Acoustic and Linguistic Features of Impromptu Speech and their Association with Anxiety: Validation Study

Bazen Gashaw Teferra, Sophie Elizabeth Borwein, Danielle D. DeSouza, William Simpson, Ludovic Rheault, Jonathan Scott Rose

Submitted to: JMIR Mental Health
on: January 28, 2022

Table of Contents

Acoustic and Linguistic Features of Impromptu Speech and their Association with Anxiety: Validation Study

Bazen Gashaw Teferra[1] BSc, MSc; Sophie Elizabeth Borwein[2] BA, MPP, PhD; Danielle D. DeSouza[3] BSc, MSc, PhD; William Simpson[4] BSc, PhD; Ludovic Rheault[5] PhD; Jonathan Scott Rose[1] BASc, MASc, PhD

[1]University of Toronto Toronto CA
[2]Simon Fraser University Vancouver CA
[3]Winterlight Labs Toronto CA
[4]Winterlight Labs McMaster University, Department of Psychiatry and Behavioural Neurosciences Toronto CA
[5]Department of Political Science Munk School of Global Affairs and Public Policy University of Toronto Toronto CA

Corresponding Author:
Bazen Gashaw Teferra BSc, MSc
University of Toronto
The Edward S Rogers Sr Department of Electrical and Computer Engineering, University of Toronto
10 King's College Road
Toronto
CA

Abstract

Background: The measurement and monitoring of Generalized Anxiety Disorder (GAD) requires frequent interaction with psychiatrists or psychologists. Access to mental health professionals is often difficult due to high costs or insufficient availability. The ability to assess GAD passively and at frequent intervals could be a useful complement to conventional treatment and help with relapse monitoring. Prior work suggests that higher anxiety levels are associated with changes in human speech. As such, monitoring speech using personal smartphones or other wearable devices may be a means to achieve passive anxiety monitoring.

Objective: To validate the association of previously suggested acoustic and linguistic features of speech with anxiety severity.

Methods: A large number of participants (N=2,000) were recruited and participated in a single online study session. Participants completed the Generalized Anxiety Disorder-7 item scale (GAD-7) assessment and provided an impromptu speech sample in response to a modified version of the Trier Social Stress Test. Acoustic and linguistic speech features were a-priori selected based on the existing speech and anxiety literature, together with related features. Associations between speech features and anxiety levels were assessed using age and personal income included as covariates.

Results: Word count and speaking duration were negatively correlated with anxiety scores (r=-0.12; P<.001), indicating that participants with higher anxiety scores spoke less. Several acoustic features were also significantly (P<.05) associated with anxiety including the Mel Frequency Cepstral Coefficients (MFCCs), Linear Prediction Cepstral Coefficients (LPCCs), Shimmer, Fundamental Frequency, and first formant. In contrast to previous literature, the acoustic features, second and third formant, Jitter, and ZCR-zPSD were not significantly associated with anxiety. Linguistic features, including negative emotion words, were also associated with anxiety (r=0.10; P<.001). Additionally, some linguistic relationships were sex-dependent. The number of sentences produced was strongly associated with anxiety in females (r=0.12; P<.001). The use of personal pronouns was strongly associated with anxiety in males (r=0.11; P<.001).

Conclusions: Both acoustic and linguistic speech measures are associated with anxiety scores. The amount of speech, acoustic quality of speech, and gender-specific linguistic characteristics of speech may be useful as part of a system to screen for anxiety, detect relapse, or treatment monitoring.

(JMIR Preprints 28/01/2022:36828)
DOI: https://doi.org/10.2196/preprints.36828

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Acoustic and Linguistic Features of Impromptu Speech and their Association with Anxiety: Validation Study

Abstract

Background: The measurement and monitoring of Generalized Anxiety Disorder (GAD) requires frequent interaction with psychiatrists or psychologists. Access to mental health professionals is often difficult due to high costs or insufficient availability. The ability to assess GAD passively and at frequent intervals could be a useful complement to conventional treatment and help with relapse monitoring. Prior work suggests that higher anxiety levels are associated with features of human speech. As such, monitoring speech using personal smartphones or other wearable devices may be a means to achieve passive anxiety monitoring.
Objective: To validate the association of previously suggested acoustic and linguistic features of speech with anxiety severity.
Methods: A large number of participants (*N = 2,000*) were recruited and participated in a single online study session. Participants completed the Generalized Anxiety Disorder-7 item scale (GAD-7) assessment and provided an impromptu speech sample in response to a modified version of the Trier Social Stress Test. Acoustic and linguistic speech features were a-priori selected based on the existing speech and anxiety literature, together with related features. Associations between speech features and anxiety levels were assessed using age and personal income included as covariates.
Results: Word count and speaking duration were negatively correlated with anxiety scores (*r = -0.12*; *P < .001*), indicating that participants with higher anxiety scores spoke less. Several acoustic features were also significantly (*P < .05*) associated with anxiety including the Mel Frequency Cepstral Coefficients (MFCCs), Linear Prediction Cepstral Coefficient (LPCCs), Shimmer, Fundamental Frequency, and first formant. In contrast to previous literature, second and third formant, Jitter, and ZCR-zPSD acoustic features were not significantly associated with anxiety. Linguistic features, including negative emotion words, were also associated with anxiety (*r = 0.10*; *P < .001*). Additionally, some linguistic relationships were sex dependent. For example, the count of words related to power was positively associated with anxiety in females (*r=0.07*, *P = .03*) while it was negatively associated in males (*r = -0.09*, *P = .01*).
Conclusions: Both acoustic and linguistic speech measures are associated with anxiety scores. The amount of speech, acoustic quality of speech, and gender-specific linguistic characteristics of speech may be useful as part of a system to screen for anxiety, detect relapse, or treatment monitoring.

Keywords: mental health; generalized anxiety disorder; impromptu speech; acoustic features; linguistic features

Introduction

Background

Anxiety disorders are among the most common mental health issues, with an incidence of roughly 10% in the Canadian population [1]. Many Canadians are unable to access psychological and psychiatric resources to help those affected [2], in part due to the cost of professional help [3]. It may be possible to address some of this deficit using methods that

automate the measurement and diagnosis of anxiety disorders. A first step in this direction is to explore methods for the automatic detection of mental health issues that could be used to trigger early intervention, monitor treatment response, or to detect relapse. In addition, frequent monitoring together with other time-series information could be used to help understand the mechanisms of Generalized Anxiety Disorder (GAD) itself. One avenue of such automation is to record a person's speech, and to look for signals of anxiety within the recordings.

In this work, we focus specifically on Generalized Anxiety Disorder (GAD) [4]. One reason that GAD may be detectable in speech is that those with anxiety disorder exhibit higher activation of the sympathetic nervous system under stress compared to those without anxiety [5], which in turn influences the production of speech [6]. The goal of this work is to collect a large set of samples of audio speech, each with a self-reported measure of anxiety scale, and to explore if acoustic and linguistic signals correlate with measured anxiety. We build on previous studies by collecting roughly ten times greater number of human subjects than previous research on the detection of anxiety in speech. Many of the signals that we explore have been previously reported as significantly correlated with anxiety in the literature, and our goal is to leverage our larger sample size to examine which signals could be most useful in identifying anxiety in speech. We also explore linguistic indicators of anxiety that have not been considered before.

This paper is organized as follows: the next section summarizes related work in anxiety detection. The Methods section describes the speech sample collection methods and the set of features considered for correlation with anxiety. The Results section reports on the demographics of participants and feature correlations, while the Discussion section discusses the results and their implications for future research on anxiety detection. A final section concludes.

Related Work

While it is important to note that some scholarship is skeptical that biomarkers correlate with emotions [7], here we review existing work exploring associations between both acoustic and linguistic speech features and anxiety severity in healthy and clinical cohorts. Note that these studies explore broader classes of anxiety disorders including internalizing disorders, Social Phobia/Social Anxiety Disorders, Panic Disorder, Agoraphobia as well as Generalized Anxiety Disorder.

McGinnis et al. [8] identify several acoustic characteristics of speech that can be used to detect anxiety disorders in children. Studying 71 participants between the ages of three and eight, the researchers were able to detect internalizing disorders - a collective term for anxiety and depression - from speech. The authors extracted and selected several acoustic features from the speech produced in a three-minute task based on the Trier Social Stress Test for children (TSST-C) [9]. These features included zero-crossing rate (ZCR), Mel Frequency Cepstral Coefficients (MFCC) [10], zero crossing rate of the z-score of the power spectral density (ZCR-zPSD), dominant frequency, mean frequency, perceptual spectral centroid, spectral flatness, and the skew and kurtosis of the power spectral density. Using the Davies-Bouldin index based feature selection [11], the MFCC features and ZCR of the zPSD had the highest Davies-Bouldin score. Several models were built to predict which children had an internalizing disorder (n=43 out of 71) or were healthy. Both a logistic regression and an SVM [12] achieved a classification accuracy of 80%.

Özseven et al. [13] conducted a study of the speech of 43 adults between the ages of 17 and 55. Of these 43 adults, 21 were clinically diagnosed with GAD, two were diagnosed with Panic Disorder, and 20 were healthy controls. The study explored 122 acoustic features derived from the participants' speech to determine the correlation between these features and anxiety. Their results showed that 42 of the features (including MFCCs, LPCC, F0, F1,F2,F3 jitter, and shimmer) showed a significant change between a neutral state and an anxious state in the anxious participants.

Weeks et al. [14] found a relationship between anxiety and alterations in voice. Specifically, their study shows a link between vocal pitch (characterized by the fundamental frequency) and social anxiety disorder (SAD). They collected impromptu speech samples from 46 undergraduate students, 25 with a diagnosis of SAD and 21 healthy controls. Participants also completed the Beck Anxiety Scale as measure of self-reported anxiety severity [15]. Their results indicated that mean fundamental frequency was positively correlated ($r = 0.72$, $P = .002$) with anxiety severity across all male participants. However, the correlation for female participants was weaker ($r = 0.02$, $P = .92$), indicating possible sex differences in the relationship between anxiety severity and vocal pitch.

Laukka et al. [16] explore the relationship between anxiety and the acoustic features of speech. They collected speech data from 71 patients with social phobia delivering public speeches and extracted four types of speech features: pitch (F0 mean, F0 std, F0 max), loudness (intensity mean), voice quality (HF 500, relative proportion of spectral energy above vs below 500), and temporal aspects of speech (articulation rate and percentage of silence). The researchers observed a significant change from pre-treatment to post-treatment (a pharmacological anxiolytic treatment for social anxiety) in F0 mean, F0 max, HF500 and percentage of silence. They also calculated Pearson's correlation coefficient between state anxiety measured by STAI-S [17] and the speech features. Those with a significant correlation were standard deviation of F0 ($r = -0.24$, $P < .05$) and percentage of silence ($r = 0.36$, $P < .01$).

