/ajp.145.4.464

Ozdas, A., Shiavi, R. G., Wilkes, D. M., Silverman, M. K., & Silverman, S. E. (2004). Analysis of Vocal Tract Characteristics for Near-term Suicidal Risk Assessment. *Methods of Information in Medicine*, *43*(01), 36–38. https://doi.org/10.1055/s-0038-1633420

Pan, Z., Gui, C., Zhang, J., Zhu, J., & Cui, D. (2018). Detecting Manic State of Bipolar Disorder Based on Support Vector Machine and Gaussian Mixture Model Using Spontaneous Speech. *Psychiatry Investigation*, *15*(7), 695–700. https://doi.org/10.30773/pi.2017.12.15

Parrott, W. G. (Ed.). (2001). *Emotions in social psychology: Essential readings* (pp. xiv, 378). Psychology Press.

Paruk, S., Jhazbhay, K., Singh, K., Sartorius, B., & Burns, J. K. (2017). The clinical impact of a positive family history of psychosis or mental illness in psychotic and non-psychotic mentally ill adolescents. *Journal of Child & Adolescent Mental Health*, *29*(3), 219–229. https://doi.org/10.2989/17280583.2017.1389741

Pauselli, L., Halpern, B., Cleary, S. D., Ku, B., Covington, M. A., & Compton, M. T. (2018). Computational linguistic analysis applied to a semantic fluency task to measure derailment and tangentiality in schizophrenia. *Psychiatry Research*, *263*, 74–79. https://doi.org/10.1016/j.psychres.2018.02.037

Pavlova, B., & Uher, R. (2020a). Assessment of Psychopathology: Is Asking Questions Good Enough? *JAMA Psychiatry*. https://doi.org/10.1001/jamapsychiatry.2020.0108

Pavlova, B., & Uher, R. (2020b). Assessment of Psychopathology: Is Asking Questions Good Enough? *JAMA Psychiatry*, *77*(6), 557. https://doi.org/10.1001/jamapsychiatry.2020.0108

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological Aspects of Natural Language Use: Our Words, Our Selves. *Annual Review of Psychology*, *54*(1), 547–577. https://doi.org/10.1146/annurev.psych.54.101601.145041

Perlini, C., Marini, A., Garzitto, M., Isola, M., Cerruti, S., Marinelli, V., Rambaldelli, G., Ferro, A., Tomelleri, L., Dusi, N., Bellani, M., Tansella, M., Fabbro, F., & Brambilla, P. (2012). Linguistic production and syntactic comprehension in schizophrenia and bipolar disorder. *Acta Psychiatrica Scandinavica*, *126*(5), 363–376. https://doi.org/10.1111/j.1600-0447.2012.01864.x

Perlis, R. H., Fava, M., Trivedi, M. H., Alpert, J., Luther, J. F., Wisniewski, S. R., & Rush, A. J. (2009). Irritability is associated with anxiety and greater severity, but not bipolar spectrum features, in major depressive disorder. *Acta Psychiatrica Scandinavica*, *119*(4), 282–289. https://doi.org/10.1111/j.1600-0447.2008.01298.x

Perlis, R. H., Smoller, J. W., Fava, M., Rosenbaum, J. F., Nierenberg, A. A., & Sachs, G. S. (2004). The prevalence and clinical correlates of anger attacks during depressive episodes in bipolar disorder. *Journal of Affective Disorders*, *79*(1–3), 291–295. https://doi.org/10.1016/S0165-0327(02)00451-2

Pestian, J. P., Grupp- Phelan, J., Cohen, K. B., Meyers, G., Richey, L. A., Matykiewicz, P., & Sorter, M. T. (2016). A Controlled Trial Using Natural Language Processing to Examine the Language of Suicidal Adolescents in the Emergency Department. *Suicide and Life-Threatening Behavior*, *46*(2), 154–159. https://doi.org/10.1111/sltb.12180

Pies, R. (2007). How "Objective" Are Psychiatric Diagnoses? *Psychiatry (Edgmont)*, *4*(10), 18–22.

Poletti, S., Vai, B., Mazza, M. G., Zanardi, R., Lorenzi, C., Calesella, F., Cazzetta, S., Branchi, I., Colombo, C., Furlan, R., & Benedetti, F. (2021). A peripheral inflammatory signature discriminates bipolar from unipolar depression: A machine learning approach. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *105*, 110136. https://doi.org/10.1016/j.pnpbp.2020.110136

Post, R. M. (2005). The Impact of Bipolar Depression. *J Clin Psychiatry*, *66*(5), 5–10.

Purper-Ouakil, D., Michael, G., & Mouren-Simeoni, M. C. (2002). Vulnerability to Depression in Children and Adolescents: Update and Perspectives. *Encephale*, *28*(3), 234–240.

Quatieri, T. F., & Malyska, N. (2012). Vocal-Source Biomarkers for Depression: A Link to Psychomotor Activity. *Interspeech 2012*, 1059–1062.

Ratana, R., Sharifzadeh, H., Krishnan, J., & Pang, S. (2019). A Comprehensive Review of Computational Methods for Automatic Prediction of Schizophrenia With Insight Into Indigenous Populations. *Frontiers in Psychiatry*, *10*. https://doi.org/10.3389/fpsyt.2019.00659

Raucher-Chéné, D., Achim, A. M., Kaladjian, A., & Besche-Richard, C. (2017). Verbal fluency in bipolar disorders: A systematic review and meta-analysis. *Journal of Affective Disorders*, *207*, 359–366. https://doi.org/10.1016/j.jad.2016.09.039

Rezaii, N., Walker, E., & Wolff, P. (2019). A machine learning approach to predicting psychosis using semantic density and latent content analysis. *NPJ Schizophrenia*, *5*(9). https://doi.org/10.1038/s41537-019-0077-9

Robin, J., Harrison, J. E., Kaufman, L. D., Rudzicz, F., Simpson, W., & Yancheva, M. (2020). Evaluation of Speech-Based Digital Biomarkers: Review and Recommendations. *Digital Biomarkers*, *4*(3), 99–108. https://doi.org/10.1159/000510820

Rosenberg, S. D. (1979). Verbal Behavior and Schizophrenia: The Semantic Dimension. *Archives of General Psychiatry*, *36*(12), 1331. https://doi.org/10.1001/archpsyc.1979.01780120061008

Rosenberg, S. D., Schnurr, P. P., & Oxman, T. E. (1990). Content Analysis: A Comparison of Manual and Computerized Systems. *Journal of Personality Assessment*, *54*(1 & 2), 298–310.

Rozenzweig, N., Vandenberg, S. G., Moore, K., & Dukay, A. (1961). A Study of the Reliability of the Mental Status Examination. *The American Journal of Psychiatry*, *117*(12), 1102–1108.

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., Markowitz, J. C., Ninan, P. T., Kornstein, S., Manber, R., Thase, M. E., Kocsis, J. H., & Keller, M. B. (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. *Biological Psychiatry*, *54*(5), 573–583. https://doi.org/10.1016/S0006-3223(02)01866-8

Sani, G., Vöhringer, P. A., Barroilhet, S. A., Koukopoulos, A. E., & Ghaemi, S. N. (2018). The Koukopoulos Mixed Depression Rating Scale (KMDRS): An International Mood Network (IMN) validation study of a new mixed mood rating scale. *Journal of Affective Disorders*, *232*, 9–16. https://doi.org/10.1016/j.jad.2018.01.025

Scherer, S., Morency, L.-P., Gratch, J., & Pestian, J. (2015). Reduced vowel space is a robust indicator of psychological distress: A cross-corpus analysis. *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4789–4793. https://doi.org/10.1109/ICASSP.2015.7178880

Schweizer, T. H., Snyder, H. R., Young, J. F., & Hankin, B. L. (2020). The Breadth and Potency of Transdiagnostic Cognitive Risks for Psychopathology in Youth. *Journal of Consulting and Clinical Psychology*, *88*(3), 196–211. https://doi.org/10.1037/ccp0000470

Schweizer, T. H., Snyder, H. R., Young, J. F., & Hankin, B. L. (2021). Prospective Prediction of Depression and Anxiety by Integrating Negative Emotionality and Cognitive Vulnerabilities in Children and Adolescents. *Research on Child and Adolescent Psychopathology*. https://doi.org/10.1007/s10802-021-00839-4

Scott, E. M., Hermens, D. F., Naismith, S. L., Guastella, A. J., De Regt, T., White, D., Lagopoulos, J., & Hickie, I. B. (2013). Distinguishing young people with emerging bipolar disorders from those with unipolar depression. *Journal of Affective Disorders*, *144*(3), 208–215. https://doi.org/10.1016/j.jad.2012.06.031

Serby, M. (2003). Psychiatric Resident Conceptualizations of Mood and Affect Within the Mental Status Examination. *American Journal of Psychiatry*, *160*(8), 1527–1529. https://doi.org/10.1176/appi.ajp.160.8.1527

Smith, M., Dietrich, B. J., Bai, E., & Bockholt, H. J. (2020). Vocal pattern detection of depression among older adults. *International Journal of Mental Health Nursing*, *29*(3), 440–449. https://doi.org/10.1111/inm.12678

Snowden, L. R. (2003). Bias in Mental Health Assessment and Intervention: Theory and Evidence. *American Journal of Public Health*, *93*(2), 239–243.

Solomon, C., Valstar, M. F., Morriss, R. K., & Crowe, J. (2015). Objective Methods for Reliable Detection of Concealed Depression. *Frontiers in ICT*, *2*. https://doi.org/10.3389/fict.2015.00005

Sonnenschein, A. R., Hofmann, S. G., Ziegelmayer, T., & Lutz, W. (2018). Linguistic analysis of patients with mood and anxiety disorders during cognitive behavioral therapy. *Cognitive Behaviour Therapy*, *47*(4), 315–327. https://doi.org/10.1080/16506073.2017.1419505

Spinhoven, P., van Hemert, A. M., & Penninx, B. W. (2018). Repetitive negative thinking as a predictor of depression and anxiety: A longitudinal cohort study. *Journal of Affective Disorders*, *241*, 216–225. https://doi.org/10.1016/j.jad.2018.08.037

Stassen, H. H., & Bomben, E. G. (1991). Speech Characteristics in Depression. *Psychopathology*, *24*(2), 88–105.

