A Human Annotation Guide for Mental Health Speech Collections

Brian Stasak[1,2], Julien Epps[2], Mark Larsen[1], and Helen Christensen[1]

[1]Black Dog Institute, UNSW, Sydney, NSW – Australia
[2]School of Elec. Eng. & Telecomm., UNSW, Sydney, NSW – Australia
b.stasak@unsw.edu.au, j.epps@unsw.edu.au, mark.larsen@blackdog.org.au, h.christensen@blackdog.org.au

Abstract

While large amounts of recorded speech audio data are collected for medical analysis, there is not a compact guideline available that outlines which human-rated annotations are important to consider when analyzing speech from individuals with potential mental illness. Herein, an annotation guideline is proposed that highlights fifty-two different speech-audio recording transcription considerations, including several new ones related to voice accent, prosodic, intelligibility, quality, auxiliary behavior, task compliance, disfluency, and noise factors. Further, a free, new software tool for recording recommended speech-based mental health annotations is provided to help scientists who may be unfamiliar with speech-audio data collection.

Index Terms: digital medicine, survey, transcription, voice

1. Introduction

Human-rated annotations derived from audio speech recordings are vital in understanding communication behavior, protocol design, real-world compliance, and noise factors that may adversely impact the quality of recorded audio. Speech annotations provide specific details about abnormal speech-language characteristics exhibited by individuals experiencing mental ill-health and may be used to help screen or predict mental health outcomes. While automatic annotation methods (e.g., automatic speech recognition, signal-to-noise ratio, syllable durations) are available to evaluate speech recordings, the human-rated annotation approach is still the 'gold standard' [1]. Moreover, automatic annotation approaches require kernels of data with human-rated annotations to generate robust modeling and to establish system test performance baselines.

While many medical speech datasets concerning neurological disorders (e.g., aphasia, dementia, Parkinson's disease) frequently contain transcripts including linguistic-level part-of-speech annotations, the few scarcely available mental health speech datasets omit this level of information [2]. Moreover, mental health speech datasets rarely include subjective human-rated insights, such as voice quality, paralinguistic traits, auxiliary behaviors, task compliance, and noise annotations. It is believed that these types of annotations may further reveal abnormalities in emotion and language patterns in individuals with mental health disorders [2].

While efforts have been made to create uniform, widely accepted annotation guidelines, most mental health research speech corpora still contain different sets of annotation taxonomies and procedures [3]. Many data collections provide vague details on the human annotation process, making it more difficult to replicate the annotation standards in future studies [1-3]. Thus, there is a need for a uniformly coded metadata annotation system and a convenient easy-to-use human-labeling software tool for mental health speech data collections. Recently, voice quality, prosodic, and speech disfluency annotations have proven useful in detecting individuals with depression [4-9]. Other reported speech annotations, such as auxiliary vocal behaviors, task compliance, and noise have been explored in the mental health literature [10-12]. Speech-language and audio-signal related annotations are of substantial interest since modern mental health smartphone collection speech recordings are often conducted in a real-world natural home environment without a clinician/administrator present.

In this paper, annotations based on six different speech-audio related categories were chosen based on previous speech-based mental health and voice studies [4-12]. Per speech annotation category, this paper includes a brief literature review, a description of known connections to mental illness, and proposed annotation approaches to help capture relevant speech behaviors for further analysis. A newly proposed human annotation system and uniform software tool provides groundwork for scientists labeling speech-audio metadata to help discover new speech-related sub-symptoms in mental illness. Annotated dataset results may lead to the development of automated near-real-time annotation systems, which can be implemented on a large scale – reducing annotation time and costs. Further, automated annotation systems could also help to reduce possible human-rater bias, improve health annotator safety, and allow greater patient anonymity.