Albuquerque et al. [18] investigated the relationship between acoustic speech features and anxiety. They recruited 112 adult Portuguese speakers who performed two tasks: reading vowels in disyllabic words and picture description. The authors extracted 18 acoustic features including F0, F1, F2, speech duration, number of pauses, and articulation rate. They measured the percentage change between non-anxious (HADS-A [19] score ≤ 7) and anxious (HADS-A >7) participants and observed a change of more than 10% in speech duration.

Wörtwein et al. [20] assessed the behaviours of participants experiencing anxiety caused by public speaking through audiovisual features. A total of 45 participants were recruited from Craigslist. These participants were asked to complete the Personal Report of Confidence as a Speaker (PRCS) scale [21], which estimates public speaking anxiety levels. Several audio features were extracted from the audio and their results show significant relationships between PRCS and Standard deviation of MFCC0 [22] ($r = -0.36$, $P < .05$), Standard deviation of the first formant ($r = -0.41$, $P < .01$), and the total pause duration ($r = 0.35$, $P < .05$).

Hagenaars et al. [23] explores if the activation of fear is manifest in the speech of 25 female patients diagnosed with panic disorder. Their results show that patients with panic disorder have significantly higher pitch ($P < .001$) during autobiographical fear memory. Respondents also spoke significantly slower ($P < .001$) during autobiographical talking compared to a script

talking.

Di Matteo et al. [24] explored the relationship between linguistic features of speech and anxiety. Their work used *passively* collected intermittent samples of audio data from participants' smartphones, collected over a 2-week period, as input. The study had 84 non-clinical participants recruited from an online recruitment platform. The audio was converted to text, and the authors used the Linguistic Inquiry and word Count (LIWC) approach [25] to classify the words into 67 different categories. They calculated correlations with four self-report measures: social anxiety disorder, GAD, depression, and functional impairment. They observed a significant correlation between words related to perceptual process ('See' in the LIWC) with social anxiety disorder ($r = 0.31$, $P = .003$) and words related to rewards with generalized anxiety disorder ($r = -0.29$, $P = .007$).

In a similar study that used LIWC features, Anderson et al. [26] recruited 42 participants diagnosed with SAD and 27 healthy controls to explore the differences in the words used between these two groups. The participants were asked to write a distinct autobiographical and socially painful passage. They used LIWC to extract word count in each of the LIWC categories, such as first person singular, anxiety related words and fear related words. Their results showed that SAD patients used more first-person singular pronouns (I, me, mine), anxiety-related words, sensory/perceptual words, words denoting physical touch, and fewer references to other people.

Overall, previous work identifies several audio features that are correlated with anxiety. However, the results are mixed due to differences in participants recruited, speech measures assessed, statistical methods employed, and amount of mood induction. Additionally, the largest sample size among these studies was 112, which limits the potential for generalizability to the larger population; a necessary step before considering the deployment of technologies for passive anxiety monitoring. In this study, we recruit a substantially larger cohort ($N = 2,000$) to explore features of speech from previous findings at a greater scale.

Methods

Data collection

Participants from a nonclinical population were recruited for a 10-to-15-minute task implemented through a custom website. Self-report measures of anxiety were collected once at the beginning of the study and at the end of each of two specific tasks. In the sub-sections below, we describe the recruitment of participants, the data collection procedure, and the assessment of anxiety and speech measures.

The study was approved by the University of Toronto Research Ethics Board under protocol #37584.

Recruitment and Demographics

A total of 2,000 participants were recruited using the Prolific [27] online human subject recruitment platform. Prolific maintains a list of registered participants and, for each participant, many characteristics including age, income, sex, primary language spoken, country of birth, and residence. The inclusion criteria for this study were: age range 18–65 years,

fluency in English, English as a first language, at least ten previous studies completed on Prolific with 95% of these previous Prolific tasks completed satisfactorily, as labelled by the study author. The dataset was also balanced for sex (50% Female, 50% Male). The Prolific platform provides us with some relevant demographics of the participants, including their age and income.

Participants who completed the study were paid £2 (approximately $3.41 CAD). They were able to complete the entire study remotely, using their personal computers.

Study Procedure

Subjects were presented with the opportunity to participate in this study on Prolific if they met the inclusion criteria given above. Those who wished to participate clicked on the study link, which brought them to a consent form that described the procedure and goals of the study and provided information on data privacy. After they gave consent, a hyperlink brought participants to an external web application (a screenshot of which is shown in Multimedia Appendix 1) that implemented the tasks described below.

Participants were first asked to fill out the standard GAD-7 questionnaire [28] described in more detail in the Anxiety Measures section. Then, they were asked to complete two speech tasks, which were recorded using their computer's internal microphone. Note that our protocol also recorded a video of the participants' faces during both speech tasks. Although that video is not used in the work reported here, the fact that the video was requested may have influenced the set of participants willing to continue participation, as discussed later in this paper.

For the first speech task (Task 1) participants were asked to read aloud a specific passage called "My Grandfather," which is a public domain passage that contains nearly all the phonemes of American English [29]. The full script of this passage appears in Multimedia Appendix 2. This passage is not intended to induce stress or anxiety, but to provide a baseline speech sample for each participant. It was used in this work to test the quality of the speech-to-text transcription.

For the second speech task (Task 2), the participant followed a modified version of the widely-used Trier Social Stress Test (TSST) [30], for the purpose of inducing a moderate amount of stress. We have chosen to base our anxiety stimulus on the TSST because previous studies ([31,32]) have shown a higher activation in participants with relatively higher anxiety after exposure to moderate stress induced by the TSST.

In this modified version of the TSST, participants were told to imagine that they were a job applicant for a job that they really want (their 'dream' job), and they were invited for interview with a hiring manager. They were given a few minutes to prepare – to decide what their 'dream' job is – and how they would convince an interviewer that they are the right person for the position. Participants were also told that the recorded video will be viewed by researchers studying their behaviour and language. Participants were then asked to speak for five minutes, making the case for themselves to be hired for that dream job.

Note that, in the original TSST [30], participants would normally deliver their speech in front of a live panel of judges. If a participant finished their delivery in less than five minutes, the judges in the original TSST design would encourage the participant to keep speaking for the full five

minutes. An example statement of encouragement is: “What are your personal strengths?” In our modified TSST, we implemented a similar method to encourage participants to speak for the full five minutes: When our software detects silence (the absence of speech for more than 6 seconds), it will display several different prompts, which are reproduced in Multimedia Appendix 3, inviting participants to keep speaking on different topics relating to the task. Finally, note that the modified TSST only made use of the first part of the original TSST, and not the second task involving mental arithmetic.

Anxiety Measures

Our goal is to examine possible correlations between features of speech and Generalized Anxiety Disorder (GAD), based largely on previously suggested features. To measure the severity of GAD, we used the GAD-7 [28] scale, which is a seven-item questionnaire that asks participants how often they were bothered by anxiety-related problems during the previous two weeks. While the two week time period suggests that the GAD-7 measures a temporary condition, this seems in contradiction with the fact that a GAD diagnosis requires six months duration of symptoms [33,34]. However, the GAD-7 has been validated as a diagnostic tool for GAD (using a value of 10 as the cut-off threshold) with a sensitivity of 89% and a specificity of 82% [28]. Thus, we choose to use the GAD-7 to obtain a binary label of GAD (using the same threshold of 10) as our main indicator of anxiety.

Each of the seven questions on the GAD-7 has four options for the participant to select from, indicating how often they have been bothered by the seven problems in the scale. These options and their numerical ratings are: 0-Not at all, 1-Several days, 2-More than half the days, and 3-Nearly every day. The final GAD-7 score is a summation of the values for each question, giving a severity measure for GAD in the range from 0 (no anxiety symptoms) to 21 (severe anxiety symptoms).

We also employ a second, informal anxiety measure in this study to serve as an internal check to measure how much, on average, the modified TSST (Task 2) has induced stress and anxiety compared to Task 1 (the reading/speaking of the Grandfather passage). Here we use a single question to measure self-reported levels of anxiety, on a four-point scale. We ask participants how anxious they felt during the task, and to choose from the following numerical rating: 0-Not anxious at all, 1-Somewhat anxious, 2-Very anxious, and 3-Extremely anxious. This question is deployed immediately after the first and second tasks.

Selection of Acoustic and Linguistic Features

Prior work suggested that information about the mental state of a person may be acquired from the signals within speech acoustics [35] as well as the language used [36]. We refer to each kind of this extracted information as a *feature* using the terminology employed in the field of machine learning.

In this work, we consider both acoustic and linguistic features, which are described in the following sections. These features were extracted from each of the 5-minute speech samples in which the subject responded to the modified TSST task. Note that all the participants are prompted to speak for the full five minutes, as described in the Study Procedure section, although the total speech duration of each participant may vary.

Acoustic Features

Previous research has identified several acoustic features that are correlated with anxiety, as described in Section Related Work. Using these previous findings as a reference point, we select the following acoustic features for our empirical analysis:

- **Mel Frequency Cepstral Coefficients (MFCC):** Coefficients derived from a mel-scale cepstral representation of an audio signal. We include 13 MFCCs, a common set of acoustic signals that are designed to reflect changes in perceivable pitch. The MFCC features were shown to be related to anxiety in [8,13,20]. Descriptive statistics (mean and standard deviation) of the 13 MFCC features were used in the current study. Note that not all MFCC features included in the current study were determined to be significant in prior work; however, these 13 are most commonly assessed together so we included them all as features of interest. The parameters we used when extracting these 13 MFCC features are: window length = 2048 samples; length of FFT window = 2048 samples; samples advance between successive frames = 512 samples; Window type = Hanning; Number of Mel bands = 128.

- **Linear Prediction Cepstral Coefficients (LPCC):** Coefficients derived from a linear prediction cepstral representation of an audio signal. The first 13 cepstrum coefficients are used here. The LPCC features were shown to be related with anxiety in [13]. Descriptive statistics (mean and standard deviation) of the 13 LPCC were used in the current study.

- **ZCR zPSD:** The Zero Crossing Rate (ZCR) of the z-score of the Power Spectral Density (PSD). In [8], ZCR-zPSD was one of the top features selected using Davies-Bouldin index based feature selection [11] for an anxiety prediction task.

- **Amount of Speech:** The amount of speech and related metrics such as the percentage of silence. These features have been shown to be related with anxiety in [16,18,20]. Our specific feature will be the amount of time, in seconds, that speech was present. We will also count the total number of words present in a speech-to-text transcript, as a separate measure of amount of speech.

- **Articulation rate:** Indicates how fast the participant spoke. Work by [23] suggests that patients with panic disorder spoke significantly slower ($P < .001$) during autobiographical talking as compared to reading a script.

- **Fundamental frequency (F0):** is the frequency at which the glottis vibrates or, also known as *pitch* of the voice. Multiple studies have shown F0 to be one of the acoustic features that are affected by anxiety [13,14,16,23]. The fundamental frequency varies throughout a person's speech, so both the mean and standard deviation of F0 are used as features.