Stassen, H. H., Kuny, S., & Hell, D. (1998). The speech analysis approach to determining onset of improvement under antidepressants. *European Neuropsychopharmacology*, *8*(4), 303–310. https://doi.org/10.1016/S0924-977X(97)00090-4

Straw, I., & Callison-Burch, C. (2020). Artificial Intelligence in mental health and the biases of language based models. *PLoS ONE*, *15*(12). https://doi.org/10.1371/journal.pone.0240376

Struijs, S. Y., Lamers, F., Spinhoven, P., van der Does, W., & Penninx, B. W. J. H. (2018). The predictive specificity of psychological vulnerability markers for the course of affective disorders. *Journal of Psychiatric Research*, *103*, 10–17. https://doi.org/10.1016/j.jpsychires.2018.04.017

Su, M.-H., Wu, C.-H., Huang, K.-Y., & Yang, T.-H. (2020). Cell-Coupled Long Short-Term Memory With L -Skip Fusion Mechanism for Mood Disorder Detection Through Elicited Audiovisual Features. *IEEE Transactions on Neural Networks and Learning Systems*, *31*(1), 124–135. https://doi.org/10.1109/TNNLS.2019.2899884

Sun, T., Gaut, A., Tang, S., Huang, Y., ElSherief, M., Zhao, J., Mirza, D., Belding, E., Chang, K.-W., & Wang, W. Y. (2019). Mitigating Gender Bias in Natural Language Processing: Literature Review. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 1630–1640. https://doi.org/10.18653/v1/P19-1159

Swann, A. C. (2013). Activated Depression: Mixed Bipolar Disorder or Agitated Unipolar Depression? *Current Psychiatry Reports*, *15*(8), 376. https://doi.org/10.1007/s11920-013-0376-1

Sweeney, P. D., Anderson, K., & Bailey, S. (1986). Attributional Style in Depression: A Meta-Analytic Review. *Journal of Personality and Social Psychology*, *50*(5), 974–991.

Szabadi, E., Bradshaw, C. M., & Besson, J. A. O. (1976). Elongation of Pause-Time in Speech: A Simple, Objective Measure of Motor Retardation in Depression. *British Journal of Psychiatry*, *129*(6), 592–597. https://doi.org/10.1192/bjp.129.6.592

Tasnim, M., & Stroulia, E. (2019). Detecting Depression from Voice. In M.-J. Meurs & F. Rudzicz (Eds.), *Advances in Artificial Intelligence* (Vol. 11489, pp. 472–478). Springer International Publishing. https://doi.org/10.1007/978-3-030-18305-9_47

Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. https://doi.org/10.1177/0261927X09351676

Teixeira, J. P., Oliveira, C., & Lopes, C. (2013). Vocal Acoustic Analysis – Jitter, Shimmer and HNR Parameters. *Procedia Technology*, *9*, 1112–1122. https://doi.org/10.1016/j.protcy.2013.12.124

Thieme, A., Belgrave, D., & Doherty, G. (2020). Machine Learning in Mental Health: A Systematic Review of the HCI Literature to Support the Development of Effective and Implementable ML Systems. *ACM Transactions on Computer-Human Interaction*, *27*(5), 1–53. https://doi.org/10.1145/3398069

Toazza, R., Salum, G. A., Jarros, R. B., DeSousa, D., Salles, J. F. de, & Manfro, G. G. (2016). Phonemic verbal fluency and severity of anxiety disorders in young children. *Trends in Psychiatry and Psychotherapy*, *38*(2), 100–104. https://doi.org/10.1590/2237-6089-2016-0018

Uher, R. (2016). Person- centered measurement- based care for depression. *World Psychiatry*, *15*(3), 238–239. https://doi.org/10.1002/wps.20363

Uher, R., Cumby, J., MacKenzie, L. E., Morash-Conway, J., Glover, J. M., Aylott, A., Propper, L., Abidi, S., Bagnell, A., Pavlova, B., Hajek, T., Lovas, D., Pajer, K., Gardner, W., Levy, A., & Alda, M. (2014). A familial risk enriched cohort as a platform for testing early interventions to prevent severe mental illness. *BMC Psychiatry*, *14*(1), 344. https://doi.org/10.1186/s12888-014-0344-2

Vai, B., Parenti, L., Bollettini, I., Cara, C., Verga, C., Melloni, E., Mazza, E., Poletti, S., Colombo, C., & Benedetti, F. (2020). Predicting differential diagnosis between bipolar and unipolar depression with multiple kernel learning on multimodal structural neuroimaging. *European Neuropsychopharmacology*, *34*, 28–38. https://doi.org/10.1016/j.euroneuro.2020.03.008

Vanello, N., Guidi, A., Gentili, C., Werner, S., Bertschy, G., Valenza, G., Lanata, A., & Scilingo, E. P. (2012). Speech analysis for mood state characterization in bipolar patients. *2012*

Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2104–2107. https://doi.org/10.1109/EMBC.2012.6346375

Wang, J., Zhang, L., Liu, T., Pan, W., Hu, B., & Zhu, T. (2019). Acoustic differences between healthy and depressed people: A cross-situation study. *BMC Psychiatry*, *19*(1). https://doi.org/10.1186/s12888-019-2300-7

Wehr, T. A., Sack, D. A., Rosenthal, N. E., & Cowdry, R. W. (1988). Rapid Cycling Affective Disorder: Contributing Factors and Treatment Responses in 51 Patients. *American Journal of Psychiatry*, *145*(2), 179–184.

Willits, J. A., Rubin, T., Jones, M. N., Minor, K. S., & Lysaker, P. H. (2018). Evidence of disturbances of deep levels of semantic cohesion within personal narratives in schizophrenia. *Schizophrenia Research*, *197*, 365–369. https://doi.org/10.1016/j.schres.2017.11.014

Young, R. C., Biggs, J. T., Ziegler, V. E., & Meyer, D. A. (1978). A Rating Scale for Mania: Reliability, Validity and Sensitivity. *British Journal of Psychiatry*, *133*(5), 429–435. https://doi.org/10.1192/bjp.133.5.429

Zhang, J., Pan, Z., Gui, C., Xue, T., Lin, Y., Zhu, J., & Cui, D. (2018). Analysis on speech signal features of manic patients. *Journal of Psychiatric Research*, *98*, 59–63. https://doi.org/10.1016/j.jpsychires.2017.12.012

Zhang, X., Li, W., Zhao, N., Cheung, T., Ungvari, G. S., Wang, G., Xiang, Y.-T., & Angst, J. (2021). Use of the 33-Item Hypomania Checklist (HCL-33) to Distinguish Bipolar Disorder From Major Depressive Disorder in Older Adults. *Journal of Geriatric Psychiatry and Neurology*, 08919887211016065. https://doi.org/10.1177/08919887211016065

Zheng, Y., He, S., Zhang, T., Lin, Z., Shi, S., Fang, Y., Jiang, K., & Liu, X. (2019). Detection Study of Bipolar Depression Through the Application of a Model-Based Algorithm in

Terms of Clinical Feature and Peripheral Biomarkers. *Frontiers in Psychiatry*, *10*. https://doi.org/10.3389/fpsyt.2019.00266

Zimmermann, J., Brockmeyer, T., Hunn, M., Schauenburg, H., & Wolf, M. (2017). First-person Pronoun Use in Spoken Language as a Predictor of Future Depressive Symptoms: Preliminary Evidence from a Clinical Sample of Depressed Patients: First-person Pronoun Use in Spoken Language. *Clinical Psychology & Psychotherapy*, *24*(2), 384–391. https://doi.org/10.1002/cpp.2006

Appendix A

Section A: Systematic Search Methods

Table 3

Inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
Sample of participants with a psychiatric disorder	Only control or population sample (no psychiatric diagnoses), substance abuse, participants with neurological illness, brain injury or trauma, intellectual disability or dementia
Focused on speech as it related to one or more psychiatric disorders	Looked at speech as an indicator of mood in a non-psychiatric population, speech used to assess language disorders
Incorporated speech collection and/or analysis, or analysis of transcriptions of speech samples	Used only text-based analysis – participants' writing, clinician's notes, suicide notes, hospital charts, discharge reports
Original-data, peer-reviewed articles	Reviews, case studies, narrative analyses

Search strategy:

The search was conducted by amalgamating the search strategies for four key concepts and with the boolean operator "AND" on PubMed and PsycInfo.

Concept 1: Diagnosis and prediction

(TI (diagnos* OR detect* OR predict* OR prognos* OR assess* OR Sensitiv*) OR AB (diagnos* OR detect* OR predict* OR prognos* OR assess* OR Sensitiv*)) OR ((((DE "Clinical Judgment (Not Diagnosis)") OR (DE "Prognosis")) OR (DE "Diagnosis") OR (DE "Psychodiagnosis")) OR (DE "Psychological Assessment" OR DE "Behavioral Assessment" OR DE "Cognitive Assessment" OR DE "Emotional Assessment" OR DE "Motivation Measures" OR DE "Neuropsychological Assessment" OR DE "Q-Sort" OR DE "Stress and Coping Measures"))

Concept 2: Mood disorders

(((((DE "Mental Health" OR DE "Community Mental Health") OR (DE "Affective Disorders" OR DE "Bipolar Disorder" OR DE "Disruptive Mood Dysregulation Disorder" OR DE "Major Depression" OR DE "Mania" OR DE "Seasonal Affective Disorder")) OR (DE "Schizophrenia" OR DE "Acute Schizophrenia" OR DE "Catatonic Schizophrenia" OR DE "Childhood Schizophrenia" OR DE "Paranoid Schizophrenia" OR DE "Process Schizophrenia" OR DE "Schizophrenia (Disorganized Type)" OR DE "Schizophreniform Disorder" OR DE "Undifferentiated Schizophrenia")) AND (DE "Suicide" OR DE "Attempted Suicide")) OR (DE "Mental Disorders" OR DE "Adjustment Disorders" OR DE "Affective Disorders" OR DE "Anxiety Disorders" OR DE "Chronic Mental Illness" OR DE "Dissociative Disorders" OR DE "Eating Disorders" OR DE "Factitious Disorders" OR DE "Gender Identity Disorder" OR DE "Neurosis" OR DE "Personality Disorders" OR DE "Psychosis" OR DE "Schizoaffective Disorder"))

OR (TI (depress* OR bipolar* OR mania* OR manic* OR schizophren* OR psychos* OR psychotic* OR "mood disorder*" OR "psychiatric illness*") OR AB (depress* OR bipolar* OR mania* OR manic* OR schizophren* OR psychos* OR psychotic* OR

"mood disorder*" OR "psychiatric illness*" OR obsessive compulsive disorder* OR "OCD"))

