Measuring neuropsychiatric symptoms in patients with early cognitive decline using speech analysis

Authors

Alexandra König1,2,3, Elisa Mallick1,2,3, Johannes Tröger1,2,3, Nicklas Linz1,2,3, Radia Zeghari1,2,3, Valeria Manera1,2,3, Philippe Robert1,2,3

1 Institut national de recherche en informatique et en automatique (INRIA). Stars Team. Sophia Antipolis. Valbonne. France
2 ki:elements. Germany
3 CoBTeK (Cognition-Behaviour-Technology) Lab, FRIS-University Côte d’azur, Nice, France

Corresponding Author

Alexandra König. PhD
Institut national de recherche en informatique et en automatique (INRIA),
Stars Team,
2004 Route des Lucioles, 06902 Valbonne, France E-mail: alexandra.konig@inria.fr

Keywords

Neuropsychiatric symptoms; apathy; anxiety; depression; speech analysis; mild neurocognitive disorders; vocal parameters

This is an Open Access article, distributed under the terms of the Creative Commons Attribution- NonCommercial-NoDerivatives licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.
Abstract

Objectives: Certain neuropsychiatric symptoms (NPS), namely apathy, depression and anxiety demonstrated great value in predicting dementia progression, representing eventually an opportunity window for timely diagnosis and treatment. However, sensitive and objective markers of these symptoms are still missing. Therefore, the present study aims to investigate the association between automatically extracted speech features and NPS in patients with mild neurocognitive disorders.

Methods: Speech of 141 patients aged 65 or older with neurocognitive disorder was recorded while performing two short narrative speech tasks. NPS were assessed by the Neuropsychiatric Inventory (NPI). Paralinguistic markers relating to prosodic, formant, source, and temporal qualities of speech were automatically extracted, correlated with NPS. Machine learning experiments were carried out to validate the diagnostic power of extracted markers.

Results: Different speech variables are associated with specific neuropsychiatric symptoms; apathy correlates with temporal aspects, and anxiety with voice quality - and this was mostly consistent between male and female after correction for cognitive impairment. Machine learning regressors are able to extract information from speech features and perform above baseline in predicting anxiety, apathy and depression scores.

Conclusion: Different NPS seem to be characterized by distinct speech features which are easily extractable automatically from short vocal tasks. These findings support the use of speech analysis for detecting subtypes of NPS in patients with cognitive impairment. This could have great implications for the design of future clinical trials as this cost-effective method could allow more continuous and even remote monitoring of symptoms.
Introduction

The amount of patients with cognitive decline and thus, the risk to develop dementia continues to rise as the population ages [1]. Neuropsychiatric symptoms (NPS) can be defined as behavioural non-cognitive disturbances such as depression or agitation and are extremely common in dementia appearing in up to 98% of patients at one point during the disease [2]. It has been argued that NPS can appear relatively early even prior to dementia diagnosis, representing eventually an opportunity window for timely detection and treatment [3]. Since pathological changes seem to occur in the brain often before the onset of clinical dementia, the damages due to neurodegeneration may alter in a subtle manner emotion regulation and behaviour as cognition starts to decline [4]. This suggests that underlying neurobiological mechanisms might underpin the appearance of these symptoms [5] contributing to an increased vulnerability and risk for conversion [6].

NPS can cause a great burden for patients and caregivers affecting severely disease course and quality of life [7]. Interestingly, even in cognitively normal subjects it has been reported that certain types of NPS may represent promising prognostic factors for predicting cognitive deterioration, especially the affective ones such as depression, apathy and anxiety [8-10]. Anxiety seems to be an important predictor of amyloid pathology. The presence of apathy may increase the risk of conversion from Mild Cognitive Impairment (MCI) to dementia [11] and a similar association is found with depression [12]. Differential utilities of NPS for even predicting dementia subtypes are currently intensively investigated [13] since it could lead to earlier tailored treatment.