- **F1, F2, F3:** The first, second and third formants [37]. The work in [13] shows a significant relation with anxiety. The mean and standard deviation of each format were used as features.

- **Jitter:** The cycle-to-cycle F0 variation of the sound wave. *Jitter* have been shown to be an

indicator of anxiety [13,38,39].

- **Shimmer:** The cycle-to-cycle amplitude variation of the sound wave. *Shimmer* has been shown to be related with anxiety severity [13].

- **Intensity:** The mean squared of the amplitude of the sound wave within a given frame, also known as *intensity* has been shown to be related with anxiety [16]. Since the amplitude of a sound wave varies during speech, the mean and standard deviation were used as features.

The features listed above were extracted using the following software packages: My Voice Analysis [40], Surfboard [41], and Librosa [42].

Linguistic Features

Using Amazon's AWS speech-to-text [43], a transcript was produced from the audio recordings. From the transcripts, linguistic features were extracted using the Linguistic Inquiry and Word Count (LIWC) software [25], which places words into dictionaries based on semantic categories. For example, one category is called 'negemo' and contains words that relate to negative emotions, such as hurt, ugly and nasty. Another category is called 'health' and contains words such as clinic, flu and pill. There is also a category called 'anxiety' which includes words such as anxiety and fearful. Some categories are contained within others - for example anxiety is contained within negemo.

To apply the LIWC dictionaries, one simply counts the number of words that belong to each category, and each count becomes a feature. There are a total of 93 categories in the LIWC, but not all are relevant for a speech-to-text transcript. We have removed those features that are not relevant - for example informal language words such as 'lol' and 'btw.' Other excluded categories include those relating to some punctuation (e.g., colons, quotation marks, parentheses). Removing these, a total of 80 linguistic features remained. Prior work [24,26] that was discussed in Section Related Work has shown that LIWC categories related to perceptual processes (see, hear, feel), words related to reward, the use of first-person singular pronoun, and anxiety related words were associated with anxiety.

Separation of Data for Analysis

The overarching objective of this study is to gain an understanding of which features of speech – both acoustic and linguistic – are correlated with the GAD-7 scale. It is known, however, that certain demographic attributes are directly indicative of anxiety. For example, sex is known to influence prevalence of anxiety [44]. Also, both age [45] and income [46] influence anxiety which suggests the need to control for these demographics. An additional reason to control for the demographics is that both age and income have been shown to be related to speech features [47][48]. Due to the strong effect of sex on GAD-7 score, we create a separate dataset for analysis of female and male samples, in addition to the combined dataset. We chose to do this, rather than correcting for sex computationally, as it leaves the data intact.

Statistical Analysis

The partial Pearson's correlation coefficients [49] was computed between each of the features

and the GAD-7 (controlling for the effect of age and personal income). Correlations were examined for three versions of the dataset: the entire sample data set, and separately by sex for male and female participants. We have considered a result statistically significant at a *P* value significance level of *.05*. The *P* values were not corrected to account for the large number of tests, since we are attempting to use features that were determined to be significant in previous works.

Results

This section reports the main empirical results. We begin by discussing the recruitment yield, the demographic characteristics of the participants, and the relationship between demographic attributes and the reported GAD-7 scale. Next, we report correlations for the features described in Section Selection of Acoustic and Linguistic Features.

Recruitment and Data Inclusion

A total of 4,542 participants accepted the offer from the Prolific recruitment platform to participate in the study. From those, 2,212 participants finished the study, giving a recruitment yield of 49%.

From the 2,212 participants who completed the study, 2,000 provided acceptable submissions (and thus received payment), giving a submission-to-approval yield of 90%. To be clear, the recruitment continued until 2,000 acceptable submissions were received. The reasons that submissions were deemed unacceptable include: a missing video, a missing or grossly imperfect audio, or a failure to complete one or both tasks. These acceptability criteria are distinct from those used in the subsequent review of audio quality that is described below. The period of the recruitment ranged from November 23, 2020 to May 28, 2021. We note that recruitment took place during the global COVID-19 pandemic.

In addition to the above submission approval criterion, we reviewed the input data and audio for acceptability using the following procedure. To begin, we computed all acoustic and linguistic features described in Section Selection of Acoustic and Linguistic Features. Recordings with poor quality were filtered out for manual review based on the following criteria:

1. A Task 2 word count lower than 125;
2. A speaking duration for Task 2 lower than 60 seconds (as compared to the full five minutes);
3. *Any* other feature value being beyond three standard deviations from the mean, in either direction (outliers).

A total of 193 participant recordings were flagged based on these criteria. For each of these, a researcher listened to their Task 2 audio recordings. The researcher discarded any samples that were deemed, subjectively, to be of insufficient audio quality, or those whose response to Task 2 was not responsive to the task itself. A total of 123 out of the 193 flagged participants were rejected through this manual review, leaving 1,877 samples.

Finally, the samples were checked for missing data, with 133 participants having missing

demographic info. Consequently, the final number of participants included in our analysis is 1,744. The flow chart of the study recruitment and quality control is given in Figure 1. We have also explored correlations of the excluded data with the GAD-7, often called missingness analysis, and is presented in Multimedia Appendix 4.

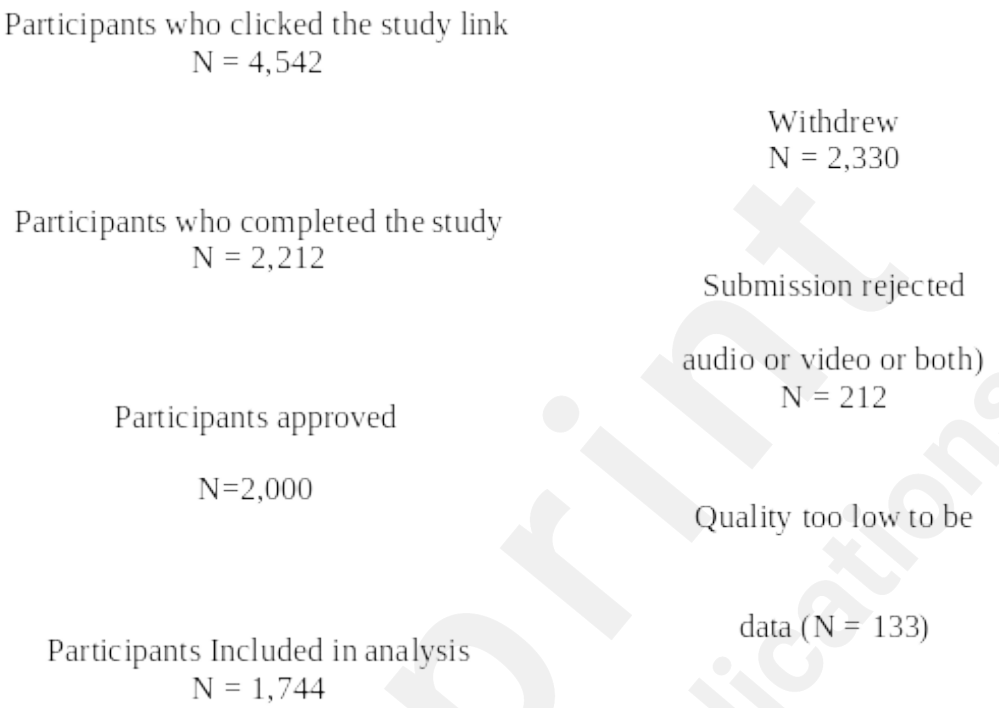


Figure 1: Study recruitment flow chart

Data Overview and Demographics of Participants

Table 1 shows how many of the 1,744 participants are above and below the GAD-7 screening threshold of 10. In the foregoing, we will refer to those participants who have a GAD-7 score ≥ 10 as the *anxious* group, and those with a GAD-7 score < 10 as the *non-anxious* group.

Table 1: GAD Classification 1,744 Accepted Participants

GAD-7 Category	
Above threshold (≥10) Anxious Group	540 (31%)
Below threshold (<10) Non-Anxious Group	1204 (69%)

Table 2 shows participants' demographics, obtained from the Prolific recruitment platform. Columns 1 and 2 of the table shows the name of demographic attributes and each category, while columns 3 and 4 give the number (and percentage) of participants with that attribute in the anxious and non-anxious groups, respectively. Column 5 gives the *P*-value for a chi-square test of the null of independence, to determine if there is a significant difference between the anxious and non-anxious groups, for each categorical factor.

Table 2:Demographic of participants for anxious and non-anxious groups

Demographic Factors		Anxious (N=540)	Non-anxious (N=1204)	*P* value
Sex	M	229 (26.0%)	653 (74.0%)	<.001 (χ^2)
	F	311 (36.1%)	551 (63.9%)	
Self-reported ongoing mental health illness/condition	Yes	297 (48.8%)	311 (51.2%)	<.001 (χ^2)
	No	243 (21.4%)	893 (78.6%)	
Personal Income (GBP)	Less than £10,000	181 (39.2%)	281 (60.8%)	<.001 (χ^2)
	£10,000 - £19,999	112 (35.0%)	208 (65.0%)	
	£20,000 - £29,999	92 (26.2%)	259 (73.8%)	
	£30,000 - £39,999	60 (24.6%)	184 (75.4%)	
	£40,000 - £49,999	36 (24.8%)	109 (75.2%)	
	£50,000 - £59,999	20 (21.3%)	74 (78.7%)	
	≥ £60,000	39 (30.5%)	89 (69.5%)	
Age	18-19	27 (38.0%)	44 (62.0%)	<.001 (χ^2)
	20-29	239 (38.7%)	379 (61.3%)	
	30-39	162 (32.7%)	334 (67.3%)	
	40-49	67 (23.4%)	219 (76.6%)	
	50-59	39 (22.8%)	132 (77.2%)	
	≥ 60	6 (5.9%)	96 (94.1%)	

Post-Task Self-Report Anxiety Measure

As described in Section Anxiety Measures, participants were asked to rate their state anxiety after each task on a scale from 0 to 3, where 3 is the highest level of anxiety. Table 3 gives the average value of this informal rating after Task 1 and Task 2. We report a paired t-test to assess the difference between the two measurements. The test validates that the modified TSST task successfully induced some anxiety in participants, with the average score on the self-reported state anxiety measure increasing from 0.5 to 1.6 ($P < .001$), before and after completing Task 2.

Table 3: Post-task Self-Report Anxiety Measure and Paired t-test

	Task 1	Task 2	Paired t-test *P*-value
Average Post-Task Self-Reported Anxiety measure (SD)	0.5 (0.6)	1.5 (0.9)	< .001

Feature Correlations

Section Selection of Acoustic and Linguistic Features, describes the set of acoustic and linguistic features that were selected. These were features that were reported as significant in prior work on anxiety and speech, as well as closely associated features. These features were computed on the speech samples of participants performing Task 2, the modified TSST. The subsections below summarize the main empirical results. Correlation between demographics and the acoustic/linguistic features is presented in Multimedia Appendix 5 and inter-correlation between the significant features is presented in Multimedia Appendix 6, 7, and 8 for the all-sample, female sample and male samples, respectively.