Concept 3: Speech

(((DE "Verbal Communication" OR DE "Articulation (Speech)" OR DE "Conversation" OR DE "Language Proficiency" OR DE "Oral Communication" OR DE "Pragmatics") OR (DE "Verbal Meaning" OR DE "Word Meaning")) OR (DE "Language" OR DE "Figurative Language" OR DE "Form Classes (Language)" OR DE "Native Language" OR DE "Natural Language" OR DE "Phrases" OR DE "Rhetoric" OR DE "Sentences" OR DE "Spelling" OR DE "Vocabulary")) OR (TI (speech* OR "natural language" OR "language processing" OR "language deficit*" OR sentiment* OR "linguistic analys*" OR "voice data") OR AB (speech* OR "natural language" OR "language processing" OR "language deficit*" OR sentiment* OR "linguistic analys*" OR "voice data"))

Concept 4: Automation

DE "Computer Assisted Diagnosis" OR TI (comput* OR "machine learning" OR automat* OR software OR "graph analysis" OR "artificial intelligence") OR AB (comput* OR "machine learning" OR automat* OR software OR "graph analysis" OR "artificial intelligence")

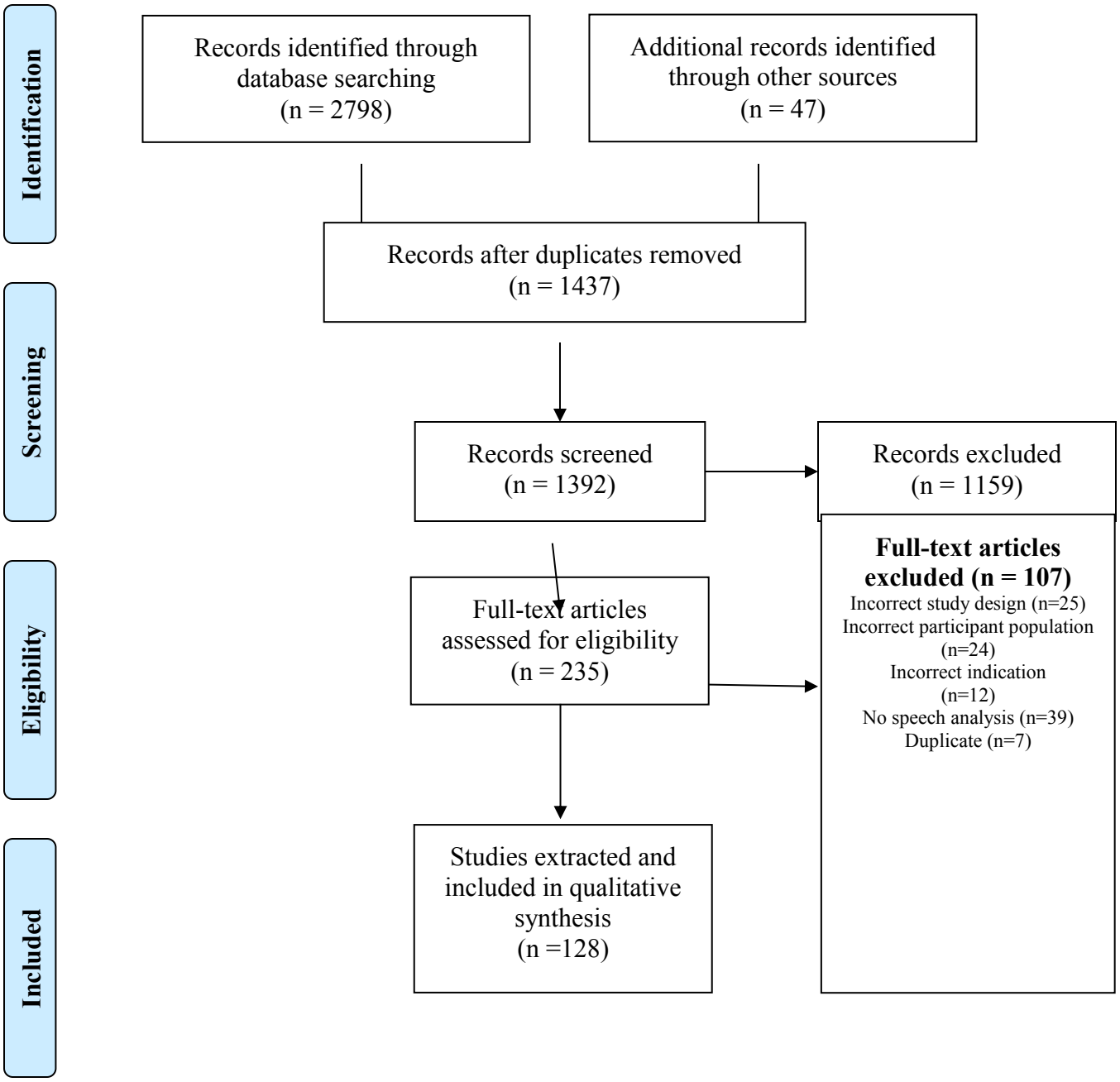
Data Extraction:

Search results from both databases were uploaded into Covidence, an online screening application. The first author screened all articles for eligibility with title and abstract, and the first and second author screened each subsequently included full-text for any exclusion criteria. The search yielded 2,776 unique publications. After title and abstract screening, 215 underwent full text review and 123 met eligibility criteria. After

screening, the following variables were extracted from each article: title, author(s), year of publication, journal, disorder (depression, schizophrenia/psychosis, bipolar disorder, anxiety, mixed diagnosis), outcome, speech feature, number of participants (patient and control, male and female), analysis methods, disorder feature (e.g. speech disorganization, self-focused attention, affect, loosening associations), language spoken by participants, and whether or not analysis included machine learning classification. Outcomes were divided into six categories depending on the focus of study: Classifying diagnoses, assessing severity, predicting illness onset, predicting treatment outcome, prognosis, and other (e.g. assessing cognition). Speech features were originally divided into four categories (acoustics and prosody were later collapsed into one category): Acoustics, prosody, semantics, and content. Analysis methods were extracted for two different stages: speech sampling methods and speech analysis methods.

Figure 3

Consort diagram showing flow of searching, screening, and extraction of studies



Section B: Reference list of systematic search results

Alghowinem, S. (2015). *Multimodal Analysis of Verbal and Nonverbal Behaviour on the Example of Clinical Depression* [PhD]. The Australian National University.

Alhanai, T., Ghassemi, M., & Glass, J. (2018). Detecting Depression with Audio/Text Sequence Modeling of Interviews. *Interspeech 2018*, 1716–1720. https://doi.org/10.21437/Interspeech.2018-2522

Allen, H. A. (1983). Do positive symptom and negative symptom subtypes of schizophrenia show qualitative differences in language production. *Psychological Medicine*, *13*(4), 787–797.

Alpert, M., Pouget, E. R., & Silva, R. R. (1995). Cues to the Assessment of Affects and Moods: Speech Fluency and Pausing. *Psychopharmacological Bulletin*, *31*(2), 421–424.

Alpert, M., Pouget, E. R., & Silva, R. R. (2001). Reflections of depression in acoustic measures of the patient's speech. *Journal of Affective Disorders*, *66*(1), 59–69. https://doi.org/10.1016/S0165-0327(00)00335-9

Alpert, M., Rosenberg, S. D., Pouget, E. R., & Shaw, R. J. (2000). Prosody and lexical accuracy in flat affect schizophrenia. *Psychiatry Research*, *97*(2–3), 107–118. https://doi.org/10.1016/S0165-1781(00)00231-6

Arevian, A. C., Bone, D., Malandrakis, N., Martinez, V. R., Wells, K. B., Miklowitz, D. J., & Narayanan, S. (2020). Clinical state tracking in serious mental illness through computational analysis of speech. *PLoS ONE*, *15*(1). https://doi.org/10.1371/journal.pone.0225695

Ayer, A., Yalınçetin, B., Aydınlı, E., Sevilmiş, Ş., Ulaş, H., Binbay, T., Akdede, B. B., & Alptekin, K. (2016). Formal thought disorder in first-episode psychosis. *Comprehensive Psychiatry*, *70*, 209–215. https://doi.org/10.1016/j.comppsych.2016.08.005

Bearden, C. E., Rosso, I. M., Hollister, J. M., Sanchez, L. E., Hadley, T., & Cannon, T. D. (2000). A Prospective Cohort Study of Childhood Behavioral Deviance and Language Abnormalities as Predictors of Adult Schizophrenia. *Schizophrenia Bulletin*, *26*(2), 395–410. https://doi.org/10.1093/oxfordjournals.schbul.a033461

Bedi, G., Carrillo, F., Cecchi, G. A., Slezak, D. F., Sigman, M., Mota, N. B., Ribeiro, S., Javitt, D. C., Copelli, M., & Corcoran, C. M. (2015). Automated analysis of free speech predicts psychosis onset in high-risk youths. *Npj Schizophrenia*, *1*(1). https://doi.org/10.1038/npjschz.2015.30

Bonfils, K. A., Luther, L., Firmin, R. L., Lysaker, P. H., Minor, K. S., & Salyers, M. P. (2016). Language and hope in schizophrenia-spectrum disorders. *Psychiatry Research*, *245*, 8–14. https://doi.org/10.1016/j.psychres.2016.08.013

Braun, S., Annovazzi, C., Botella, C., Bridler, R., Camussi, E., Delfino, J. P., Mohr, C., Moragrega, I., Papagno, C., Pisoni, A., Soler, C., Seifritz, E., & Stassen, H. H. (2016). Assessing Chronic Stress, Coping Skills, and Mood Disorders through Speech Analysis: A Self-Assessment "Voice App" for laptops, Tablets and Smartphones. *Psychopathology*, *49*(6), 406–419. https://doi.org/10.1159/000450959

Breznitz, Z. (1992). Verbal Indicators of Depression. *The Journal of General Psychology*, *119*(4), 351–363.