Despite the immense clinical impact of NPS, these clinical dimensions remain insufficiently identified and therefore untreated. Evaluation and management of NPS remains challenging since its detection is not made systematically in current clinical practice in patients with
cognitive decline [9], and these are often assessed only when family members bring their
observation to the attention of health care professionals. However, there is evidence that if
eye early treatment is provided, namely in the form of non-pharmacological intervention, it can be
effective to reduce the severity of NPS [7], and even to slow down cognitive decline [11].
Current assessment methods rely mostly on clinical scales such as the Neuropsychiatric
Inventory [14], which in turn depends heavily on the informants’ objective reporting ability
[15]. The clinician rating may be influenced as well by caregivers’ personal experiences, and
by limited access to reliable information. This is why clinicians and researchers are looking
for ways to complement classical assessment scales with more objective indexes, which are
able to detect subtle variations in the presentation of NPS as well as its different subtypes
[16,17].
Mobile technologies may be a promising solution for detecting and monitoring objectively
fine-grained changes in behaviour and thus, the appearance of NPS in a continuous manner
[18]. The use of digital biomarkers which are measures of behavioural data collected by
means of digital devices are increasingly studied in patients with cognitive decline as a
potential alternative assessment approach [19]. One promising avenue lies in recent advances
in computational linguistics and language processing that have led to the use of automatic
speech analysis in the assessment of various clinical manifestations [20,21] making it
potentially useful for assessing different NPS.
Until now, it has been clearly demonstrated that changes in speech and language patterns can
be indicative for cognitive decline due to neuropathological processes. Different affective
states including depression, anxiety and apathy alter as well mechanisms involved in speech
production, namely variations in muscle tension and tonality which can affect prosody and the
quality of speech. Reduced or increased muscle tension will influence vocal tract dynamics as
well as articulation behaviour [22] which may be today with recent advances in speech analysis technologies be easily detectable.

In depression, which is characterized by persistent sadness and anhedonia, accompanied by disturbed sleep/appetite, fatigue, and even suicidal ideation, it is notable by ear that patients show a reduced speech rate and prosody spectrum and sound rather monotonous which could serve as markers, if objective measurements can quantify these observations [23]. Until now, several groups investigated the use of automatic analysis of speech as an additional assessment tool with an extensive review published by Cummins et al. [22] outlining the interest of using speech as a key objective marker for disease progression.

Prosodic, articulatory, and acoustic features of speech seem affected by depression severity and thus can easily be identified and used for continuously monitoring patients. With a considerable overlap of symptoms between depression and apathy, namely the lack of interest and goal-oriented behavior, we anticipate similar results when applying speech technology methods to apathy with a slightly different pattern in regards to emotionally triggered speech.

In anxiety, which is characterized by excessive fear or worrying, many studies found a significant increase in mean fundamental frequency (f0) [24]. Jitter and shimmer were also significantly higher in anxious patients [25,26]. In a previous study, we showed that certain speech characteristics extracted over the phone correlated with stress levels in both genders; mainly, spectral (ie, formant) features, such as the mel-frequency cepstral coefficient, and prosodic characteristics, such as the fundamental frequency, appeared to be sensitive to stress [27].

In apathy, which is characterized by lack of motivation, decreased initiative, and emotional indifference, we demonstrated that certain paralinguistic features, namely temporal aspects of
speech as well as prosodic characteristics correlate significantly with levels of symptom
severity [28].

Hence, given the increasingly important role of NPS in early dementia diagnosis [29], it
seems worthwhile investigating the additional value of this method further for differential
diagnosis of these most frequently found NPI symptoms. Therefore, the purpose of this study
is to determine whether automatic speech analysis and the automatic extraction of speech
features can be useful for the early detection of specific NPS - namely depression, apathy and
anxiety - in patients with mild neurocognitive disorders.

Materials and Methods

Participants

All participants were recruited through the Memory Clinic located at the Institut Claude
Pompidou in the Nice University Hospital. 141 patients aged 65 or older with mild
neurocognitive disorder according to the Diagnostic Statistical Manual 5 (DSM-5) [30] were
included in this study. For this, the presence of cognitive decline in memory and/or executive
function with or without interference with performance in activity of daily living was required
based on previously performed evaluations. Patients coming to the Memory clinic underwent
a clinical assessment including, among others, the Mini-Mental State Examination (MMSE)
[31] and the Neuropsychiatric Inventory interview (NPI) [14].