2. Speech Annotation Categories

2.1. Voice Quality

There are many factors that can impact voice quality, such as an individual's height, stress, personal habits (e.g., alcohol consumption, smoking, vaping), amount of vocal use/abuse, trauma injury, and diagnosed illness (e.g., voice, neurological, psychogenic disorders). In general populations, it is estimated that up to 9% of the population has a clinical voice disorder [13, 14]. Recent speech-based studies [8, 9] have discovered strong correlations between mental health disorders and voice quality. For example, [8] found that when given read sentence tasks, individuals with recent suicidal ideation or attempt exhibited significantly poorer voice quality scores than non-suicidal individuals. Further, in [9], it was found that despite different speaker age groups (e.g., 18-34, 35-48, 49-62, 63-79), individuals with higher depression severity scores demonstrated poorer voice quality.

Individuals' perception of their voice quality often reflects their degree of self-esteem and recognized changes in voice due to medical conditions [15]. Based on human perceptual ratings, undesirable vocal characteristics include unusual nasality and monotony [16]. For example, according to [17], individuals

with chronic sinusitis and nasal polyp disorders have nearly twice the risk of clinical depression disorders than healthy controls possibly in part due to self-esteem, voice quality, and other contributing factors (e.g., sleep habit, stress).

Studies [18, 19] have highlighted that the GRBASI subjective voice quality assessment is widely used in the field of clinical voice evaluation and research. Currently, the GRBASI is the evaluation standard, even when compared to more objective automatic approaches [18]. Further, GRBASI ratings were shown to have good interrater reliability across many different types of voice disorders [19].

While all individuals have unique resonance/production factors that result from a blend of voice quality attributes, most are normal/slight and do not negatively impact verbal expressiveness and intelligibility. The GRBASI utilizes a four-point scale (e.g., 0-normal, 1-slight 2-moderate, 3-severe) and it requires a pre-trained listener to subjectively rate a speaker based on six different voice quality attributes. Individual GRBASI scores are then added together to produce a GRBASI composite score. The higher the GRBASI composite score, the poorer an individual's voice quality. The six different GRBASI voice quality attributes are described briefly as follows: (1) *grade*: hoarseness, raspy; (2) *roughness*: low frequency vibration irregularity; (3) *breathiness*: air escaping, leakage, whispery; (4) *asthenia*: loss of power, weakness; (5) *strain*: hyper-functional phonation, rattle; and (6) *instability*: inconsistent quality, voice fluctuation. For more details regarding the GRBASI voice quality assessment, see [18, 19].

In addition to the GRBASI voice quality ratings, another important voice quality is nasal resonance. Nasality is unrelated to laryngeal function (e.g., vocal folds) and requires independent activation of the nasopharynx. In English, only the /n, m, ŋ/ sounds have nasality. Both hypo/hyper nasality can cause a loss of speech clarity and linguistic stress between syllables [20]. Hyponasality means that for nasal sounds /n, m, ŋ/, no acoustic vocal energy passes through the nasal passage (i.e., no air can escape via the nostrils). This creates a voice quality found in individuals with a cold or allergies with a stuffed-up nose. Hypernasality means that vocal energy is accidently escaping into the nasal passage while making all sounds, rather than only nasal sounds /n, m, ŋ/. In hypernasality, there is a faulty control in palato-pharyngeal valving, which results in excessive nasalization of vowels and non-nasal consonants. Hypernasality causes a voice quality often found in individuals with Down Syndrome (i.e., due to alteration in velopharyngeal sphincter mobility and epipharynx narrowing) and cleft palate.

In addition to the recommended GRBASI voice quality assessment, a nasality annotation is also proposed using three possible presence indicators (0-normal, 1-hyponasality, 2-hypernasality). A score of '0' is normal nasopharynx function without noticeable hypo/hyper traits.

2.2. Accent, Prosody, Rate, and Confidence

Refugee/immigrant mental health studies have shown a high risk of mental distress due to various traumas (e.g., war, natural disaster, famine) and struggles (e.g., crime, discrimination, financial, employment) when compared with the general local population [21]. It has also been reported that some ethnicities/cultures have historically and significantly higher incidence of mental health issues. For example, in the United States, Native Americans have the highest rates of any mental health diagnosis than any other ethnic population [22]. Due to the high prevalence of mental health concerns in refugee/immigrant populations and potential non-native language skill deficits, speech accent is an important factor to consider when analyzing speech recordings.