Amount of speech

The features with one of the highest correlations, for both the male and female datasets, are those related to how much the participant spoke during Task 2. Two specific features used to estimate speech length are speaking duration (the number of seconds of speech present within the five-minutes speech task) and the word count derived from a speech-to-text transcript. Table 4 gives the correlation for the all-sample dataset (controlling for sex, age, and income) and for separated female and male datasets (controlling for age and income). Figure 2 presents a scatter plot of speaking duration vs. GAD-7 as well as the distribution of both variables, for all three datasets. The scatter plot is coloured to give a better sense of the density of data points. Figure 3 provides the same kind of scatter plots/distributions for the word count metric of Task 2.

Table 4: Correlation of Amount of Speech features with the GAD-7

All Samples (N=1877)		
Feature	**r**	***P***
Speaking Duration	-0.12	<.001
Word Count	-0.12	<.001

Female Samples (N=935)		
Feature	**r**	***P***
Word Count	-0.13	<.001
Speaking Duration	-0.11	<.001

Male Samples (N=942)		
Feature	**r**	***P***
Speaking Duration	-0.13	<.001
Word Count	-0.12	<.001

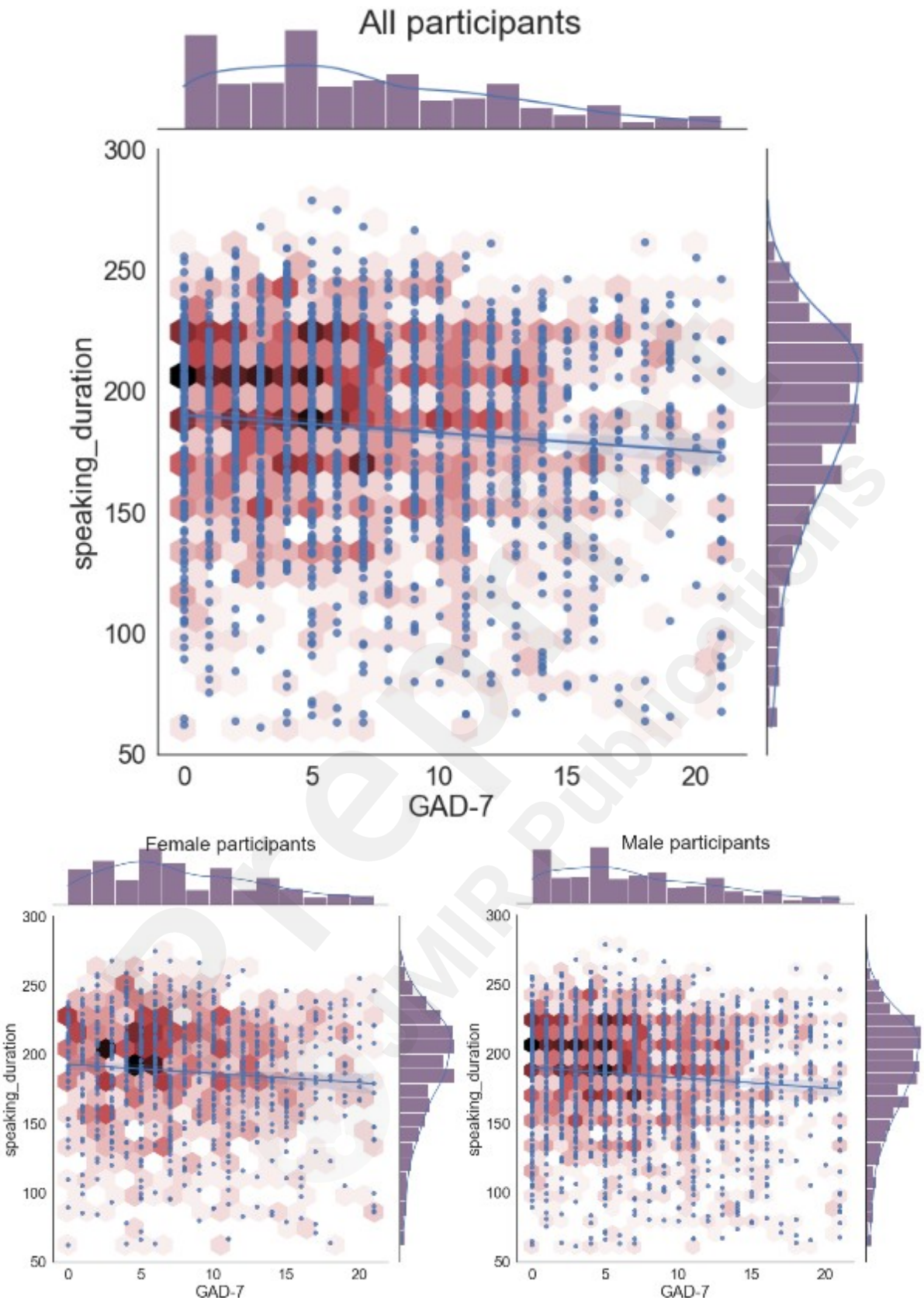


Figure 2: Speaking Duration vs. GAD-7 Scatter plot and distributions

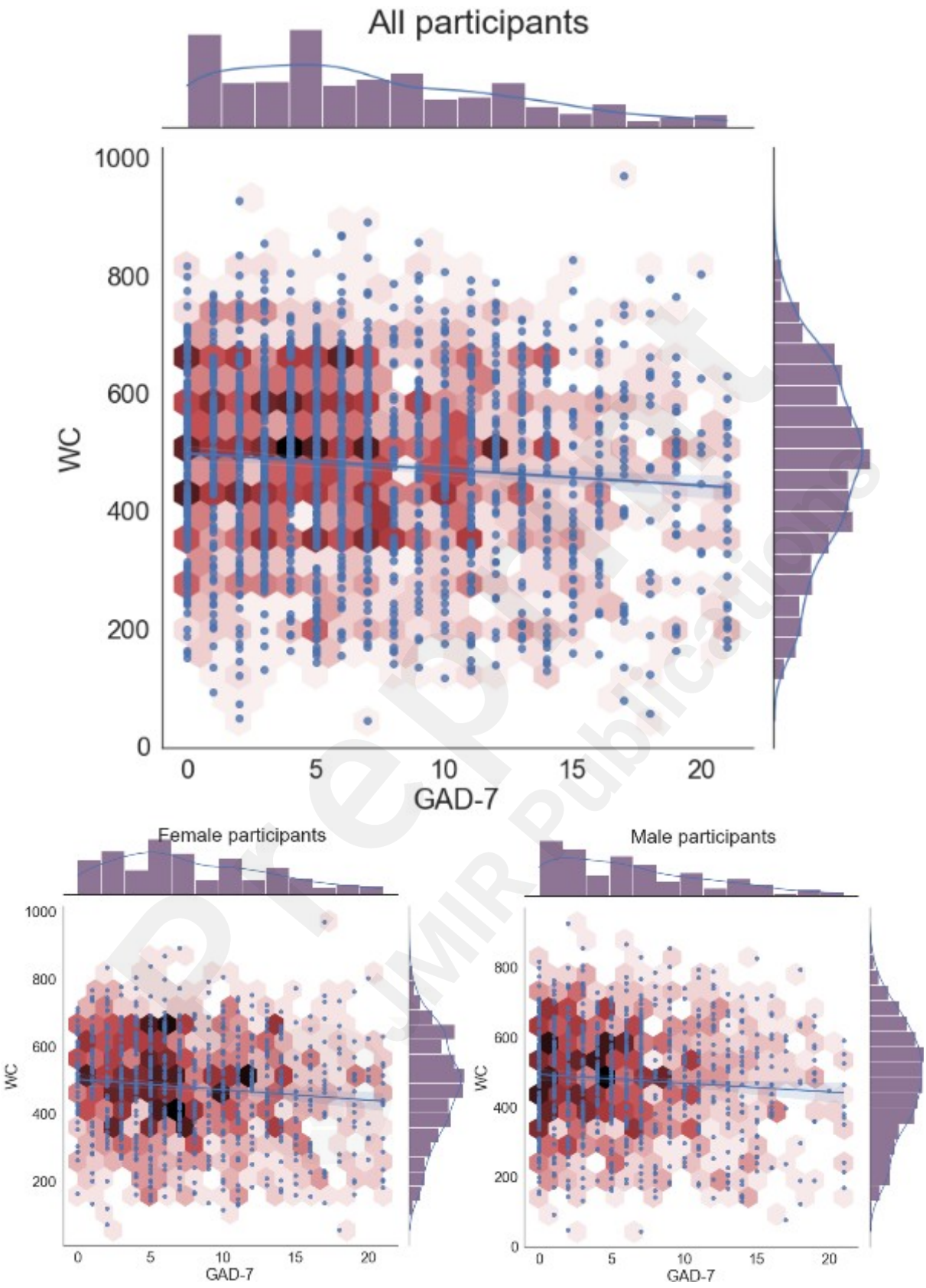


Figure 3: Word Count vs. GAD-7 Scatter plot and distributions

Acoustic Feature Correlation with GAD-7

Table 5 present the correlation and *P*-values for all of the acoustic features (presented in section Acoustic Features) that had *P*-values above the 95% confidence level for the three

datasets – all participants, and then female-only and male-only participants. Again, note that all correlations were computed after controlling for age and personal income, while the calculations involving all participants also controlled for sex.
Table 6 reports results for features that previous work found to be statistically significant, but for which we find no correlation in our sample. In our results, these features were not significantly associated with anxiety in any of the three datasets: all, female, and male.

Table 7 makes a direct comparison between previous work on the specific features (and their relation to anxiety) and results from the present study.