Brockmeyer, T., Zimmermann, J., Kulessa, D., Hautzinger, M., Bents, H., Friederich, H.-C., Herzog, W., & Backenstrass, M. (2015). Me, myself, and I: Self-referent word use as an indicator of self-focused attention in relation to depression and anxiety. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01564

Buck, B., Minor, K. S., & Lysaker, P. H. (2015a). Differential lexical correlates of social cognition and metacognition in schizophrenia; a study of spontaneously-generated life narratives. *Comprehensive Psychiatry*, *58*, 138–145. https://doi.org/10.1016/j.comppsych.2014.12.015

Buck, B., Minor, K. S., & Lysaker, P. H. (2015b). Lexical Characteristics of Anticipatory and Consummatory Anhedonia in Schizophrenia: A Study of Language in Spontaneous Life Narratives: Lexical Characteristics of Anhedonia in Schizophrenia. *Journal of Clinical Psychology*, *71*(7), 696–706. https://doi.org/10.1002/jclp.22160

Buck, B., & Penn, D. L. (2015). Lexical Characteristics of Emotional Narratives in Schizophrenia: Relationships With Symptoms, Functioning, and Social Cognition. *The Journal of Nervous and Mental Disease*, *203*(9), 702–708. https://doi.org/10.1097/NMD.0000000000000354

Cabana, Á., Valle-Lisboa, J. C., Elvevåg, B., & Mizraji, E. (2011). Detecting order–disorder transitions in discourse: Implications for schizophrenia. *Schizophrenia Research*, *131*(1–3), 157–164. https://doi.org/10.1016/j.schres.2011.04.026

Carrillo, F., Sigman, M., Fernández Slezak, D., Ashton, P., Fitzgerald, L., Stroud, J., Nutt, D. J., & Carhart-Harris, R. L. (2018). Natural speech algorithm applied to baseline interview data can predict which patients will respond to psilocybin for treatment-resistant depression. *Journal of Affective Disorders*, *230*, 84–86. https://doi.org/10.1016/j.jad.2018.01.006

Ciccarelli, G. A. (2009). Characterization of Phoneme Rate as a Vocal Biomarker of Depression. *Electrical Engineering*, 162.

Cohen, A. S., Alpert, M., Nienow, T. M., Dinzeo, T. J., & Docherty, N. M. (2008). Computerized measurement of negative symptoms in schizophrenia. *Journal of Psychiatric Research*, *42*(10), 827–836. https://doi.org/10.1016/j.jpsychires.2007.08.008

Cohen, A. S., Kim, Y., & Najolia, G. M. (2013). Psychiatric symptom versus neurocognitive correlates of diminished expressivity in schizophrenia and mood disorders. *Schizophrenia Research*, *146*(1–3), 249–253. https://doi.org/10.1016/j.schres.2013.02.002

Cohen, A. S., McGovern, J. E., Dinzeo, T. J., & Covington, M. A. (2014). Speech deficits in serious mental illness: A cognitive resource issue? *Schizophrenia Research*, *160*(1–3), 173–179. https://doi.org/10.1016/j.schres.2014.10.032

Cohen, A. S., Mitchell, K. R., Docherty, N. M., & Horan, W. P. (2016). Vocal expression in schizophrenia: Less than meets the ear. *Journal of Abnormal Psychology*, *125*(2), 299–309. https://doi.org/10.1037/abn0000136

Cohen, A. S., Najolia, G. M., Kim, Y., & Dinzeo, T. J. (2012). On the boundaries of blunt affect/alogia across severe mental illness: Implications for Research Domain Criteria. *Schizophrenia Research*, *140*(1–3), 41–45. https://doi.org/10.1016/j.schres.2012.07.001

Compton, M. T., Lunden, A., Cleary, S. D., Pauselli, L., Alolayan, Y., Halpern, B., Broussard, B., Crisafio, A., Capulong, L., Balducci, P. M., Bernardini, F., & Covington, M. A. (2018). The aprosody of schizophrenia: Computationally derived acoustic phonetic underpinnings of monotone speech. *Schizophrenia Research*, *197*, 392–399. https://doi.org/10.1016/j.schres.2018.01.007

Corcoran, C. M., Carrillo, F., Fernández-Slezak, D., Bedi, G., Klim, C., Javitt, D. C., Bearden, C. E., & Cecchi, G. A. (2018). Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World Psychiatry*, *17*(1), 67–75. https://doi.org/10.1002/wps.20491

Cummins, N., Epps, J., & Ambikairajah, E. (2013). Spectro-temporal analysis of speech affected by depression and psychomotor retardation. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 7542–7546. https://doi.org/10.1109/ICASSP.2013.6639129

Cummins, N., Epps, J., Breakspear, M., & Goecke, R. (2011). An Investigation of Depressed Speech Detection: Features and Normalization. *Interspeech 2011*, 2997–3000.

Cummins, N., Epps, J., Sethu, V., Breakspear, M., & Goecke, R. (2013). Modeling Spectral Variability for the Classification of Depressed Speech. *Interspeech 2013*, 857–861.

Cummins, N., Sethu, V., Epps, J., Schnieder, S., & Krajewski, J. (2015). Analysis of acoustic space variability in speech affected by depression. *Speech Communication*, *75*, 27–49. https://doi.org/10.1016/j.specom.2015.09.003

de la Parra, G., Mergenthaler, E., & Kaechele, H. (1988). Computarized analysis of patients', and therapists' talking behavior in first-interview. *Acta psiquitrica y psicoliga de America Latina*, *34*(4), 309–320.

Dwyer, K., David, A. S., McCarthy, R., McKenna, P., & Peters, E. (2019). Linguistic alignment and theory of mind impairments in schizophrenia patients' dialogic interactions. *Psychological Medicine*, 1–9. https://doi.org/10.1017/S0033291719002289

Elvevåg, B., Foltz, P. W., Rosenstein, M., & DeLisi, L. E. (2010). An automated method to analyze language use in patients with schizophrenia and their first-degree relatives. *Journal of Neurolinguistics*, *23*(3), 270–284. https://doi.org/10.1016/j.jneuroling.2009.05.002

Elvevåg, B., Foltz, P. W., Weinberger, D. R., & Goldberg, T. E. (2007). Quantifying incoherence in speech: An automated methodology and novel application to schizophrenia. *Schizophrenia Research*, *93*(1–3), 304–316. https://doi.org/10.1016/j.schres.2007.03.001

Erol, A., Mete, L., Doganer, I., & Sorias, S. (1995). The Use of Computerized Content Analysis Method for Psychiatric Diagnosis. *Turkish Journal of Psychiatry*, *6*(4), 263–268.

Erol, A., Sorias, O., Mete, L., & Sorias, S. (1995). Comparison of TAT Responses and Free Speech Examples with Computerized Content Analysis. *Turkish Journal of Psychiatry*, *6*(2), 89–97.

Faurholt-Jepsen, M., Busk, J., Frost, M., Vinberg, M., Christensen, E. M., Winther, O., Bardram, J. E., & Kessing, L. V. (2016). Voice analysis as an objective state marker in bipolar disorder. *Translational Psychiatry*, *6*(7), e856–e856. https://doi.org/10.1038/tp.2016.123

Fineberg, S. K., Leavitt, J., Deutsch-Link, S., Dealy, S., Landry, C. D., Pirruccio, K., Shea, S., Trent, S., Cecchi, G., & Corlett, P. R. (2016). Self-reference in psychosis and depression: A language marker of illness. *Psychological Medicine*, *46*(12), 2605–2615. https://doi.org/10.1017/S0033291716001215

Flint, A. J., Black, S. E., Campbell-Taylor, I., Gailey, G. F., & Levinton, C. (1992). Acoustic analysis in the differentiation of Parkinson's disease and major depression. *Journal of Psycholinguistic Research*, *21*(5), 383–399. https://doi.org/10.1007/BF01067922

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. *Journal of Psychiatric Research*, *27*(3), 309–319.

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 Transactions on Biomedical Engineering*, *47*(7), 829–837. https://doi.org/10.1109/10.846676

Fraser, W. I., King, K. M., Thomas, P., & Kendell, R. E. (1986). The Diagnosis of Schizophrenia by Language Analysis. *British Journal of Psychiatry*, *148*(3), 275–278. https://doi.org/10.1192/bjp.148.3.275

Gideon, J., Provost, E. M., & McInnis, M. (2016). Mood state prediction from speech of varying acoustic quality for individuals with bipolar disorder. *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2359–2363. https://doi.org/10.1109/ICASSP.2016.7472099

Gottschalk, L. A., Stein, M. K., & Shapiro, D. H. (1997). The application of computerized content analysis of speech to the diagnostic process in a psychiatric outpatient clinic. *Journal of Clinical Psychology*, 15.

Haas, S. S., Doucet, G. E., Garg, S., Herrera, S. N., Sarac, C., Bilgrami, Z. R., Shaik, R. B., & Corcoran, C. M. (2020). Linking language features to clinical symptoms and multimodal imaging in individuals at clinical high risk for psychosis. *European Psychiatry*, *63*(1). https://doi.org/10.1192/j.eurpsy.2020.73

Hagenaars, M. A., & van Minnen, A. (2005). The effect of fear on paralinguistic aspects of speech in patients with panic disorder with agoraphobia. *Journal of Anxiety Disorders*, *19*(5), 521–537. https://doi.org/10.1016/j.janxdis.2004.04.008

Hanakawa, Y. (2004a). *Comparison of Linguistic and Acoustic Analyses of Schizophrenic Language* [PhD]. Adelphi University.

Hanakawa, Y. (2004b). Patterns of schizophrenic language, analyzed by computerized content analysis. *Journal of the American Psychoanalytic Association*, *52*(2), 475–476.

Harati, S., Crowell, A., Mayberg, H., & Nemati, S. (2018). Depression Severity Classification from Speech Emotion. *2018 40th Annual International Conference of the IEEE*

Engineering in Medicine and Biology Society (EMBC), 5763–5766. https://doi.org/10.1109/EMBC.2018.8513610

Hashim, N. W., Wilkes, M., Salomon, R., Meggs, J., & France, D. J. (2017). Evaluation of Voice Acoustics as Predictors of Clinical Depression Scores. *Journal of Voice*, *31*(2), 256.e1-256.e6. https://doi.org/10.1016/j.jvoice.2016.06.006

He, L., & Cao, C. (2018). Automated depression analysis using convolutional neural networks from speech. *Journal of Biomedical Informatics*, *83*, 103–111. https://doi.org/10.1016/j.jbi.2018.05.007

Himmelstein, P., Barb, S., Finlayson, M. A., & Young, K. D. (2018). Linguistic analysis of the autobiographical memories of individuals with major depressive disorder. *PLOS ONE*, *13*(11), e0207814. https://doi.org/10.1371/journal.pone.0207814

Hoffman, G. M. A., Gonze, J. C., & Mendlewicz, J. (1985). Speech Pause Time as a Method for the Evaluation of Psychomotor Retardation in Depressive Illness. *British Journal of Psychiatry*, *146*(5), 535–538. https://doi.org/10.1192/bjp.146.5.535

Hoffman, R. E. (1986). A Comparative Study of Manic vs Schizophrenic Speech Disorganization. *Archives of General Psychiatry*, *43*(9), 831–838. https://doi.org/10.1001/archpsyc.1986.01800090017003

Holshausen, K., Harvey, P. D., Elvevåg, B., Foltz, P. W., & Bowie, C. R. (2014). Latent semantic variables are associated with formal thought disorder and adaptive behavior in older inpatients with schizophrenia. *Cortex*, *55*, 88–96. https://doi.org/10.1016/j.cortex.2013.02.006

Hong, K., Nenkova, A., March, M. E., Parker, A. P., Verma, R., & Kohler, C. G. (2015). Lexical use in emotional autobiographical narratives of persons with schizophrenia and healthy controls. *Psychiatry Research*, *225*(1–2), 40–49. https://doi.org/10.1016/j.psychres.2014.10.002

Honig, F., Batliner, A., Noth, E., Schnieder, S., & Krajewski, J. (2014). Automatic Modelling of Depressed Speech: Relevant Features and Relevance of Gender. *Interspeech 2014*, 1248–1252.