Early subjective observational paralinguistic depression studies [23, 24] by clinicians described patients with depression as having abnormal speech patterns, such as weaker loudness, slower rate, flattened pitch, uniform rhythm, less verbosity, and an unusual lifeless or hollow sounding timbre. Modern acoustic speech studies [7, 25] have also indicated decreases in depressed speakers' loudness, stress-prosodic characteristics, verbal fluency, and overall rate-of-speech. A recent study [7] found that depressed populations exhibit more uniform syllable durations and flatter amplitude dynamics than healthy populations, resulting in less prosodic word stress and overall poorer intelligibility of speech. Under experimental conditions, researchers found that that individuals with social anxiety disorder subsequently exhibited decreased vocal confidence in contrast to a control group [26].

Rate-of-speech annotations have been explored in mental health studies [27-29], which have indicated changes in speech fluency during depressive, anxiety, or schizophrenic episodes. It is known that depression can impair fine-motor skills and lead to symptomatic psychomotor retardation (i.e., decreased rate-of-speech, increased syllable lengths, shorter sentence utterances) or agitation (i.e., excessive rapid gesturing, accelerated motor activity, and verbose activity).

It is expected in a speech collection that most speakers will use a native accent. Thus, a native accent binary annotation is proposed using the indicators 0-native and 1-foreign. The native accent annotation rating is subjectively based on the native production of a speaker's phonemes (i.e., often non-native speakers will use their native language phoneme prototypes). A new degree of accentedness annotation rating range is also proposed by the following: 0-native; 1-light foreign accent; 2-medium foreign accent; and 3-heavy foreign accent. Further, a new speech intelligibility rating based on five ratings is also proposed using: 0-very low; 1-low; 2-moderate; 3-high; and 4-very high. A 4-very high speech intelligibility rating is described as clear flawless enunciated speech (i.e., high comprehension), whereas a 0-very low intelligibility rating implies very little of what was spoken could be comprehended by a listener (i.e., accurately repeated/written).

The concept of natural speech continuum is rated using a binary whereby any recording with a natural stream of speech (i.e., without unnatural pauses/breaths) is given a '0'. Conversely, a recording with a speech pattern that has abnormal disruptions is given a '1' (i.e., abnormally broken or disrupted). Per recording, similarly to [8], an overall rate-of-speech is recommended using the following ratings: 0-very slow; 1-slow; 2-moderate; 3-fast; and 4-very fast. The rate-of-speech is rated based on the average of the entire recording. If the participant speaks very slowly during one sentence and then very fast for the next, a rating of 'moderate' (2) is given for the recording.

A prosodic annotation rating indicates the level of paralinguistic dynamics per recording. Prosodic annotation ratings are proposed as follows: 0-very flat; 1-flat; 2-moderate; 3-dynamic; and 4-very dynamic. Similar to the rate-of-speech annotation, the prosody annotation is rated based on the average of the entire recording. An example of 0-very flat prosody annotation rating includes a speaker that uses a monotone voice with minimal amplitude variation, rate of speech, and pitch contour (e.g., Ben Stein). On the contrary, a 4-very dynamic prosody annotation rating includes a speaker that uses excessively wide amplitude, rate of speech, and pitch contour ranges (e.g., Robin Williams).

The speech task confidence rating is related to how confident a speaker sounds (i.e., did the participant sound more assertive or unsure during his/her speech task recording?). A speaker's confidence is inferred by a combination of vocal amplitude, speed, directness, dominance, and recorded response fluidity [31]. The proposed task confidence annotation relies on five ratings (0-very low; 1-low; 2-moderate; 3-high; 4-very high). An example of a 0-very low rating is an individual that perceptually sounds uncertain of during an utterance. A 4-very high rated confident individual will produce utterances that sound more like a factual statement with increased speed/loudness. Note that an individuals may sound confident even though they may utter incorrect responses during speech tasks (e.g., picture naming, read sentences, Stroop color test).