Table 5: Correlation of Significant Acoustic features with GAD-7

All Samples (N=1744)		
Feature	**r**	***P***
Shimmer	0.08	<.001
mfcc_std_2	-0.08	.002
mfcc_std_3	-0.07	.002
mfcc_mean_2	-0.07	.004
f0_std	0.06	.01
mfcc_std_5	-0.06	.01
mfcc_std_4	-0.05	.03

Female samples (N=862)		
Feature	**r**	***P***
mfcc_std_3	-0.10	.002
Shimmer	0.10	.004
lpcc_std_6	-0.09	.008
lpcc_std_4	-0.09	.008
mfcc_mean_2	-0.09	.01
Intensity_mean	-0.09	.01
mfcc_mean_1	-0.09	.01
lpcc_std_10	-0.07	.03
intensity_std	-0.07	.03
lpcc_std_12	-0.07	.04
mfcc_mean_8	0.07	.04
lpcc_mean_4	0.07	.049

Male samples (N=882)		
Feature	**r**	***P***
mfcc_std_2	-0.09	.005
mfcc_std_5	-0.09	.01
mfcc_mean_5	-0.08	.01
f0_std	0.07	.03
mfcc_std_4	-0.07	.04
Shimmer	0.07	.04
mfcc_std_11	-0.07	.046
f1_mean	0.07	.047

Table 6: Correlation of Acoustic Features Not Found Significant

Feature	**Previous works**	**Current study**					
		All samples		**Female samples**		**Male samples**	
		r	***P***	**r**	***P***	**r**	***P***
Jitter	Showed a significant increase from a neutral state to anxious state [13]	0.03	.18	-0.01	.76	0.06	.06
ZCR_zPSD	ZCR_zPSD was one of the top selected features using Davies-Bouldin Index based feature selection [8]	0.01	.67	-0.04	.29	0.05	.14
Articulation rate	Patients with panic disorder spoke significantly slower (P < .001) during autobiographical talking compared to a script talking [23]	-0.01	.64	-0.05	.12	0.02	.55
F1 std	Showed a significant change between a neutral state and anxious state [13]	-0.03	.18	-0.02	.53	-0.04	.25
F2 mean		0.004	.85	0.04	.26	-0.04	.22
F2 std		0.01	.59	0.03	.38	-0.02	.60
F3 mean		0.02	.49	0.04	.21	-0.01	.72

Table 7: Comparison of Previous Work Correlations with Present Study

Feature	**Previous work**		**Current study**					
			All samples		**Female samples**		**Male samples**	
	r	**p**	**r**	***P***	**r**	***P***	**r**	***P***
Speaking duration	-0.36	<.01	-0.12	<.001	-0.11	<.001	-0.13	<.001
MFCC_std_1	-0.36	<.05	0.01	.54	0.02	.61	0.02	.52
F0_mean	Female:0.02;	Female: 0.92;	0.02	.37	-0.03	.33	0.06	.06

	Male:0.72	Male:0.002						
F0_std	-0.24	<.05	0.06	.01	0.03	.30	0.07	.03
Intensity mean	-0.2	-	-0.04	.13	-0.09	.01	0.01	.72

Linguistic Feature Correlation with GAD-7

The quality of the transcript produced using Amazon's AWS speech-to-text (STT) [43] was analysed by comparing the transcript produced from Task 1 audio to the actual Grandfather Passage. The Word Error Rate (WER) was calculated and the STT transcript had an average WER of 7%.

Table 8 present the set of linguistic features (described in Section Linguistic Features) that had *P*-values lower than 0.05 for the same three datasets—all participants, and then male-only and female-only. Each table is sorted in decreasing order of absolute value of correlation. As before, the partial correlations account for age and personal income across all datasets, and we also control for sex in the full dataset.

Table 8: Correlation of Significant LIWC Linguistic features with the GAD-7

All Samples (N=1744)		
Feature	**r**	***P***
AllPunc	0.13	<.001
Period	0.12	<.001
assent	0.10	<.001
negemo	0.10	<.001
relativ	-0.09	<.001
motion	-0.08	<.001
swear	0.08	<.001
anger	0.08	<.001
focusfuture	-0.07	.003
adverb	-0.07	.004
time	-0.07	.004
function	-0.07	.005
negate	0.07	.006
prep	-0.06	.007
WPS	-0.06	.007
anx	0.06	.008
hear	0.06	0.01
death	0.06	0.01
ipron	-0.06	0.01
see	-0.06	0.01
affect	0.06	0.02
i	0.05	0.02
family	0.05	0.02
sad	0.05	0.03
ppron	0.05	0.03
space	-0.05	0.04
article	-0.05	0.04
leisure	0.05	0.04
friend	0.05	0.047

Female Samples (N=862)		
Feature	**r**	***P***
Period	0.16	<.001
AllPunc	0.14	<.001
adverb	-0.11	<.001
negemo	0.11	<.001
anger	0.11	.002
motion	-0.10	.003
assent	0.10	.004
see	-0.09	.006
relativ	-0.09	.006
sad	0.08	.01
Dic	-0.08	.02
power	0.07	.03
WPS	-0.07	.03
death	0.07	.04
percept	-0.07	.046

Male Samples (N=882)		
Feature	**r**	***P***
AllPunc	0.13	<.001
assent	0.11	.001
relativ	-0.10	.002
leisure	0.10	.002
hear	0.10	.003
swear	0.10	.004
time	-0.10	.004
Apostro	0.09	.005
power	-0.09	.01
ppron	0.09	.01
Sixltr	-0.09	.01
anx	0.08	.01
negate	0.08	.01
negemo	0.08	.01
article	-0.08	.01
Period	0.08	.02
prep	-0.08	0.02
focusfuture	-0.08	0.02
family	0.08	0.02
ipron	-0.07	0.04
affect	0.07	0.04
motion	-0.07	0.048

Discussion

Our central objective is to test specific acoustic and linguistic features of impromptu speech for their association with anxiety and to do so with a larger number of participants. In this section, we discuss the implications of the findings presented in the previous section, as well as the limitations of the study.

Principal Findings

The results presented in the Results section quantified the relationship between features computed from recorded speech and the self-reported GAD-7 scale, using Pearson correlation coefficients, controlling for age and income. Results show several significant correlations between features extracted from speech and anxiety, which can help to inform future efforts in the automatic monitoring of anxiety. We discuss these below.

Recruitment and Data Inclusion

Figure 1, the study recruitment flow chart, shows that the recruitment yield was 49%. For the 2,330 participants who dropped out after accepting the study, we can only speculate as to why. Some may have been unwilling to have their words audio recorded or their full video recorded, and even though the consent form makes this task clear, it may be that the participants who dropped out only really understood this when seeing their video on the screen.

We had also conducted a missingness analysis on the 256 samples that were excluded from the study (the data for which is presented in Multimedia Appendix 4). The results show that, in the excluded data, the mention of words related to anxiety and words related to home had a significant positive correlation with anxiety and the count of longer words (>six letters) was negatively correlated with anxiety. We found similar positive and negative correlations of these features in the 1744 samples included in our analysis. This indicates that excluding the 256 samples didn't affect the correlation results.

Demographics of Participants

Table 1 shows that the proportion of participants in the Anxious group (those above the GAD-7 screening threshold of 10) is 31%, which is much higher than the general population rate of roughly 10% [1]. This result, that English speakers recruited from Prolific have elevated rates of anxiety and depression, is consistent with our prior studies using recruits from Prolific and suggests that this population exhibits higher incidence of anxiety [24,50–52]. Table 2 does shed some light on this difference: it shows that a similar high fraction of participants self-reported on their Prolific Profile that they have an ongoing mental health condition.

The demographic data listed in Table 2 provides several interesting insights on the recruited cohort, with respect to the presence or absence of above-threshold GAD. First, there is a significantly larger proportion of females in the anxious group compared to the males. This is consistent with previous findings suggesting that anxiety is more prevalent in females than males [44]. We feel that this confirms that it is useful to consider separate female-only and male-only datasets to avoid the bias introduced by sex when exploring features that may correlate with GAD-7. For example, pitch (F0) would typically be higher for females, and as a result, sex effects could easily confound the association between pitch and anxiety.

The rows of Table 2 that show the proportion of anxious/non-anxious participants by income suggest that there is a relationship between income and anxiety: the two very lowest categories of income show a disproportionately higher amount of anxiety. There is a downward trend of anxiety with income until the very last category, which is £60,000 and above. It is interesting that above a certain income level, anxiety seems to increase, although this is consistent with prior studies on anxiety and income [46].

Similarly, with respect to age, younger participants are more likely to be in the anxious group, which is consistent with previous work [45].

Post-Task Self-Report Anxiety Measure

As described in Section Anxiety Measures, we used the post-task, self-reported anxiety measure as an internal check to see if Task 2 (the modified TSST task) induced more self-reported anxiety as compared to Task 1. Table 3 gives the paired t-test conducted on the two informal ratings of anxiety of the two tasks. It had a *P*-value < *.001*, indicating a significant difference and that Task 2 induced greater anxiety. Recall that most of the prior work discussed in Section Related Work also used mood induction tasks.

Amount of Speech

The results suggests that features relating to the amount of speech that the participants delivered in response to Task 2 had one of the highest correlation with their GAD-7 scale response across all the features explored in this work. Two features in particular captured this – *Speaking Duration* and *Word Count*, as shown in Table 4 (and their inter-correlation with each other is shown in Multimedia Appendix 6). In all cases the negative direction of the correlation suggests that participants who spoke more, tend to have lower GAD-7 scores. This result is consistent with previous work, as shown in the first data row of Table 7), but our study gives a much lower Pearson's correlation that prior work (*r = 0.12* in this study, vs. *r = 0.36* for [16]). We speculate that the more anxious a person is, the less confidence they would have about their speech, and so perhaps they speak less.

Acoustic Features

The main purpose of this work was to explore how acoustic features relate to anxiety. We wanted to determine if associations found in previous studies still hold with the larger sample size. Table 5 lists the features that have significant correlations, with *P < .05*, across all three datasets. The features with the strongest correlation in this set were *shimmer* on the all-samples dataset and the standard deviation of the 2nd and 3rd MFCCs for the male and female datasets, respectively. We note that there are multiple parameters used in the extraction of MFCC features, so a direct comparison of the specific MFCC features of our study with specific features of previous work is not possible since the prior work does not provide the exact parameters used to compute the MFCCs. The parameters used in the present study are given in Section Acoustic Features. That being said, in previous research, the 4th MFCC was the most significant from the 13 MFCC features in [13] and the standard deviation of the 1st MFCC in [20] had a significant correlation (*r = -0.36*; *P < .05*) with an anxiety scale. These results, from both our current study and previous work, suggests that signals of anxiety are present in the MFCC features.

The following features, listed as relevant in prior work, did not show significant correlations with GAD-7: the second and third formant; jitter; zero-crossing rate of the z-score of the power spectral density; and the articulation rate. Table 6 shows prior work associations with anxiety on these features and the correlation values obtained in our current study. It is important to note that, in previous research, these features were noted as significant or relevant, but no correlations with an indicator of anxiety were given. This makes it difficult to compare directly with the correlations obtained in our study.

Linguistic Features

Correlations between linguistic features extracted using the LIWC dictionaries [25] and the GAD-7 were presented in the result section. These had a higher correlation than the acoustic features, as shown in Table 8. The top LIWC category with the highest correlation in all the datasets is the count of punctuations. This includes the count of periods, which would indicate the number of separate sentences. The count of periods together with a negative correlation of words per sentence (WPS) indicates that the use of shorter sentences is positively associated with anxiety.