Huang, K.-Y., Wu, C.-H., Su, M.-H., & Kuo, Y.-T. (2020). Detecting Unipolar and Bipolar Depressive Disorders from Elicited Speech Responses Using Latent Affective Structure Model. *IEEE Transactions on Affective Computing*, *11*(3), 393–404. https://doi.org/10.1109/TAFFC.2018.2803178

Huston, J., Meier, S., Faith, M., & Reynolds, A. (2019). Exploratory study of automated linguistic analysis for progress monitoring and outcome assessment. *Counselling and Psychotherapy Research*, *19*(3), 321–328. https://doi.org/10.1002/capr.12219

Jiang, H., Hu, B., Liu, Z., Wang, G., Zhang, L., Li, X., & Kang, H. (2018). Detecting Depression Using an Ensemble Logistic Regression Model Based on Multiple Speech Features. *Computational and Mathematical Methods in Medicine*, *2018*. https://doi.org/10.1155/2018/6508319

Jiang, H., Hu, B., Liu, Z., Yan, L., Wang, T., Liu, F., Kang, H., & Li, X. (2017). Investigation of different speech types and emotions for detecting depression using different classifiers. *Speech Communication*, *90*, 39–46. https://doi.org/10.1016/j.specom.2017.04.001

Just, S. A., Haegert, E., Kořánová, N., Bröcker, A.-L., Nenchev, I., Funcke, J., Heinz, A., Bermpohl, F., Stede, M., & Montag, C. (2020). Modeling Incoherent Discourse in Non-Affective Psychosis. *Frontiers in Psychiatry*, *11*. https://doi.org/10.3389/fpsyt.2020.00846

King, K., Fraser, W. I., Thomas, P., & Kendell, R. E. (1990). Re-examination of the Language of Psychotic Subjects. *British Journal of Psychiatry*, *156*(2), 211–215. https://doi.org/10.1192/bjp.156.2.211

Kotsopoulos, S., & Mellor, C. (1986). Extralinguistic Speech Characteristics of Children with Conduct and Anxiety Disorders. *Journal of Child Psychology and Psychiatry*, *27*(1), 99–108. https://doi.org/10.1111/j.1469-7610.1986.tb00625.x

Kuny, S., Stassen, H. H., & Hell, D. (n.d.). Cognitive impairments in depression: A 2-year follow-up study of 30 patients. *Swiss Archives of Neurology, Psychiatry and Psychotherapy*, *148*(3), 95–102.

Leont'ev, A. A., Rokhlin, L. L., Savitskaia, B. L., & Shakhnarovich, A. M. (1973). A clinico-psycholinguistic study of the faulty speech of patients with the paranoid form of schizophrenia. *The Korsakov's Journal of Neurology and Psychiatry*, *73*(12), 1858–1863.

Loas, G., Salinas, E., Pierson, A., Guelfi, J. D., & Samuel-Lajeunesse, B. (1994). Anhedonia and Blunted Affect in Major Depressive Disorder. *Comprehensive Psychiatry*, *35*(5), 366–372. https://doi.org/10.1016/0010-440x(94)90277-1

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 Transactions on Biomedical Engineering*, *58*(3), 574–586. https://doi.org/10.1109/TBME.2010.2091640

Lundin, N. B., Hochheiser, J., Minor, K. S., Hetrick, W. P., & Lysaker, P. H. (2019). Piecing together fragments: Linguistic cohesion mediates the relationship between executive function and metacognition in schizophrenia. *Schizophrenia Research*, S0920996419305390. https://doi.org/10.1016/j.schres.2019.11.032

Maher, B. A., Manschreck, T. C., Linnet, J., & Candela, S. (2005). Quantitative assessment of the frequency of normal associations in the utterances of schizophrenia patients and healthy controls. *Schizophrenia Research*, *78*(2–3), 219–224. https://doi.org/10.1016/j.schres.2005.05.017

Manschreck, T. C., Merrill, A. M., Jabbar, G., Chun, J., & DeLisi, L. E. (2012). Frequency of normative word associations in the speech of individuals at familial high-risk for schizophrenia. *Schizophrenia Research*, *140*(1–3), 99–103. https://doi.org/10.1016/j.schres.2012.06.034

Martínez-Sánchez, F., Muela-Martínez, J. A., Cortés-Soto, P., García Meilán, J. J., Vera Ferrándiz, J. A., Egea Caparrós, A., & Pujante Valverde, I. M. (2015). Can the Acoustic Analysis of Expressive Prosody Discriminate Schizophrenia? *The Spanish Journal of Psychology*, *18*. https://doi.org/10.1017/sjp.2015.85

McGinnis, E. W., Anderau, S. P., Hruschak, J., Gurchiek, R. D., Lopez-Duran, N. L., Fitzgerald, K., Rosenblum, K. L., Muzik, M., & McGinnis, R. S. (2019). Giving Voice to Vulnerable Children: Machine Learning Analysis of Speech Detects Anxiety and Depression in Early Childhood. *IEEE Journal of Biomedical and Health Informatics*, *23*(6), 2294–2301. https://doi.org/10.1109/JBHI.2019.2913590

Meaux, L. T., Mitchell, K. R., & Cohen, A. S. (2018). Blunted vocal affect and expression is not associated with schizophrenia: A computerized acoustic analysis of speech under ambiguous conditions. *Comprehensive Psychiatry*, *83*, 84–88. https://doi.org/10.1016/j.comppsych.2018.03.009

Mete, L., Schnurr, P. P., Rosenberg, S. D., Oxman, T. E., Doganer, I., & Sorias, S. (1993). Language content and schizophrenia in acute phase Turkish patients. *Social Psychiatry and Psychiatric Epidemiology*, *28*(6), 275–280. https://doi.org/10.1007/BF00795907

Mete, Levent, Doganer, I., Erol, A., Sorias, S., & Ertas, M. (1993). Contents of Speech in Schizophrenia, Mania and Depression. *Turkish Journal of Psychiatry*, *4*(3).

Miller, N. L. (1996). *The Effect of Antidepressants on Language and Symptoms* [PhD]. Adelphi University.

Minor, K. S., Willits, J. A., Marggraf, M. P., Jones, M. N., & Lysaker, P. H. (2019). Measuring disorganized speech in schizophrenia: Automated analysis explains variance

in cognitive deficits beyond clinician-rated scales. *Psychological Medicine*, *49*(3), 440–448. https://doi.org/10.1017/S0033291718001046

Minor, Kyle S., Bonfils, K. A., Luther, L., Firmin, R. L., Kukla, M., MacLain, V. R., Buck, B., Lysaker, P. H., & Salyers, M. P. (2015). Lexical analysis in schizophrenia: How emotion and social word use informs our understanding of clinical presentation. *Journal of Psychiatric Research*, *64*, 74–78. https://doi.org/10.1016/j.jpsychires.2015.02.024

Moe, A. M., Breitborde, N. J. K., Shakeel, M. K., Gallagher, C. J., & Docherty, N. M. (2016). Idea density in the life-stories of people with schizophrenia: Associations with narrative qualities and psychiatric symptoms. *Schizophrenia Research*, *172*(1–3), 201–205. https://doi.org/10.1016/j.schres.2016.02.016

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 Transactions on Biomedical Engineering*, *55*(1), 96–107. https://doi.org/10.1109/TBME.2007.900562

Moore, E., Clements, M., Peifer, J., & Weisser, L. (2004). Comparing objective feature statistics of speech for classifying clinical depression. *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, *23*, 17–20. https://doi.org/10.1109/IEMBS.2004.1403079

Morice, R. D., & Ingram, J. C. L. (1982). Language Analysis in Schizophrenia: Diagnostic Implications. *Australian & New Zealand Journal of Psychiatry*, *16*(2), 11–21. https://doi.org/10.3109/00048678209161186

Mota, N. B., Copelli, M., & Ribeiro, S. (2017). Thought disorder measured as random speech structure classifies negative symptoms and schizophrenia diagnosis 6 months in advance. *Npj Schizophrenia*, *3*(1). https://doi.org/10.1038/s41537-017-0019-3

Mota, N. B., Vasconcelos, N. A. P., Lemos, N., Pieretti, A. C., Kinouchi, O., Cecchi, G. A., Copelli, M., & Ribeiro, S. (2012). Speech Graphs Provide a Quantitative Measure of Thought Disorder in Psychosis. *PLoS ONE*, *7*(4), e34928. https://doi.org/10.1371/journal.pone.0034928

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. *Journal of Neurolinguistics*, *20*(1), 50–64. https://doi.org/10.1016/j.jneuroling.2006.04.001

Mundt, J. C., Vogel, A. P., Feltner, D. E., & Lenderking, W. R. (2012). Vocal Acoustic Biomarkers of Depression Severity and Treatment Response. *Biological Psychiatry*, *72*(7), 580–587. https://doi.org/10.1016/j.biopsych.2012.03.015

Nilsonne, Å., Sundberg, J., Ternström, S., & Askenfelt, A. (1988). Measuring the rate of change of voice fundamental frequency in fluent speech during mental depression. *The Journal of the Acoustical Society of America*, *83*(2), 716–728. https://doi.org/10.1121/1.396114

Noël-Jorand, M.-C., Reinert, M., Giudicelli, S., & Dassa, D. (2004). Schizophrenia: The Quest for a Minimum Sense of Identity to Ward off Delusional Disorder. *The Canadian Journal of Psychiatry*, *49*(6), 394–397. https://doi.org/10.1177/070674370404900610

Novack, D. L. (2003). *Aspects of Emotional Experssion in Flat Affect Schizophrenia and Depression* [PhD]. New York University.