2.3. Speech Disfluencies

Individuals with a serious mental illness (e.g., bipolar/unipolar depression, schizophrenia) struggle in nearly every aspect of speech production when compared with control participants [14]. Studies [4-8] have shown that depression can be detected through individuals' speech disfluencies. For example, analysis of read speech tasks demonstrated statistically significant feature differences in speech disfluencies, whereby when compared with non-depressed speakers, depressed speakers showed relatively higher recorded frequencies of hesitations (55% increase) and speech errors (71% increase) [7]. Another investigation of speech recordings taken from inpatients with suicidal ideation and suicide attempt had approximately twice as many hesitations and four times as many speech errors when compared with individuals in a control group [8]. Also, when compared with the control groups, it was shown in both studies [7, 8] that individuals with depression tended to incorrectly substitute words without self-corrections.

Disfluency annotations have been applied in previous speech-based mental health studies [7, 8]. Speech repeats occur at a syllable, word, or phrase level. Raw counts are annotated to record the exact number of repeats per type. In some instances, it is possible for a spoken utterance to include a count for all three repeat types. Hesitations include two proposed annotated forms: non-speech (e.g., pause) and speech (e.g., vocal held sound). A non-speech pause hesitation typically has an abnormal gap of silence (i.e., not due to end of phrase, breath, or emphasis) that occurs abruptly, disrupting the flow of what is being said. A vocalized speech hesitation is one which is achieved by abnormally holding a sound for an unusually longer duration during an utterance. A vocal speech hesitation will frequently occur when an individual is unsure or cues that there is more to say (i.e., holding speaker-turn-taking dominance, time for recollecting thoughts).

The speech error annotation is a total raw count summary of any instance of a speech error, including *all* disfluency types (e.g., grammar, phonological, repetitions, hesitations, substitutions, deletions, insertions, malapropisms). Individuals with non-native accents or English-as-second language should not have phonological disfluency errors tracked due to normal phoneme-mapping shifts (e.g., Grimm's Law). A raw count of specific speech error types is further provided for substitutions (e.g., '*took*' / '*take*'), deletions (e.g., '*cars*' / '*car_*'), insertions (e.g., '*He left yesterday*' / '*He also left yesterday*'), and malapropisms (e.g., '*amuse*' / '*mouse*'). Substitutions are usually a variant of the target word, target sounds, and are the same word class (e.g., noun, verb, adjective), whereas malapropisms involve the substitution of the entire word often with a nonsensical effect. Another annotation proposed is self-corrections [7, 8], wherein for every self-correction opportunity evaded a raw count is recorded. It is easiest to score self-correction during read speech tasks because it is known exactly what the speaker should have said.

2.4. Auxiliary Vocal Behaviors

Auxiliary vocal behaviors are common during speech production (e.g., coughs, laughs, throat clearing, sighs, yawns). Individuals with mental health disorders often have higher incidence of additional illness (e.g., comorbidity) [31], which may impact speech production and potentially increase any number of auxiliary vocal behaviors. Abnormal sleep patterns are a common symptom among individuals with depression, whereby roughly 75% have insomnia symptoms [32].

Self-comments (i.e., externalized monologue) are generally a coping strategy used during high cognitive load competition-type tasks. However, under certain conditions, excessive self-comment verbalizations can be an indicator of mental health disorders (e.g., schizophrenia, hallucinatory episode) [34]. For example, during a Stroop color test, some individuals make comments aloud about their test performance (e.g., "*Green*, *blue*, *yellow*, *white*, black, *oh I really messed up there*, *yellow*, *damn it*."). Formulaic language is common in everyday discourse, where it contributes to nearly a quarter of all conversational speech [10, 12]. Formulaic language includes conventional word expressions, proverbs, idioms, expletives, hedges, bundles, and fillers.