Other LIWC categories with high correlation in the all-sample dataset are Negative emotion (“negemo”; e.g., hurt, ugly, nasty), Anger (“anger”; e.g., hate, kill, annoyed), Anxiety (“anx”; e.g., worried, fearful), and Sad (“sad”; e.g., crying, grief, sad
). The Anger, Anxiety, and Sad categories are constituent subsets of the Negative emotion (“negemo”) category—i.e., words counted under one of the Anger, Anxiety, or Sad categories will also be counted for the negemo category. The high inter-correlation with each other is shown in Multimedia Appendix 6. The negemo count had a higher correlation than these individual sub-categories, suggesting that words relating to anger, anxiety and sad are capturing different dimensions of self-reported anxiety.

An LIWC category with a significant correlation that is present in the male dataset but not in the female dataset is the use of apostrophes (Apostro), indicating words with contractions (such as *I’ll*) were positively associated with GAD-7. Also, only for males, function words, including personal pronouns (*ppron*) had a significant positive correlation with anxiety. We speculate that anxious male individuals might use personal pronouns (which includes I, me, mine) to divert their attention from the anxiety-inducing event and focus on themselves. More generally, the increased use of personal pronouns has been shown to occur in individuals with depression [53], a highly comorbid mental health illness with GAD (but not only for males).

Another differentiation between males and females occur in the LIWC feature for words related to "power" (e.g., superior, bully). The “power” count had a positive correlation with GAD-7 for females and a negative correlation for males. We speculate that the negative correlation is somehow related to the stereotypical dominance behaviour associated with males.

In prior work studying associations between LIWC scores and anxiety, words related to anxiety and first-person singular pronouns were shown to be significantly associated with social anxiety [26], similar to our results. The same work has also shown that perceptual process words (see, hear, feel) are significantly associated with anxiety, which does not align with our results. For example, the LIWC category for *see* has a negative correlation in both the all the sample and the female dataset (as shown in Table 8). However, in [24], the category *see* had a positive correlation ($r = 0.31$; $P = .02$) with a social anxiety measure. We speculate that the use

of perceptual process words (See) might be a differentiating factor between social anxiety and Generalized anxiety disorder since it was positively correlated in the former and negatively correlated in the latter. The LIWC category for the perceptual process *hear*, on the other hand, had a positive correlation in both the all-sample and the male dataset (also shown in Table 8). Notice that both *see* and *hear* are perceptual processes, but the category for *see* is significant for females while the category for *hear* is significant for males.

Also, in prior work, death-related words were shown to have positive correlation with anxiety [24]. Our results (as shown in Table 8) show a similar trend where death-related words had a significant positive correlation in the male and all-sample dataset. However, a significant correlation was not observed in the female dataset.

The fact that there are several single-word categories that have significant correlations suggests that techniques that are able to look at multiple word meanings may have greater potential in making predictions.

Limitations

One limitation of this study is the use of self-report measures to assess Generalized Anxiety Disorder. Self-report measures, by nature, are subjective opinions that individuals have about themselves while filling the questionnaires and may not completely capture clinical symptoms. In this study, we took these self-report questionnaires as the true label of the audio samples. However, we believe that this is a good first step which gave us encouraging preliminary results. A psychiatric diagnosis would be an improved label but is clearly much more expensive to acquire.

A further limitation of this study is the selection bias that might be introduced during the recruitment of the participants. In Figure 1, only 49% of the 4,542 participants who initially accepted the offer from Prolific to participate finished the study. We were not able to collect the GAD-7 score of these participants who did not complete the study, so we do not know their level of anxiety. It is possible that these participants have a higher levels of anxiety, which caused them to drop out of the study.

Another limitation is differences in recording devices and the recording locations of the participants performing each task. Ideally, we would want every sample to be recorded using the same microphone in the same location with the same acoustics. This would reduce the potential bias introduced by different factors such as, for example, recording quality or background noise. At the same time, in a real-life scenario where an application to detect anxiety might be deployed, the recording equipment and the location will likely differ for everyone. So, this limitation could be unavoidable, and it might even be essential to take these types of differences into consideration.

Conclusion

We have presented results from a large-*N* study examining the relationship between speech and generalized anxiety disorder. Our data collection relies on participants using home recording devices, hence capturing variations in acoustic environments that will need to be factored in when deploying tools for the detection of mental health disorders in the wild. Our goal was to provide a useful benchmark for future research, by assessing the extent to which results from previous research are generalizable to our data collection approach and larger

dataset. We tested the most common acoustic and linguistic features associated with anxiety in previous studies, and provided detailed correlation tables broken down by demographics.

Our findings are decidedly mixed. On the one hand, with our larger dataset, we find modest correlations between anxiety and several features of speech, including speaking duration and acoustic features such as MFCCs, LPCCS, Shimmer, Fundamental Frequency and first formant. However, other features elsewhere shown to correlate with anxiety—including second and third formant, Jitter, and ZCR-zPSD—were not significantly associated with anxiety in our study. Although these null findings do not entirely rule out the potential of more sophisticated learning models for this task, we believe that researchers should be wary of inherent difficulties. Readers should also note that our data collection already sidesteps additional challenges that we expect to influence the detection of anxiety disorders from speech, such as variations in accents, dialects and spoken language. On the other hand, we did find statistically significant correlations for a subset of speech features from previous research. This suggests that there may be a fundamental pathway between anxiety and the production of speech, one that is robust enough to be generalized to the population.

Future investigations could explore if features of speech from Task 1 (simple reading of a passage) exhibits correlations with GAD-7, or if those features could be used as a control for the features of a Task 2 (the modified TSST task). It may also be informative to separate out different age groups (e.g., younger, older) to see the if there is a specific impact of speech features on GAD-7.

Acknowledgements

This research was funded by a University of Toronto XSeed grant, NSERC Discovery Grant RGPIN-2019-04395, and Social Sciences and Humanities Research Council (SSHRC) Partnership Engage Grant #892-2019-0011

Conflicts of Interest

DD and BS are employees of Winterlight Labs.

Abbreviations

GAD-7: Generalized Anxiety Disorder 7-item scale
GAD: Generalized Anxiety Disorder
HADS: Hospital Anxiety and Depression Scale
LIWC: Linguistic Inquiry and word Count
LPCC: linear prediction Cepstral Coefficient
MFCC: Mel Frequency Cepstral Coefficient
PRCS: Personal Report of Confidence as a Speaker scale
SAD: Social Anxiety Disorder
STAI: State-Trait Anxiety Inventory
TSST: Trier Social Stress Test
WER: Word Error Rate
ZCR-zPSD: zero crossing rate of the power spectral density

References

1. Public Health Canada. Mental Health - Anxiety Disorders - Canada.ca [Internet]. Public Health Agency of Canada; 2009. Available from: https://www.canada.ca/en/health-canada/services/healthy-living/your-health/diseases/mental-health-anxiety-disorders.html

2. Roberge P, Fournier L, Duhoux A, Nguyen CT, Smolders M. Mental health service use and treatment adequacy for anxiety disorders in Canada. Soc Psychiatry Psychiatr Epidemiol 2011;46(4):321–330. PMID:20217041

3. Koerner N, Dugas MJ, Savard P, Gaudet A, Turcotte J, Marchand A. The Economic Burden of Anxiety Disorders in Canada. Canadian Psychology/Psychologie canadienne 2004;45(3):191–201. [doi: 10.1037/h0088236]

4. Hidalgo RB, Sheehan DV. Generalized anxiety disorder. Handb Clin Neurol 2012;106:343–362. PMID:22608630

5. Hoehn-Saric R, McLeod DR. The peripheral sympathetic nervous system. Its role in normal and pathologic anxiety. Psychiatr Clin North Am 1988 Jun;11(2):375–386. PMID:3047706

6. Thompson AR. Pharmacological agents with effects on voice. Am J Otolaryngol 1995 Feb;16(1):12–18. PMID:7717466

7. Barrett LF. How emotions are made: the secret life of the brain. First Mariner Book edition. Boston New York: Mariner Books; 2018. ISBN:978-1-328-91543-6

8. McGinnis EW, Anderau SP, Hruschak J, Gurchiek RD, Lopez-Duran NL, Fitzgerald K, Rosenblum KL, Muzik M, McGinnis RS. Giving Voice to Vulnerable Children: Machine Learning Analysis of Speech Detects Anxiety and Depression in Early Childhood. IEEE J Biomed Health Inform 2019 Nov;23(6):2294–2301. PMID:31034426

9. Buske-Kirschbaum A, Jobst S, Wustmans A, Kirschbaum C, Rauh W, Hellhammer D. Attenuated free cortisol response to psychosocial stress in children with atopic dermatitis. Psychosom Med 1997 Aug;59(4):419–426. PMID:9251162

10. Ali S, Tanweer S, Khalid S, Rao N. Mel Frequency Cepstral Coefficient: A Review. Proceedings of the 2nd International Conference on ICT for Digital, Smart, and Sustainable Development, ICIDSSD 2020, 27-28 February 2020, Jamia Hamdard, New Delhi, India [Internet] New Delhi, India: EAI; 2021 [cited 2022 Jan 25]. [doi: 10.4108/eai.27-2-2020.2303173]

11. Davies DL, Bouldin DW. A Cluster Separation Measure. IEEE Trans Pattern Anal Mach Intell 1979 Apr;PAMI-1(2):224–227. [doi: 10.1109/TPAMI.1979.4766909]

12. Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995 Sep;20(3):273–297. [doi: 10.1007/BF00994018]

13. Özseven T, Düğenci M, Doruk A, Kahraman Hİ. Voice traces of anxiety: acoustic parameters affected by anxiety disorder. Archives of Acoustics Committee on Acoustics PAS, PAS Institute of Fundamental Technological Research, Polish Acoustical Society; 2018;625–636. [doi: 10.24425/AOA.2018.125156]

14. Weeks JW, Lee C-Y, Reilly AR, Howell AN, France C, Kowalsky JM, Bush A. “The Sound of Fear”: assessing vocal fundamental frequency as a physiological indicator of social anxiety disorder. J Anxiety Disord 2012 Dec;26(8):811–822. PMID:23070030

15. Julian LJ. Measures of anxiety: State-Trait Anxiety Inventory (STAI), Beck Anxiety Inventory (BAI), and Hospital Anxiety and Depression Scale-Anxiety (HADS-A). Arthritis Care Res (Hoboken) 2011 Nov;63 Suppl 11:S467-472. PMID:22588767

16. Laukka P, Linnman C, Åhs F, Pissiota A, Frans Ö, Faria V, Michelgård Å, Appel L, Fredrikson M, Furmark T. In a Nervous Voice: Acoustic Analysis and Perception of Anxiety in Social Phobics’ Speech. J Nonverbal Behav 2008 Dec;32(4):195–214. [doi: 10.1007/s10919-008-0055-9]

17. Spielberger CD. State-Trait Anxiety Inventory. In: Weiner IB, Craighead WE, editors. The Corsini Encyclopedia of Psychology [Internet] Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2010 [cited 2022 Jan 25]. p. corpsy0943. [doi: 10.1002/9780470479216.corpsy0943]

18. Albuquerque L, Valente ARS, Teixeira A, Figueiredo D, Sa-Couto P, Oliveira C. Association between acoustic speech features and non-severe levels of anxiety and depression symptoms across lifespan. PLoS One 2021;16(4):e0248842. PMID:33831018

19. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. Acta Psychiatr Scand 1983 Jun;67(6):361–370. [doi: 10.1111/j.1600-0447.1983.tb09716.x]

20. Wortwein T, Morency L-P, Scherer S. Automatic assessment and analysis of public speaking anxiety: A virtual audience case study. 2015 International Conference on Affective Computing and Intelligent Interaction (ACII) [Internet] Xi’an, China: IEEE; 2015 [cited 2022 Jan 25]. p. 187–193. [doi: 10.1109/ACII.2015.7344570]

21. Gilkinson H. Social fears as reported by students in college speech classes. Speech Monographs 1942 Jan;9(1):141–160. [doi: 10.1080/03637754209390068]

22. Ali S, Tanweer S, Khalid SS, Rao N. Mel Frequency Cepstral Coefficient: A Review. ICIDSSD 2020 European Alliance for Innovation; 2021;92.