O'Dell, J. W., & Winder, P. (1975). Evaluation of a Content-Analysis System for Therapeutic Interviews. *Journal of Clinical Psychology*, *31*(4), 737–744. https://doi.org/10.1002/1097-4679(197510)31:4<737::AID-JCLP2270310441>3.0.CO;2-7

Ooi, K. E. B., Lech, M., & Allen, N. B. (2013). Multichannel Weighted Speech Classification System for Prediction of Major Depression in Adolescents. *IEEE Transactions on Biomedical Engineering*, *60*(2), 497–506. https://doi.org/10.1109/TBME.2012.2228646

Oxman, T. E., Rosenberg, S. D., Schnurr, P. P., & Tucker, G. J. (1988). Diagnostic classification through content analysis of patients' speech. *American Journal of Psychiatry*, *145*(4), 464–468. https://doi.org/10.1176/ajp.145.4.464

Oxman, T. E., Rosenberg, S. D., & Tucker, G. J. (1982). The Language of Paranoia. *The American Journal of Psychiatry*, *139*(3), 275–282.

Ozdas, A., Shiavi, R. G., Wilkes, D. M., Silverman, M. K., & Silverman, S. E. (2004). Analysis of Vocal Tract Characteristics for Near-term Suicidal Risk Assessment. *Methods of Information in Medicine*, *43*(01), 36–38. https://doi.org/10.1055/s-0038-1633420

Palaniyappan, L., Mota, N. B., Oowise, S., Balain, V., Copelli, M., Ribeiro, S., & Liddle, P. F. (2019). Speech structure links the neural and socio-behavioural correlates of psychotic disorders. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *88*, 112–120. https://doi.org/10.1016/j.pnpbp.2018.07.007

Pan, Z., Gui, C., Zhang, J., Zhu, J., & Cui, D. (2018). Detecting Manic State of Bipolar Disorder Based on Support Vector Machine and Gaussian Mixture Model Using Spontaneous Speech. *Psychiatry Investigation*, *15*(7), 695–700. https://doi.org/10.30773/pi.2017.12.15

Pauselli, L., Halpern, B., Cleary, S. D., Ku, B., Covington, M. A., & Compton, M. T. (2018). Computational linguistic analysis applied to a semantic fluency task to measure derailment and tangentiality in schizophrenia. *Psychiatry Research*, *263*, 74–79. https://doi.org/10.1016/j.psychres.2018.02.037

Perlini, C., Marini, A., Garzitto, M., Isola, M., Cerruti, S., Marinelli, V., Rambaldelli, G., Ferro, A., Tomelleri, L., Dusi, N., Bellani, M., Tansella, M., Fabbro, F., & Brambilla, P. (2012). Linguistic production and syntactic comprehension in schizophrenia and bipolar disorder. *Acta Psychiatrica Scandinavica*, *126*(5), 363–376. https://doi.org/10.1111/j.1600-0447.2012.01864.x

Pylyshyn, Z. W. (1970). Clinical Correlates of Some Syntactic Features of Patients' Speech. *The Journal of Nervous and Mental Disease*, *150*(4), 307–316.

Quatieri, T. F., & Malyska, N. (2012). Vocal-Source Biomarkers for Depression: A Link to Psychomotor Activity. *Interspeech 2012*, 1059–1062.

Rapcan, V., D'Arcy, S., Yeap, S., Afzal, N., Thakore, J., & Reilly, R. B. (2010). Acoustic and temporal analysis of speech: A potential biomarker for schizophrenia. *Medical Engineering & Physics*, *32*(9), 1074–1079. https://doi.org/10.1016/j.medengphy.2010.07.013

Renfordt, E. (1986). Quantitative Analysis of Speech Behaviour of Depression Patients Under a Drug Therapy. *Psychiatric Developments*, *4*(2), 135–146.

Rezaii, N., Walker, E., & Wolff, P. (2019). A machine learning approach to predicting psychosis using semantic density and latent content analysis. *NPJ Schizophrenia*, *5*(9). https://doi.org/10.1038/s41537-019-0077-9

Rosenberg, S. D. (1979). Verbal Behavior and Schizophrenia: The Semantic Dimension. *Archives of General Psychiatry*, *36*(12), 1331. https://doi.org/10.1001/archpsyc.1979.01780120061008

Rosenberg, S. D., Schnurr, P. P., & Oxman, T. E. (1990). Content Analysis: A Comparison of Manual and Computerized Systems. *Journal of Personality Assessment*, *54*(1 & 2), 298–310.

Savitskaia, A. B., & Sivoz, M. M. (1973). A Clinico-Linguistic Analysis of Neologisms in the Oral Speech of Schizophrenic Patients. *The Korsakov's Journal of Neurology and Psychiatry*, *73*(12), 1863–1869.

Scherer, S., Morency, L.-P., Gratch, J., & Pestian, J. (2015). Reduced vowel space is a robust indicator of psychological distress: A cross-corpus analysis. *2015 IEEE International*

Conference on Acoustics, Speech and Signal Processing (ICASSP), 4789–4793. https://doi.org/10.1109/ICASSP.2015.7178880

Scherer, S., Stratou, G., Gratch, J., & Morency, L.-P. (2013, August). Investigating Voice Quality as a Speaker-Independent Indicator of Depression and PTSD. *Interspeech 2013*. Interspeech 2013 14thAnnual Conference of the International Speech Communication Association, Lyon, France.

Smith, M., Dietrich, B. J., Bai, E., & Bockholt, H. J. (2020). Vocal pattern detection of depression among older adults. *International Journal of Mental Health Nursing*, *29*(3), 440–449. https://doi.org/10.1111/inm.12678

Solomon, C., Valstar, M. F., Morriss, R. K., & Crowe, J. (2015). Objective Methods for Reliable Detection of Concealed Depression. *Frontiers in ICT*, *2*. https://doi.org/10.3389/fict.2015.00005

Sonnenschein, A. R., Hofmann, S. G., Ziegelmayer, T., & Lutz, W. (2018). Linguistic analysis of patients with mood and anxiety disorders during cognitive behavioral therapy. *Cognitive Behaviour Therapy*, *47*(4), 315–327. https://doi.org/10.1080/16506073.2017.1419505

Spencer, T. J., Thompson, B., Oliver, D., Diederen, K., Demjaha, A., Weinstein, S., Morgan, S. E., Day, F., Valmaggia, L., Rutigliano, G., De Micheli, A., Mota, N. B., Fusar-Poli, P., & McGuire, P. (2020). Lower speech connectedness linked to incidence of psychosis in people at clinical high risk. *Schizophrenia Research*, S0920996420304588. https://doi.org/10.1016/j.schres.2020.09.002

Stasak, B., Epps, J., & Goecke, R. (2019). An investigation of linguistic stress and articulatory vowel characteristics for automatic depression classification. *Computer Speech & Language*, *53*, 140–155. https://doi.org/10.1016/j.csl.2018.08.001

Stassen, H. H., Albers, M., Püschel, J., Scharfetter, Ch., Tewesmeier, M., & Woggon, B. (1995). Speaking behavior and voice sound characteristics associated with negative schizophrenia. *Journal of Psychiatric Research*, *29*(4), 277–296. https://doi.org/10.1016/0022-3956(95)00004-O

Stassen, H. H., & Bomben, E. G. (1991). Speech Characteristics in Depression. *Psychopathology*, *24*(2), 88–105.

Stassen, H. H., Kuny, S., & Hell, D. (1998). The speech analysis approach to determining onset of improvement under antidepressants. *European Neuropsychopharmacology*, *8*(4), 303–310. https://doi.org/10.1016/S0924-977X(97)00090-4

St-Hilaire, A., Cohen, A. S., & Docherty, N. M. (2008). Emotion word use in the conversational speech of schizophrenia patients. *Cognitive Neuropsychiatry*, *13*(4), 343–356. https://doi.org/10.1080/13546800802250560

Szabadi, E., Bradshaw, C. M., & Besson, J. A. O. (1976). Elongation of Pause-Time in Speech: A Simple, Objective Measure of Motor Retardation in Depression. *British Journal of Psychiatry*, *129*(6), 592–597. https://doi.org/10.1192/bjp.129.6.592

Tasnim, M., & Stroulia, E. (2019). Detecting Depression from Voice. In M.-J. Meurs & F. Rudzicz (Eds.), *Advances in Artificial Intelligence* (Vol. 11489, pp. 472–478). Springer International Publishing. https://doi.org/10.1007/978-3-030-18305-9_47

Toazza, R., Salum, G. A., Jarros, R. B., DeSousa, D., Salles, J. F. de, & Manfro, G. G. (2016). Phonemic verbal fluency and severity of anxiety disorders in young children. *Trends in Psychiatry and Psychotherapy*, *38*(2), 100–104. https://doi.org/10.1590/2237-6089-2016-0018

Tucker, G. J., & Rosenberg, S. D. (1975). Computer Content Analysis of Schizophrenic Speech: A preliminary Report. *The American Journal of Psychiatry*, *132*(6), 611–616.

Vanello, N., Guidi, A., Gentili, C., Werner, S., Bertschy, G., Valenza, G., Lanata, A., & Scilingo, E. P. (2012). Speech analysis for mood state characterization in bipolar patients. *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2104–2107. https://doi.org/10.1109/EMBC.2012.6346375

Wang, J., Zhang, L., Liu, T., Pan, W., Hu, B., & Zhu, T. (2019). Acoustic differences between healthy and depressed people: A cross-situation study. *BMC Psychiatry*, *19*(1). https://doi.org/10.1186/s12888-019-2300-7

Willits, J. A., Rubin, T., Jones, M. N., Minor, K. S., & Lysaker, P. H. (2018). Evidence of disturbances of deep levels of semantic cohesion within personal narratives in schizophrenia. *Schizophrenia Research*, *197*, 365–369. https://doi.org/10.1016/j.schres.2017.11.014

Zhang, J., Pan, Z., Gui, C., Xue, T., Lin, Y., Zhu, J., & Cui, D. (2018). Analysis on speech signal features of manic patients. *Journal of Psychiatric Research*, *98*, 59–63. https://doi.org/10.1016/j.jpsychires.2017.12.012

Zimmermann, J., Brockmeyer, T., Hunn, M., Schauenburg, H., & Wolf, M. (2017). First-person Pronoun Use in Spoken Language as a Predictor of Future Depressive Symptoms: Preliminary Evidence from a Clinical Sample of Depressed Patients: First-person Pronoun Use in Spoken Language. *Clinical Psychology & Psychotherapy*, *24*(2), 384–391. https://doi.org/10.1002/cpp.2006

Appendix B

Section A: Study Setting

Families Overcoming Risks and Building Opportunities for Well-being (FORBOW) is a longitudinal study enriched for parents with mental illness and their offspring. The Canadian Biomarker Integration Network in Depression (CANBIND) is a national group conducting research in biomarkers for depression. The Canadian Depression Research and Intervention Network (CDRIN) is a group focused on lived-experience informed research and knowledge translation. The Vocal Mind Project (VMP) is speech analysis study exploring the connections between psychopathology and speech features in clinical and community populations.