Similar to [12], reporting a raw count for common word filler types (e.g., *ah*, *er*, *mm*, *uh*, *um*, *so*, *like*, *you know*) additional proposed raw count annotations are self-comments, expletives (e.g., swears), and speech auxiliary behavior types (e.g., coughs, laughs, throat clearing, sighs, yawns).

2.5. Task Compliance

Task compliance is rarely reported in speech-based data collections. Open-source speech datasets that are frequently relied upon for experimental analysis and publications often contain recordings wherein subjects did not follow the given directions properly. Task compliance is useful in understanding what percentage of recordings were completed correctly, abnormal verbal responses, and also whether or not some speaker subsets have higher compliance than others. The task compliance can also provide better insight in terms of how well a task is explained for participants, and further, whether the task instructions should be revised to increase task compliance.

The newly proposed task compliance annotation rating indicates whether the task directions were properly 0-completed or 1-non-compliant. Some examples of a marking of 1-non-compliance include: a speaker who says nothing during the recording; a speaker that says something else than what the task specifies; the speaker has another individual speak on his/her behalf (i.e., can determine based on gender/age factors); and the speaker does not fully complete the given task. Some speech tasks (e.g., held vowel, diadochokinetic, word fluency, Stroop color test, read sentence) require the speaker to generate specific target examples of related word tokens. Therefore, a proposed task-specific count annotation is also recommended (i.e., raw target word count versus total word count).

2.6. Noise Factors

During recorded speech samples, especially those collected outside of a laboratory setting (e.g., natural environment), background and channel noise level are important to annotate.

Studies [34, 35] have demonstrated that individuals living in home environments with noise pollution have greater incidence of poorer health and quality of life. Noise annotations can reveal information that may be related to higher incidence of mental health disorders (e.g., noise pollution, ability to focus on single task). This noise-based metadata can also be useful to build noise-specific modeling (i.e., increase noise robustness, feature reliability), test specific automated system performance given different noise types/levels, and help explain unusual statistical feature analysis results found in mental health populations.

For background noise level, four ratings are proposed (0-none; 1-minimal; 2-moderate; 3-severe). A background noise level of 0-none is a very quiet environment without background noise, wherein the signal-to-noise ratio is very strong, and the speech is clear. A background noise level of 1-minimal is where the signal-to-noise ratio is still strong, but there may be a small amount of background noise, and it does not impede intelligibility. A background noise level of 2-moderate is when the noise level has increased to nearing the level of the speech amplitude, possibly making some part/s of speech harder to comprehend. A background noise level of 3-severe is when the background noise is so loud it surpasses the speech signal level, making the speech difficult to understand.

A proposed noise duration annotation is as follows: 0-none; 1-intermittant; 2-continous; and 3-mixed (i.e., both intermittent and continuous). Further, specific noise type annotations are suggested the following: 0-none; 1-Radio/TV; 2-babble; 3-machinery; and 4-other. Further, the background speaker annotation (i.e., secondary speaker/s) is a binary value of either '0-absent' or '1-present'. Device noise annotation related to a recording device channel issues (e.g., bit-loss, buzz, clipping, device error, hum) is also proposed using the following scale: 0-none; 1-minimal; 2-moderate; 3-severe; and 4-very severe.

Similar to the standard mean opinion score (MOS) annotation based on audio listening speech-to-noise signal quality [36], a new rating scale for rating speech-to-noise signal quality is proposed: 1-bad (very annoying and objectionable); 2-poor (annoying but not objectionable); 3-fair (perceptible and slightly annoying); 4-good (just perceptible, but not annoying; and 5-excellent (level of distortion imperceptible). The speech-to-noise signal quality rating deals with how perceptually pleasing a recording is in terms of signal-to-noise ratio, whereas the previously mentioned noise level annotation is a perception of the clarity of the speech signal despite possible background noise or compression loss. A secondary activity annotation is also proposed as 0-none, 1-eating/drinking; 2-driving/in-vehicle; and 3-other.