23. Hagenaars MA, van Minnen A. The effect of fear on paralinguistic aspects of speech in patients with panic disorder with agoraphobia. J Anxiety Disord 2005;19(5):521–537. PMID:15749571

24. Di Matteo D, Wang W, Fotinos K, Lokuge S, Yu J, Sternat T, Katzman MA, Rose J. Smartphone-Detected Ambient Speech and Self-Reported Measures of Anxiety and Depression: Exploratory Observational Study. JMIR Form Res 2021 Jan 29;5(1):e22723. PMID:33512325

25. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The development and psychometric properties of LIWC2015. The University of Texas at Austin; 2015.

26. Anderson B, Goldin PR, Kurita K, Gross JJ. Self-representation in social anxiety disorder: linguistic analysis of autobiographical narratives. Behav Res Ther 2008 Oct;46(10):1119–1125. PMID:18722589

27. Palan S, Schitter C. Prolific.ac—A subject pool for online experiments. Journal of Behavioral and Experimental Finance 2018 Mar;17:22–27. [doi: 10.1016/j.jbef.2017.12.004]

28. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092–1097. PMID:16717171

29. Reilly J, Fisher JL. Sherlock Holmes and the Strange Case of the Missing Attribution: A Historical Note on “The Grandfather Passage.” J Speech Lang Hear Res 2012 Feb;55(1):84–88. [doi: 10.1044/1092-4388(2011/11-0158)]

30. Kirschbaum C, Pirke KM, Hellhammer DH. The ’Trier Social Stress Test’--a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology 1993;28(1–2):76–81. PMID:8255414

31. Gerra G, Zaimovic A, Zambelli U, Timpano M, Reali N, Bernasconi S, Brambilla F. Neuroendocrine responses to psychological stress in adolescents with anxiety disorder. Neuropsychobiology 2000;42(2):82–92. PMID:10940763

32. Jezova D, Makatsori A, Duncko R, Moncek F, Jakubek M. High trait anxiety in healthy subjects is associated with low neuroendocrine activity during psychosocial stress. Prog Neuropsychopharmacol Biol Psychiatry 2004 Dec;28(8):1331–1336. PMID:15588760

33. Endler NS, Kocovski NL. State and trait anxiety revisited. J Anxiety Disord 2001 Jun;15(3):231–245. PMID:11442141

34. First MB, editor. SCID-I: Structured clinical interview for DSM-IV AXIS I disorders. Clinician version ; Set. Washington, DC: American Psychiatric Press; 1997. ISBN:978-0-88048-934-8

35. Kliper R, Portuguese S, Weinshall D. Prosodic Analysis of Speech and the Underlying Mental State. In: Serino S, Matic A, Giakoumis D, Lopez G, Cipresso P, editors. Pervasive Computing Paradigms for Mental Health [Internet] Cham: Springer International Publishing; 2016 [cited 2022 Jan 25]. p. 52–62. [doi: 10.1007/978-3-319-32270-4_6]

36. Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological aspects of natural language. use: our words, our selves. Annu Rev Psychol 2003;54:547–577. PMID:12185209

37. Aalto D, Malinen J, Vainio M. Formants. Oxford Research Encyclopedia of Linguistics [Internet] Oxford University Press; 2018 [cited 2022 Jan 25]. [doi: 10.1093/acrefore/9780199384655.013.419]

38. Silber-Varod V, Kreiner H, Lovett R, Levi-Belz Y, Amir N. Do social anxiety individuals hesitate more? The prosodic profile of hesitation disfluencies in Social Anxiety Disorder individuals. Speech Prosody 2016 [Internet] ISCA; 2016 [cited 2022 Jan 25]. p. 1211–1215. [doi: 10.21437/SpeechProsody.2016-249]

39. Fuller BF, Horii Y, Conner DA. Validity and reliability of nonverbal voice measures as indicators of stressor-provoked anxiety. Res Nurs Health 1992;15(5):379–389. PMID:1529122

40. Sabahi S. my-voice-analysis [Internet]. GitHub repository. GitHub; 2019. Available from: https://github.com/Shahabks/my-voice-analysis

41. Lenain R, Weston J, Shivkumar A, Fristed E. Surfboard: Audio Feature Extraction for Modern Machine Learning. Interspeech 2020 [Internet] ISCA; 2020 [cited 2022 Jan 25]. p. 2917–2921.

[doi: 10.21437/Interspeech.2020-2879]

42. McFee B, Raffel C, Liang D, Ellis D, McVicar M, Battenberg E, Nieto O. librosa: Audio and Music Signal Analysis in Python. Proceedings of the 14th python in science conference [Internet] Austin, Texas; 2015 [cited 2022 Jan 25]. p. 18–24. [doi: 10.25080/Majora-7b98e3ed-003]

43. Hashemipour S, Ali M. Amazon Web Services (AWS) – An Overview of the On-Demand Cloud Computing Platform. In: Miraz MH, Excell PS, Ware A, Soomro S, Ali M, editors. Emerging Technologies in Computing [Internet] Cham: Springer International Publishing; 2020 [cited 2022 Apr 1]. p. 40–47. [doi: 10.1007/978-3-030-60036-5_3]

44. Cameron OG, Hill EM. Women and Anxiety. Psychiatric Clinics of North America 1989 Mar;12(1):175–186. [doi: 10.1016/S0193-953X(18)30459-3]

45. Krasucki C, Howard R, Mann A. The relationship between anxiety disorders and age. Int J Geriatr Psychiatry 1998 Feb;13(2):79–99. PMID:9526178

46. Dijkstra-Kersten SMA, Biesheuvel-Leliefeld KEM, van der Wouden JC, Penninx BWJH, van Marwijk HWJ. Associations of financial strain and income with depressive and anxiety disorders. J Epidemiol Community Health 2015 Jul;69(7):660–665. PMID:25636322

47. Pellegrini T, Hämäläinen A, Mareüil PB de, Tjalve M, Trancoso I, Candeias S, Dias MS, Braga D. A corpus-based study of elderly and young speakers of European Portuguese: acoustic correlates and their impact on speech recognition performance. Interspeech 2013 [Internet] ISCA; 2013 [cited 2022 Apr 11]. p. 852–856. [doi: 10.21437/Interspeech.2013-241]

48. Farrow K, Grolleau G, Mzoughi N. What in the Word! The Scope for the Effect of Word Choice on Economic Behavior: What in the Word! Kyklos 2018 Nov;71(4):557–580. [doi: 10.1111/kykl.12186]

49. Baba K, Shibata R, Sibuya M. Partial Correlation and Conditional Correlation as Measures of Conditional Independence. Aust NZ J Stat 2004 Dec;46(4):657–664. [doi: 10.1111/j.1467-842X.2004.00360.x]

50. Di Matteo D, Fotinos K, Lokuge S, Yu J, Sternat T, Katzman MA, Rose J. The Relationship Between Smartphone-Recorded Environmental Audio and Symptomatology of Anxiety and Depression: Exploratory Study. JMIR Form Res 2020 Aug 13;4(8):e18751. PMID:32788153

51. Di Matteo D, Fotinos K, Lokuge S, Mason G, Sternat T, Katzman MA, Rose J. Automated Screening for Social Anxiety, Generalized Anxiety, and Depression From Objective Smartphone-Collected Data: Cross-sectional Study. J Med Internet Res 2021 Aug 13;23(8):e28918. PMID:34397386

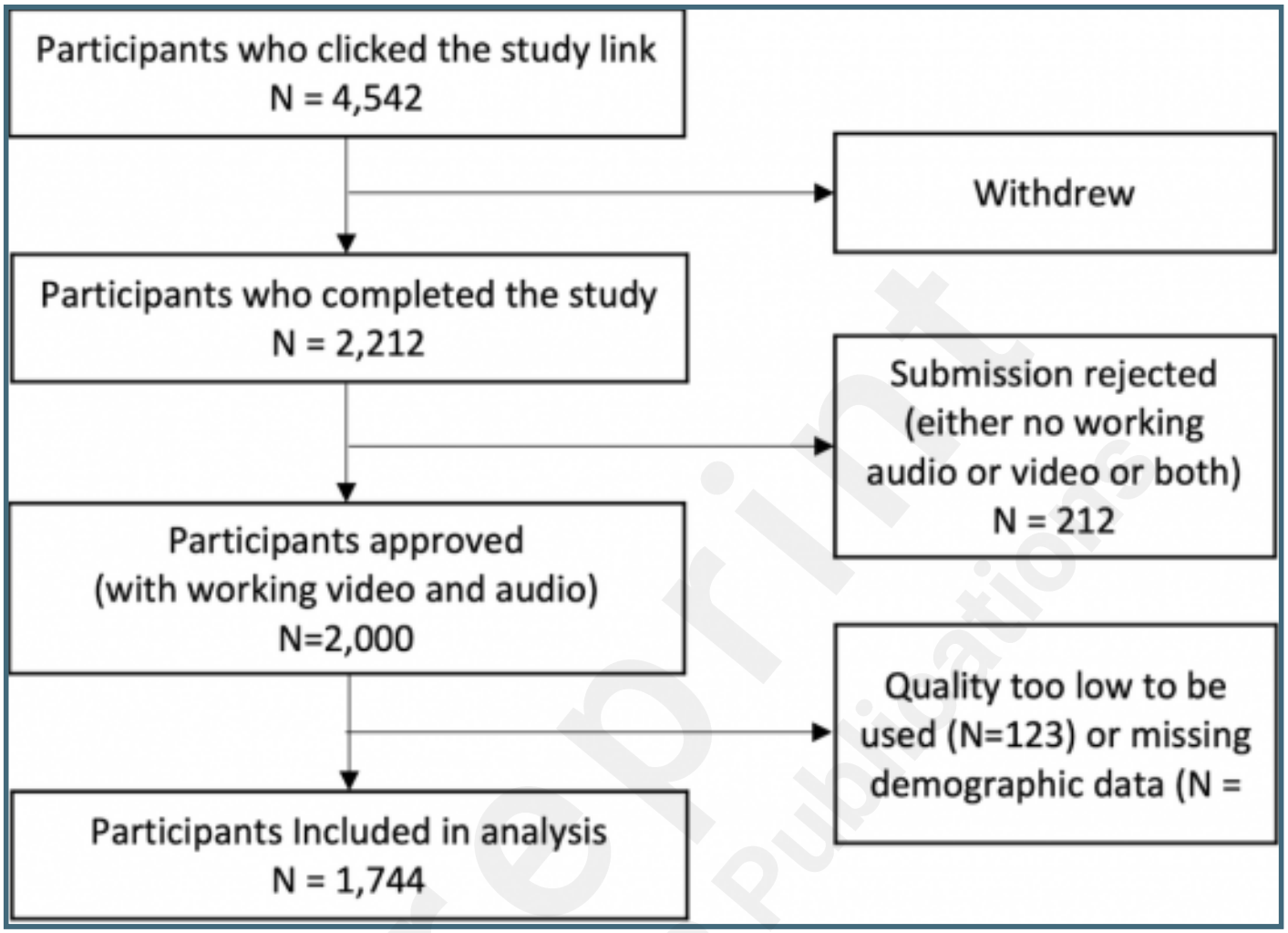
52. Di Matteo D. Inference of Anxiety and Depression from Smartphone-collected Data [PhD Thesis]. University of Toronto; 2021.

53. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preoţiuc-Pietro D, Asch DA, Schwartz HA. Facebook language predicts depression in medical records. Proc Natl Acad Sci U S A 2018 Oct 30;115(44):11203–11208. PMID:30322910

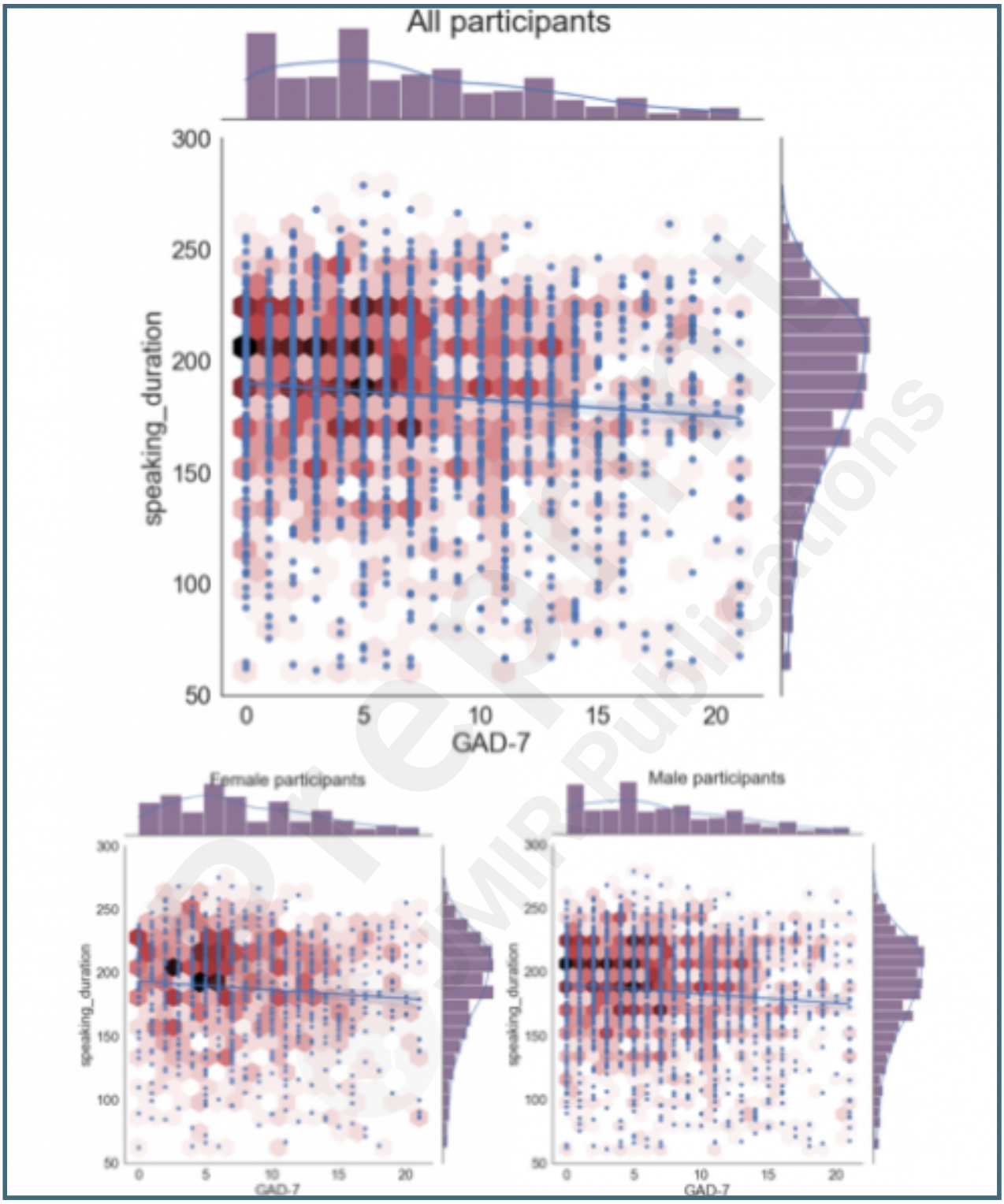
Supplementary Files

Figures

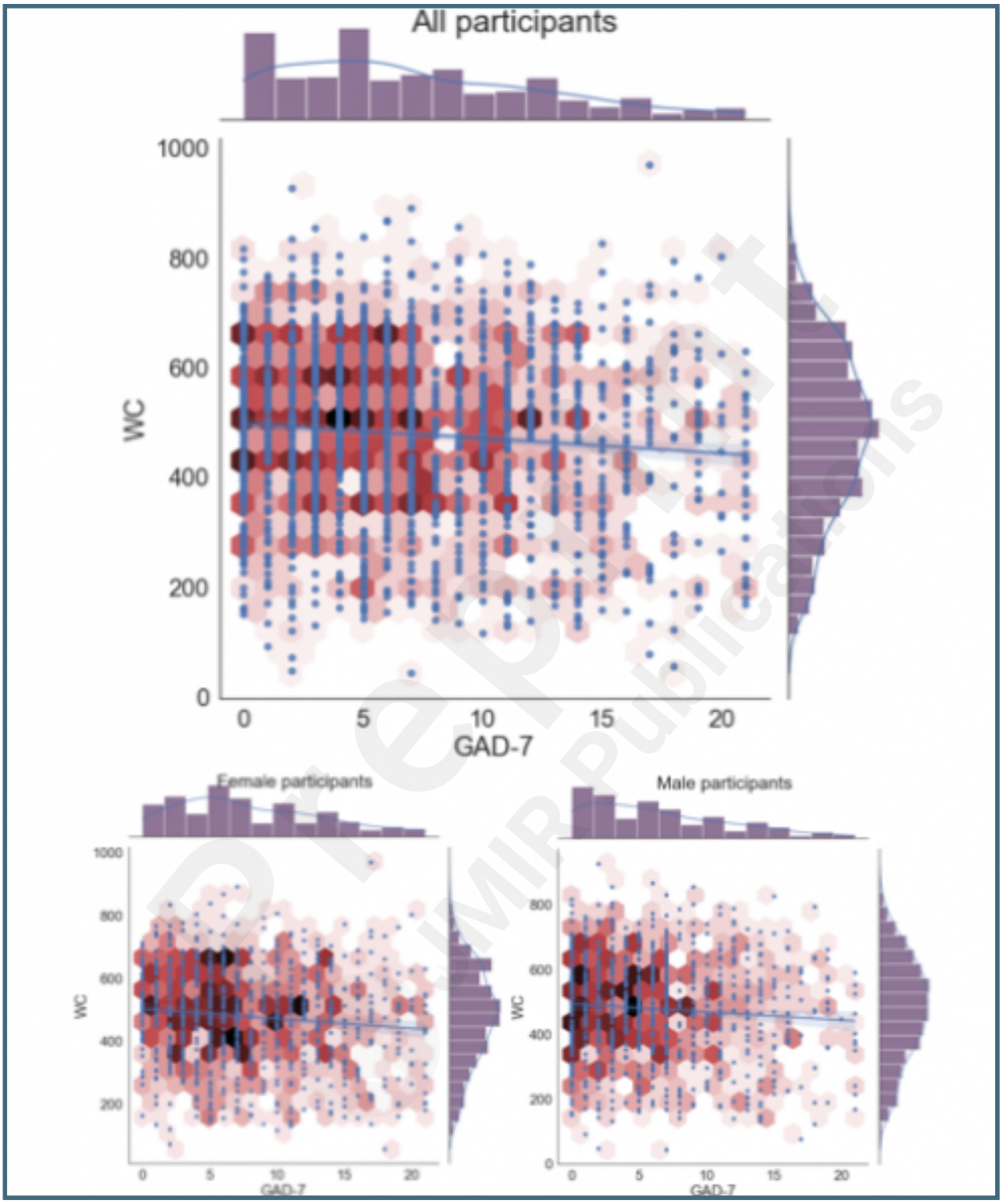
Study recruitment flow chart.



Speaking Duration vs. GAD-7 Scatter plot and distributions.



Word Count vs. GAD-7 Scatter plot and distributions.



Multimedia Appendixes

Web application screenshot.
URL: http://asset.jmir.pub/assets/a535e091aa80d8b203d12d2c22f93463.pdf

My Grandfather passage.
URL: http://asset.jmir.pub/assets/60846f39119c14b64b19c343b760428c.pdf

Speech encouragement statements.
URL: http://asset.jmir.pub/assets/42d211f1bc9a5a255a1028fc938c0665.pdf

Excluded data analysis.
URL: http://asset.jmir.pub/assets/603be40dd36018f39d45c6b755c52161.pdf

Correlation between Demographics and acoustic/Linguistic features.
URL: http://asset.jmir.pub/assets/80f1e3ca7ea2ccc3def43dc63803d638.pdf

All-sample dataset significant features inter-correlation.
URL: http://asset.jmir.pub/assets/3993f5e314ebc6d8efd47f9693104e9f.xlsx

Female-sample dataset significant features inter-correlation.
URL: http://asset.jmir.pub/assets/a8b0c7c596d8d84e2b1c33bea81cf882.xlsx

Male-sample dataset significant features inter-correlation.
URL: http://asset.jmir.pub/assets/e7f72fd806924e30c2649d874dbf49ef.xlsx