Study procedure

Speech features vary naturally between males and females. These differences have been
leveraged in gender classification through speech analysis based on pitch and formant
frequencies [32], Harmonic- to-Noise ratio [33], and linear predictive components and mel
frequency cepstral coefficients (MFCC) [34]. Previous work found differences in speech
depending on gender in the effects of apathy [35], as well as depression and the effectiveness of classifiers for its detection [36]. This is why this study considers males and females separately resulting in two experimental groups consisting of 92 females and 49 males.

Participants were all native speakers of French and excluded if they had any major auditory or language problems, history of head trauma, loss of consciousness, psychotic or aberrant motor behavior, or history of drug abuse. Written informed consent was obtained from all subjects prior to the experiments. The study was approved by the Nice Ethics Committee (ELEMENT ID RCB 2017-A01896-45, MoTap ID RCB 2017-A01366-47) and was conducted according to the Declaration of Helsinki.

Speech task (positive and negative story)

Free and natural speech tasks are capable of eliciting emotional reactions (or a lack thereof) by asking to describe events that triggered recent affective arousal. Affective arousal is expected to impact speech in NPS since simple vocal exercises or reading tasks do not allow to capture the acoustic effects of alterations in affective states [22]. Using emotional induced free speech tasks allows for a greater range of emotional effects such as describing events that have aroused significant emotions [37]. Therefore, the participants were asked to: 1) talk about a positive event in their life and 2) to talk about a negative event in their life. Instructions for the speech tasks (“Can you tell me in one minute about a positive/negative event in your life?”) were pre-recorded by one of the psychologists and played from a tablet computer ensuring standardized instruction over both experiments. The answers were recorded with the tablet’s internal microphone.
Audio features were extracted directly and automatically from the speech signal. This form of speech analysis does not consider the semantic content of what a participant said, thus increasing the applicability of results in a clinical scenario, as no prior processing, such as transcription of what has been said, is required. For each speech task (positive and negative story), features were extracted separately from different main areas: temporal features including measures of speech proportion (e.g. length of pauses, length of speaking segments), the connectivity of speech segments and general speaking rate; prosodic, relating to long-time variations in perceived stress and rhythm in speech. Prosodic features also measure alterations in personal speaking style (e.g. perceived pitch, intonation of speech); formant features represent the dominant components of the speech spectrum and carry information about the acoustic resonance of the vocal tract and its use. These markers are often indicative of problems with articulatory coordination in speech motor control disorders (ref Sapir); source features relate to the source of voice production, the airflow through the glottal speech production system. Spectral features characterise the speech spectrum; the frequency distribution of the speech signal at a specific time instance information in some high dimensional representation. These features operationalize irregularities in vocal fold movement (e.g. measures of voice quality). An overview and explanation of the extracted speech features can be found in table 1.
Statistical analysis was run using R software (software version 3.4.02). QQ-plots as well as Shapiro-Wilk tests indicated a violation of the normal distribution for the depression scores. Therefore, we used Spearman rank correlations to test whether speech as well as transcript data was correlated with NPI scores (only for the items apathy, depression and anxiety). We applied the Benjamini - Hochberg procedure for each of the feature categories when correcting for multiple comparisons. Three variable groups were excluded from this analysis. MFCCs, Delta and Delta Delta values were not considered in group comparisons, as they do not have directly human understandable explanations and would therefore provide little insight. These variables were included in the machine learning experiments, as explainability is not required.

To predict NPI scores, we train different regression models and evaluate their performance using MAE. Regression models are trained including Support Vector Regression (SVR) and Lasso (Linear Regression with L1 regularisation). Implementations were provided by the scikit-learn python framework. Features were normalised by subtraction of their mean and division through their standard deviation. Because of the small data set size a separate validation/test set could not be used. Instead, Leave-One-Out-Cross-Validation was employed. This is a method where N different models are evaluated: train a model on N-1 observations and test it on 1 observation. The model is evaluated for every held-out observation. The final result is then calculated by taking the mean of all the individual evaluations.
Results

141 participants, of which 92 were female and 49 were male, have been included in the analysis. Demographic and clinical information split by gender is available in Table 2. Both groups had an average age around 75 and showed on average some signs of mild cognitive impairment, with a mean MMSE of around 24. The male group scored significantly higher on the apathy domain of the NPI. Correlations between the NPI subscales and the MMSE showed significant correlations for both groups (see Table 3). MMSE was significantly correlated with the NPI apathy subscale for both groups, while only females also showed significant correlations between the anxiety subscale and the MMSE. Both normal and partial correlations, corrected for MMSE, of speech variables and NPI subscales by gender group and speech task are displayed in Figure 1 (a detailed listing of correlations is provided in the supplementary materials; Table S1 & S2).