Speech sampling happens on a repeated basis. Time periods between samples vary depending on study of origin:

FORBOW – 1x/year

CANBIND – 3x/first 16 weeks, 2x/1 year follow-up

CDRIN – 1x

VMP – 2x/16 weeks

Section B: Speech Sample Standard Operating Procedures

Materials:

Recorder:

TASCAM Linear PCM Recorder DR-05

Accessories:

- Tripod – Polaroid 8” Heavy Duty Mini Tripod or AmazonBasics Lightweight Mini Tripod
- USB connector chord - included with recorder
- MicroSD chip - included with the recorder
- Batteries – 2 AA’s, included with recorder (keep extras on hand)

Clipboard:

A clipboard will be used for the prompting sheet and for the experimenter to take notes while the participant is talking. The experimenter will refrain from writing on the table to minimize acoustic interruption to the recording.

Stopwatch:

A stopwatch (or timekeeping device such as a smartphone or digital watch) will be used during the recording to time the participants' talking. If using a smartphone, the experimenter must have all sending and receiving functions turned off (airplane mode) during sample collection to avoid interruption.

Recorder:

The settings for the recorder will change slightly depending on the acoustic conditions of the room where the sample is being collected. The experimenter should give enough time to calibrate the recorder settings before the participant arrives. The recommendations below are meant to serve as a guide to lessen variability of audio quality across study sites. Recorders provided by the Dalhousie (Halifax) site will already be calibrated in accordance with the instructions below. Any additional information about recorder settings that is not mentioned below can be found in the recorder manual:

https://tascam.com/content/downloads/products/558/e_dr-05_rm_va.pdf

Recorder functioning:

The automatic sample label name and recording time are displayed on the main screen. Pressing the *RECORD* button once will put the recorder on recording standby, showing the input levels for the microphones and the *REC* light at the top of the screen will flash red. The input levels can be calibrated when the recorder is in *standby* mode. The recording has not begun until the record button is pressed a second time – the time will start running and the *REC* light will stay lit.

The input bars for the left and right microphones move in accordance to the loudness of the signal input to the microphones. The input level can be controlled by using the fast forward and rewind buttons on the recorder. The input volume will need to be

recalibrated when any conditions of the recording change – the setting (room, non-human noise in the room), the distance of the recorder from the participant, etc. When properly calibrated, the input levels should fluctuate around the middle or upper middle of their range while someone is talking. If the bars reach their maximum, the *PEAK* light at the top of the screen will light up. Input levels should be maintained under the *PEAK* maximum level throughout the recording.

Recorder settings:

The home screen can be reached by using the *HOME/ POWER* button. From here, the recorder settings can be reached by using the *MENU* button. From the menu, calibrate settings in *REC SETTINGS*. The file format (*FORMAT)* should be set as WAV 24bit. The *SAMPLE* should be set to 48k. *TYPE* should be left at the default setting of *STEREO. SIZE* can be left at the default setting of *2G.* The *MIC POWER* should be *OFF.* The *LOW CUT* must be *OFF*. *PRE REC* and *AUTO TONE* should also be set to *OFF*. *TONE SECS* can be left at the default setting of *1sec.* Under *AUTO REC, MODE* should be set to *OFF,* and *LEVEL* can be left at its default of *-12dB*.

The recorder's speaker can be used to listen to samples after they have been recorded. To assess quality of recordings, it is best to listen to the samples after they have been transferred off the recorders, through headphones connected to the computer. The recorder's speaker should only be used to check basic things (did the sample actually record, is the volume too quiet/loud, etc.). The speaker can be turned on by accessing the *SPEAKER* option from the *MENU* dropdown. Turn the *SPEAKER* to *ON.*

The *MENU* dropdown also has an *OTHERS* tab. In *OTHERS,* under *MONITOR MIX,* the *INPUT* should be *OFF.* The *PB INPUT* level will not matter when the *INPUT* is *OFF*.

Under the *EFFECT* tab, *EFFECT* should be *OFF,* which will render the other options unimportant. Also under *OTHERS* the format of the file labels can be changed. As well, *OTHERS* is where the *DATE/TIME* option will be found.

It is very important to set the correct date and time, as they are saved with the file when the file is transferred to the computer.

The recommendations above are important to control as much audio and acoustic variation as possible across sites. If for any reason the settings above do not work under the site testing circumstances, experimenters should either refer to the recorder manual or contact one of the Dalhousie site speech project coordinators.

Method:

Overview:

One speech sample will be collected for each participant at three time points (Baseline (Week 0), Week 8 and Week 16). Each sample contains three different sections, and each section is prompted differently to elicit different emotions from the participant. The first is a neutral sample, in which the participant gives some background about themselves and how they spend their time. The second sample is about a positive experience, and the third sample is about a negative experience (see exact prompts below).

Each recording will be between 10 and 15 minutes long (prompts and between 3-4 minutes of participant speech per prompt). Each recording will include the audio of the introduction to the procedure, the experimenter's prompts, and all three answers from the participant.

Recorder:

The recorder will be positioned on a tripod stand with the microphones facing the participant, and the recorder display facing the experimenter. The recorder should be placed ~2 feet away from the participant.

Procedure:

The procedure begins before the participant comes into the room. The experimenter will calibrate the recorder (see "Recorder functioning" above), and then record 20 seconds of silence before the participant enters the room, under the same circumstances that the sample will be taken (same recorder settings, with any extraneous noise in the room). This sample should be part of the larger recording, and will be used to identify and filter out non-human noise in the background at the time of analysis.

The procedure should be completed in a quiet room with only the experimenter and the participant present. After the experimenter greets the participant, they may begin the recording and read the introduction. Any questions the participant has will be answered by the experimenter after the introduction, before the first prompt is given. The experimenter will give the first prompt only after ensuring the participant's understanding of the instructions. The experimenter should use a stopwatch (smartphone, digital watch, etc.) to time how long the participant has been speaking. The experimenter will start the stopwatch when the participant begins talking after the neutral prompt, and again for each of the subsequent answers.

The experimenter should not rely on the recorder for timekeeping purposes. Looking at the recorder may draw the participant's attention away from speaking, or may make them focus on the time remaining rather than on what they are saying.

The answer to each prompt will be a minimum of 3 minutes long and a maximum of 4 minutes long. If the participant stops talking before 3 minutes have passed, the experimenter will allow 15 seconds of silence before prompting for more speech with one of the appropriate prompts. If the participant stops talking after 3 minutes have elapsed, the experimenter may move on to the next prompt or cut the recording in the case of the last prompt. If the participant continues talking, the experimenter should find the next natural moment (ideally before 4 minutes have passed) to interject and close the sample (i.e. "you're all done", "ok great, thanks").

The participant must feel as though they are being listened to, so as to maintain the naturalistic traits of the speech. The recorder and recording should not be emphasized, as an uninterrupted sample is ideal for analysis. While the recording is being done, the experimenter should make minimal eye contact and provide minimal non-verbal conversation cues (i.e. smiling, nodding, etc.) to the participant, so as not to distract them or give them the impression that a back and forth discourse may take place. The experimenter may take notes on a clipboard (not on the table- minimizing excess recording noise) during the sample.

Prompting:

The prompts will be read to the participant as they are stated below.

Begin with an introduction:

"We are looking at speech as part of today's visit. To do that, I'm going to have you talk about yourself. I would like to get a sense of who you are, in your own words, so I will give you a prompt and then let you speak for 3 minutes. We will do this three times, with

three different prompts, and I would prefer not to answer any questions until you're done. Do you have any questions before we start?"

At this point, any questions the participant has can be answered by the experimenter. The less information you can give to satisfy the question, the better. If the participant is aware of the specificities of the analyses that will be carried out, they are more likely to alter their speaking behaviour to fulfill a perceived profile. For example, the participant may ask something like "What about speech are you looking at?" The experimenter should respond by saying "We are looking at what people say and the way they say it. Just speak as you would if you were telling a story to a friend/ just speak naturally."

<u>Neutral sample:</u>

"First, I would like to hear about how your last couple of weeks have been, and how you've been spending your time. Tell me how you've been feeling and what you've been up to lately."

During each sample, if the participant stops talking before three minutes have elapsed, the person gathering the sample must wait 15 seconds before using one of the two prompts below:

"Please tell me more"

"Keep going"

If these prompts are not sufficient to elicit more speech from a participant (e.g. they say "that's it", or "I don't have anything else to say"), the experimenter can use the prompt below:

"Tell me anything about yourself or what you've been up to."

Positive sample:

"Next, I want you to think about a time in the past few weeks when things went well for you. Think about when you had a positive experience or when something good may have happened to you. Take your time to think about it, and you can go ahead whenever you're ready."

If the participant stops talking, use one of the two prompts below after waiting 15 seconds:

"Please tell me more"

"Keep going"

If the above prompts do not work in eliciting more speech, or the participant asks what to talk about, use the following prompt:

"You can talk about a time when you were happy or when something good happened."

"You can talk about a time when things worked out or went your way"

In the case that a participant responds to the prompt by expressing that they haven't had a good experience in the last couple of weeks, the experimenter may use the prompt below:

"What about a time in the last couple of weeks when you were closest to feeling okay?"

Negative sample:

"Lastly, I now want you to think about a time in the past few weeks when things didn't go well for you. Think about when you had a negative experience or when something bad

may have happened to you. Take your time to think about it, and you can go ahead whenever you're ready."

If the participant stops talking, use one of the two prompts below after waiting 15 seconds:

"Please tell me more"

"Keep going"

If the above prompts do not work in eliciting more speech, or the participant asks what to talk about, use the following prompt:

"You can talk about a time when you were sad, angry, or when something bad happened."

"You can talk about a time when things didn't work out or didn't go as planned"

In any of the samples, if the participant asks how much time is left in the recording, use one of the prompts below:

"Just a couple more minutes"

"About a minute left"

The experimenter should refrain from using prompts or talking within the recording unless entirely necessary as outlined in the instructions above. However, it is important that the experimenter acknowledge what the participant said in the previous recording. Between each prompt, the experimenter can thank the participant for sharing their account, using the appropriate comment from the list below:

"That's great, thank you."

"Thank you for sharing that. That sounds like it was a great time/experience/vacation/etc…."

"Thank you for sharing that. I know it's hard to talk about those things, so thank you for opening up."