3. Annotation Software Design

A new survey tool for human annotations and audio listening capability was created using python code which is freely available at: https://www.researchgate.net/profile/brian_stasak. This includes a GUI for batch processing of audio file recording playback and file-by-file survey reporting. A digital survey was created that included all 52 annotations described in this paper. Further, metadata regarding the annotator, such as gender, age, first-language, headphone/speaker type, years annotator experience, and listening environment loudness, were also included in the digital survey app. The app exports the annotation survey information into a .CVS file format for later analysis. Currently, this app tool is helping to annotate an ongoing Black Dog Institute mental health data collection that to-date includes over 7k recordings from more than 1k school-aged participants in naturalistic conditions [37].

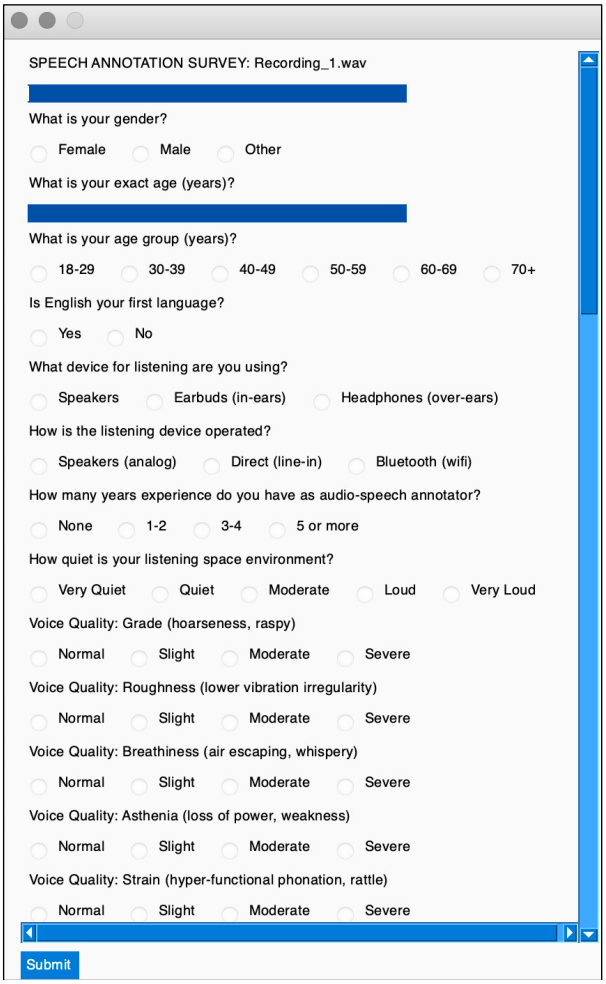

Figure 2: Image of python-based human-rated annotation GUI for speech-audio playback listening and survey notations.

4. Conclusion

While previous speech-based mental health data collections have evaluated one or two human annotation categories, none have conducted a comprehensive examination that includes *all* six categories described in this paper. Due to the limited number of speech-based mental health databases available, it is imperative that maximum information is gleaned for analysis. This proposed speech-based mental health annotation guide was streamlined into a novel annotation survey tool, which is useful in unifying annotation standards and comparing future datasets. While transcription annotation software (e.g., Praat, SpeechTools) is commonly used in audio-speech notation, annotation software guides readily available that implement broad categorical annotations are limited. This annotation tool is helpful to non-audio-speech experts who are interested in evaluating recorded audio (e.g., medical professionals, linguists); and/or digital health experts designing automatic speech-based machine learning models, whereby human-rated annotations are required to generate 'ground-truth' performance baseline and test model parameter optimization.

5. Acknowledgements

Funding for this project came from a NHMRC Project Grant Awarded to HC (APP1120646).

6. References

[1] Li, X., Liu, H., Kury, F., Yuan, C., Butler, A., Sun, Y., Ostropolets, A., Xu, H., and Weng, C., "A comparison between human and NLP-based annotation of clinical trial eligibility criteria text using the OMOP common data model", In: Proc. AMIA Joint Summits on Transl. Science, pp. 394–403, 2021.

[2] Li, Y., Ding, H., and Li, D., "Speech databases for mental disorders: a systematic review", General Psych., vol. 32(3), pp. 1–10, 2019.

[3] Delais-Roussarie, E. and Post, B., "Corpus annotation: methodology and transcript systems", The Oxford Handbook of Corpus Phonology, Oxford University Press, Oxford - UK, 2014.

[4] Esposito, A., Esposito, A.M., Likforman-Sulem, L., Maldonato, M.N., Vinciarelli, A., "On the significance of speech pauses in depressive disorders: results on read and spontaneous narratives", Recent Advances in Nonlinear Speech Processing, Smart Innovation, Systems and Technologies, vol. 48, Springer, 2016.

[5] Oxman, T. E., Rosenberg, S. D., Schnurr, P. P., and Tucker, G. J., "Diagnostic classification through content analysis of patients' speech", American J. of Psych., vol. 145(4), pp. 464–468, 1988.

[6] Rubino, A., D'Agostino, L., Sarchiola, L., Romeo, D., Siracusano, A., and Docherty, N.M., "Referential failures and affective reactivity of language in schizophrenia and unipolar depression", Schizophrenia Bulletin, vol. 37(3), pp. 554–560, 2011.

[7] Stasak, B., Epps, J. and Goecke, R., "Automatic depression classification based on affective read sentences: opportunities for text-dependent analysis", Speech Comm., vol. 115, pp. 1–14, 2019.

[8] Stasak, B., Epps, J., Schatten, H.T., Miller, I.W., Provost, E.M. and Armey, M.F., "Read speech voice quality and disfluency in individuals with recent suicidal ideation or suicide attempt", Speech Comm., vol. 132, pp.10–20, 2021.

[9] Stasak, B., Joachim, D. and Epps, J., "Breaking age barriers with automatic voice-based depression detection, IEEE Pervasive Computing, pp. 1–5, 2022.

[10] Bridges, K.A., "Prosody and formulaic language in treatment-resistant depression: effects of deep brain stimulation", Doctoral Thesis - New York University, 2014.

[11] Pope, B., Blass, T., Siegman, A.W., and Raher, J., "Anxiety and depression in speech", J. of Consulting and Clinical Psych., vol. 35(1), pp. 128–138, 1970.

[12] Stasak, B., Epps, J., and Cummins, N., "Depression prediction via acoustic analysis of formulaic word fillers", Polar, vol. 77(74), pp. 230–234, 2016.

[13] Roy, N., Merrill, R.M., Gray, S.D., and Smith, E.M., "Voice disorders in the general population: prevalence, risk factors, and occupational impact", The Laryngoscope, vol. 115(11), pp. 1988–1995, 2005.

[14] Cohen, A.S., McGovern, J.E., Dinzeo, T.J., and Covington, M.A., "Speech deficits in serious mental illness: a cognitive resource issue?", Schizophrenia Res., vol. 160(1-3), pp.173–179, 2014.

[15] Berto, V., "The relationship between perceptions of vocal quality and function with self-esteem in older adults", Theses and Dissertations - Illinois State University, pp. 1–65, 2018.

[16] Zuckerman, M. and Miyake, K., "The attractive voice: what makes it so?", J. Nonverbal Behavior, 17(2), pp. 119–135, 1993.

[17] Kim, J.Y., Ko, I., Kim, M.S., Yu, M.S., Cho, B.J., and Kim, D.K., "Association of chronic rhinosinusitis with depression and anxiety in a nationwide insurance population", JAMA Otolaryngology–Head & Neck Surgery, vol. 145(4), pp. 313–319, 2019.

[18] Barsties, B. and De Bodt, M., "Assessment of voice quality: current state-of-the-art", Auris Nasus Larynx, vol. 42(3), pp. 183–188, 2015.

[19] Yamaguchi, H., Shrivastav, R., Andrews, M.L., and Niimi, S., "A comparison of voice quality ratings made by Japanese and American listeners using the GRBAS scale", Folia Phoniatrica et Logopaedica, vol. 55(3), pp. 147–157, 2003.