Important Notes:

1. When prompting for more speech during a sample, it is important to avoid wording the prompts as yes or no questions. For instance, "Can you tell me more?" or "Can you tell me anything else about that?" allows a participant to say "no" and finish the sample before the three minute minimum time has been reached.

2. Additional prompts can be modified conservatively as seen fit by the experimenter. The ideal sample will follow the procedure insofar that the naturalistic aspect of the speech is preserved. Modest accommodations that must be made to the script to achieve this aim are permissible.

3. Participants will always underestimate the time remaining, and overestimate their ability to fill it. For this reason, it is preferred to use the time prompts specifically to correct for this. For example, if a participant asks how much time is left, and the time remaining is 30 seconds, the experimenter can use the prompt "About a minute left".

4. The samples should always be elicited in the order they are listed above: neutral, positive, negative.

5. The experimenter should not cut the recording off in the middle of a sentence. Full and complete utterances are better for analysis.

6. It is a good idea to have tissues accessible for participants who may get emotional while providing their sample.

7. After the negative sample is collected, it is important for the experimenter to check in with the participant to assess how they are feeling after the recording has finished. The well-being of the participant must be evaluated before finishing the appointment. If they are upset after having given their negative sample, the experimenter should attempt to neutralize their mood before finishing the appointment.

Section C: Prompting Script

Introduction	"We are looking at speech behaviour as part of today's visit. To do that, I'm going to have you talk about yourself. I would like to get a sense of who you are, in your own words. I will give you a prompt and then let you speak for 3 minutes. We will do this three times, with three different prompts, and I would prefer not to answer any questions until you're done. Do you have any questions before we start?"		
Initial prompt	"First, I would like to hear about how your last couple of weeks have been, and how you've been spending your time. Tell me how you've been feeling and what you've been up to lately."	"Next, I want you to think about a time in the past few weeks when things went well for you. Think about when you had a positive experience or when something good may have happened to you. Take your time to think about it, and you can go ahead whenever you're ready."	"Lastly, I now want you to think about a time in the past few weeks when things didn't go well for you. Think about when you had a negative experience or when something bad may have happened to you. Take your time to think about it, and you can go ahead whenever you're ready."
First prompt	"Please tell me more" "Keep going"		
Second prompt	"Tell me anything about yourself or what you've been up to."	"You can talk about a time when you were happy or when something good happened." "You can talk about a time when things worked out or went your way"	"You can talk about a time when you were upset, or when something bad happened." "You can talk about a time when things didn't work out or didn't go as planned"
Time prompt	"Just a couple more minutes" "About a minute left"		
Special prompts		"What about a time in the last couple of weeks when you were closest to feeling okay?"	

Section D: Processing and Segmenting

Each segmented transcription contains coded variables for "speaker", indicating who, participant or experimenter, is speaking during the segment. Segments of experimenter speech (introduction and prompts), or segments in which the participant and experimenter both speak, are given a sentiment coding of "9". The sentiment value of "9" indicates to raters that these segments should not be coded, and they are omitted from further analysis. Only segments where the participant is speaking are coded and analyzed.

Each sample also contains "identifying" and "notes" variables. All segments that contain identifying information (names of people, places) are flagged using the "identifying variable", so that identifying content can be removed from the recordings. The "notes" variable includes rater comments regarding sound artifacts in the audio files that are not speech, including but not limited to: environmental noises, laughter, crying, sighing and sniffing.

Section E: Speech Variables

Sentiment:

The sentiment variable was created to identify the polarity of each segment in a participant speech sample. Sentiment has 3 levels: negative, neutral and positive. Sentiment ratings are based on transcribed text meaning. Measuring levels of negative sentiment can help assess depression (Breznitz, 1992; Himmelstein et al., 2018). Ratings of sentiment can be combined with emotion ratings to assess overall feeling per segment.

Richness:

Richness refers to the lexical and semantic complexity of speech. Each segment is rated as being impoverished, expected, or rich in terms of speech complexity. Impoverished speech lacks content or concreteness, and the segments may lack new information or contain reiterations or repetitions of previously uttered speech. Expected speech is characterized by phrasing that is considered conventional given the subject matter and level of arousal. Rich speech is novel or surprising, and may use humour or non-literal language such as idioms, metaphors, or sarcasm. Speech richness has been implicated as a part of depressed and psychotic illness (Cohen et al., 2014; Corcoran & Cecchi, 2020), as well as attention deficit (Gallardo-Paúls et al., 2012) (Machado-Nascimento et al., 2016).

Reference:

Each speech segment is coded according to who the speaker is referencing. There are 4 categories of reference: “0” - the speaker does not mention a relationship to any living being in the segment; “1” - the speaker references themselves within the segment; “2” - the speaker mentions themselves in relation with one or more people (or living beings);

"3" - the speaker exclusively speaks about one or more people other than themselves in the segment. Research has shown that more frequent self-reference in speech can be indicative of depression or schizophrenia (Himmelstein et al., 2018; Hong et al., 2015; Zimmermann et al., 2017).

Emotions – Anger, Fear, Sadness, Joy, Neutral:

The emotions were rated using Parrot's Emotion Classification (2001). Emotions were identified by raters using a combination of transcript content (reading) and acoustic tone (listening). Each segment was rated with only one emotion: anger, fear, sadness or joy. Those segments that did not contain any emotion were rated as neutral. Emotions that fell under Parrot's definitions of anger, fear, sadness and joy were coded as such. Speech emotion has been linked to severity and illness improvement in depression and schizophrenia (Harati et al., 2018; Minor et al., 2015; Novack, 2003).

Primary emotion	Secondary emotion	Tertiary emotions
Love	Affection	Adoration, affection, love, fondness, liking, attraction, caring, tenderness, compassion, sentimentality
	Lust	Arousal, desire, lust, passion, infatuation
	Longing	Longing
Joy	Cheerfulness	Amusement, bliss, cheerfulness, gaiety, glee, jolliness, joviality, joy, delight, enjoyment, gladness, happiness, jubilation, elation, satisfaction, ecstasy, euphoria
	Zest	Enthusiasm, zeal, zest, excitement, thrill, exhilaration
	Contentment	Contentment, pleasure
	Pride	Pride, triumph
	Optimism	Eagerness, hope, optimism
	Enthrallment	Enthrallment, rapture
	Relief	Relief
Surprise	Surprise	Amazement, surprise, astonishment
Anger	Irritation	Aggravation, irritation, agitation, annoyance, grouchiness, grumpiness
	Exasperation	Exasperation, frustration
	Rage	Anger, rage, outrage, fury, wrath, hostility, ferocity, bitterness, hate, loathing, scorn, spite, vengefulness, dislike, resentment
	Disgust	Disgust, revulsion, contempt
	Envy	Envy, jealousy
	Torment	Torment
Sadness	Suffering	Agony, suffering, hurt, anguish
	Sadness	Depression, despair, hopelessness, gloom, glumness, sadness, unhappiness, grief, sorrow, woe, misery, melancholy
	Disappointment	Dismay, disappointment, displeasure
	Shame	Guilt, shame, regret, remorse
	Neglect	Alienation, isolation, neglect, loneliness, rejection, homesickness, defeat, dejection, insecurity, embarrassment, humiliation, insult
	Sympathy	Pity, sympathy
Fear	Horror	Alarm, shock, fear, fright, horror, terror, panic, hysteria, mortification
	Nervousness	Anxiety, nervousness, tenseness, uneasiness, apprehension, worry, distress, dread

Parrott, W. (2001). *Emotions in Social Psychology*. Psychology Press: Philadelphia.

Coherence:

The measure of coherence quantifies a speaker's ability to follow a logical narrative. Breaks in coherence are characterized by speech content that deviates from the topic or train of thought; these statements may be unexpected based on the previous dialogue, and the listener may have difficulty following the speaker's narrative. Coherence is a feature that is used frequently to quantify disorganization and formal thought disorder in schizophrenia (Ayer et al., 2016; Elvevåg et al., 2007; Pauselli et al., 2018), and coherence measures have the ability to differentiate individuals with schizophrenia from controls (Elvevåg et al., 2007, 2010; Willits et al., 2018). Additionally, coherence breaks may shed light on attention deficit and distractibility, clinical features present in many disorders (Engelhardt et al., 2010; Mota et al., 2012; Raucher-Chéné et al., 2017).

Worry:

Worry manifests in the content of speech. Worry is coded in segments where "What if?" statements are present, or where the speaker repeatedly returns to a future-oriented narrative with the anticipation of a negative outcome. Higher levels of worry can be a predictor of negative mental health outcomes (Drost et al., 2014; Spinhoven et al., 2018; Struijs et al., 2018).

Anxiousness:

Unlike worry, presence of anxiousness is determined by acoustic features, or speech sounds, such as rate of speech, stop/restarts (stuttering), inappropriate laughter, or unusually high pitch. Limited literature on the manifestation of anxiety in speech points to more hesitation, more variable speech rate, and more breathy sounds (Kotsopoulos & Mellor, 1986; Toazza et al., 2016).

Criticism:

The rating of criticism was derived from the original FMSS protocol. Criticism was designed to detect negative attitudes of caregivers towards their relatives (Magaña et al., 1986). More recent research shows that higher levels of outward-directed criticism are linked to depression (Breznitz, 1992). Here, criticism is coded for any segment that includes critical statements about anything or anyone other than oneself.

Self-criticism:

The self-criticism variable is coded on all segments that contain statements containing self-critical or self-deprecating comments. Self-criticism has been implicated as a predictor of illness outcomes for various disorders, with higher levels of self-criticism linked to worse outcomes (Löw et al., 2020).

Affect (global):

The affect variable is rated based on acoustic indicators. Affect is a measure of prosodic modulation in speech, and includes pitch variability, volume and rate of speech. There are 4 levels of affect: "4" – **normal** variability in the range of expression in voice; "3" – the range of vocal expression and modulation is **restricted**, ex. range of intonation may be more restricted; "2" – the range of vocal expression and modulation is **blunted**, ex. range of intonation is minimal; "1" – vocal expression is **flat**, intonation is absent, speech is monotonous. Psychological affective states are routinely assessed as part of clinical practice. Affect is rated globally (per-sample) as a quantification for clinical impression of affect, which is generally made up of many factors present throughout a clinical interaction. Restricted, blunted, or flat affective states can be associated with both depressive and psychotic illness (Cohen et al., 2012).

Richness and Coherence (global ratings):

Global richness and coherences variables were created for the purpose of obtaining an overall impressionistic value of the sample as a whole, which may or may not correspond with the frequency of ratings of their segment-level counterparts. The aim is to create a qualification of what would be a clinician's perception of a patients' speech.