[20] Oren, L., Kummer, A., and Boyce, S., "Understanding nasal emission during speech production: a review of types, terminology, and causality", Cleft Palate Craniofac. J., vol. 57(1), pp. 123–126, 2020.

[21] Pumariega, A.J., Rothe, E., and Pumariega, J.B., "Mental health of immigrants and refugees, Comm. Mental Health J., vol. 41(5), pp. 581–597, 2005.

[22] Coleman, K.J., et al., "Racial/ethnic differences in diagnosis and treatment of mental health conditions across healthcare system participants in the mental health research network", vol. 67(7), pp. 749–757, 2016

[23] Newman, S. and Mather, V.G., "Analysis of spoken language of patients with affective disorders", American J. of Psych., vol. 94(4), pp. 913–942, 1938.

[24] Stinchfield, S.M., "Speech disorders: a psychoanalytical study of the various defects in speech", New York, NY - USA, 1933.

[25] Alpert, M., Pouget, E.R., and Silva, R.R., "Reflections of depression in acoustic measures of the patient's speech", J. Affective Disorders, vol. 66(1), pp. 59–69, 2001.

[26] Gilboa-Schechtman, E., Galili, L., Sahar, Y., and Amir, O., "Being "in" or "out" of the game: subjective and acoustic reactions to exclusion and popularity in social anxiety", Frontiers in Human Neuroscience, vol. 8, pp. 147–160, 2014.

[27] Flint, A.J., Black, S.E., Campbell-Taylor, I., Gailey, G.F., and Levinton, C., "Abnormal speech articulation, psychomotor retardation, and subcortical dysfunction in major depression", J. Psychiatric Res., vol. 27(3), pp. 309–319, 1993.

[28] Kuny, S.T. and Stassen, H.H., "Speaking behavior and voice sound characteristics in depressive patients during recovery", J. of Psychiatric Res., vol. 27(3), pp. 289–307, 1993.

[29] Nilsonne, A., "Speech characteristics as indicators of depressive illness", Acta Psychiatrica Scandinavica, vol. 77, pp. 253–263, 1988.

[30] Kimble, C.E. and Seidel, S.D., "Vocal signs of confidence", J. Nonverbal Behavior, vol. 15(2), pp. 99–105, 1991.

[31] Goodell, S., Druss, B.G., Walker, E.R., and Mat, M.J.R.W.J.F.P., "Mental disorders and medical comorbidity", report, Robert Wood Johnson Foundation, vol. 2(21), pp. 1–6, 2011.

[32] Nutt, D., Wilson, S., and Paterson, L., "Sleep disorders as core symptoms of depression", Dialogues Clinical Neurosci., vol 10(3), pp. 329–336, 2008.

[33] Tovar, A., Fuentes-Claramonte, P., Soler-Vidal, J., Ramiro-Sousa, N., Rodriguez-Martinez, A., Sarri-Closa, C., Sarró, S., Larrubia, J., Andrés-Bergareche, H., Miguel-Cesma, M.C., and Padilla, P.P., "The linguistic signature of hallucinated voice talk in schizophrenia", Schizophrenia Res., vol. 206, pp. 111–117, 2019.

[34] Passchier-Vermeer ,W. and Passchier, W.F., "Noise exposure and public health", Environ. Health Perspect., vol. 108(1), pp. 123–131, 2000.

[35] Stansfeld, S.A. and Matheson, M.P., "Noise pollution: non-auditory effects on health", British Medical Bulletin, vol. 68(1), pp. 243–257, 2003.

[36] Letowski, T.R. and Scharine, A.A., "Correctional analysis of speech intelligibility tests and metrics for speech transmission", report, U.S. Army Research Laboratory, pp. 1–50, 2017.

[37] Werner-Seidler, A., Huckvale, K., Larsen, M.E. et al., "A trial protocol for the effectiveness of digital interventions for preventing depression in adolescents: the future proofing study", Trials, vol. 21(1), pp.1–21, 2020.