The Sound to noise ratio shows significant correlations with the NPI anxiety subscale for both genders and tasks, suggesting that the higher the sound to noise ratio, the stronger symptoms of anxiety are present. This effect is stable after correcting for the influence of MMSE. Before and after correcting for MMSE, the Speech Ratio shows significant correlations with the NPI apathy subscale for both genders and across both tasks, indicating that less speaking during the recording is associated with higher levels of apathy. The same is visible for the Total Phonation Time and Number of Pauses, but effects are not sustained after correcting for MMSE using partial correlations. The Amplitude Kurtosis shows significant correlations with the anxiety and apathy NPI subscales, which are not retained in the partial correlation.

Correlations between all NPI scales and measures of power (Mean power and Total power) are visible for females in the negative story. When correcting for the influence of MMSE,
only weaker correlations with the anxiety and depression subscales remain. Performance of trained machine learning models are listed in Table 4 as the MAE of predictions. The baseline for each subscale and gender is reported. For each subscale and gender, a model could be trained that outperforms this baseline by at least 0.3 points on the NPI scale. Features selected by the machine learning models are available in the supplementary material (Table S4).

Discussion

Identifying early signs of different NPS by the means of objective measurement tools could ultimately lead to better management and treatment. This is particularly important in patients with mild neurocognitive disorders. Indeed, the assessment of NPS is not systematically performed in clinical practice, despite the fact that they represent risk factors for dementia conversion, and thus are increasingly becoming targets of clinical trials [38,39]. As in the early stages of neurocognitive disorders NPS do not interfere substantially with activities of daily living, patients and caregivers may not report them unless prompted, which suggests the need for non-invasive, objective and reliable measures that can help detect subtle changes in NPS.

The results of the present study show that certain changes in spontaneous speech characteristics seem associated with specific NPS. Namely, apathy correlates with temporal aspects of speech such as speech ratio, which means that patients with more severe apathy, speak slower and less. This is consistent with the diagnostic criteria for apathy [40], which identified reduction of self-initiated verbal production as one of the core apathy dimensions. Similarly, we found previously strong correlations between this type of features (sound duration, syllable count, etc.) and subdomains of the apathy inventory [28].
In turn, anxiety correlates with voice quality features such as Sound to noise ratio. These results are even mostly consistent between male and females. Importantly, these correlations between apathy and anxiety and the distinct speech variables were still significant after correcting for cognitive impairment, meaning that these specific speech features may be relevant for assessing NPS in patients with different degrees of cognitive deterioration, and do not capture simply cognitive decline. As the presence of NPS increases with cognitive decline progression [29], it is important to control for the degree of cognitive decline to verify if speech features are relevant specifically to detect NPS. The fact that we did not find any strong correlation between speech features and depression might be due to the small variance on the depression NPI subscale in our sample.

Our findings are consistent with other research in which a significant increase in mean f0 (prosodic related feature) was detected in anxiety disorders [41–43]; however, we found this effect only in females during the negative storytelling. In another study, in which emotional responses were induced by watching videos, females showed higher emotional expressivity for negative stimuli, whereas males showed overall more intense emotional experiences, meaning that the differences in gender seem to depend on the specific emotion type [44].

Moreover, greater F0 has been reported mostly in males with social anxiety disorder compared to controls and only in females during in vivo social exposure [43]. This could explain why we found a difference only in the negative task since it may elicit a more emotionally loaded reaction. Jitter and shimmer were also significantly higher in patients with an elevated score on the NPI anxiety subdomain which is in line with previous studies [25,26].
Machine learning regressors were able to extract information from speech features and perform above baseline in predicting anxiety, apathy and depression scores, despite NPS were globally of quite low intensity in our sample. When adding MMSE as a feature, only the model predicting apathy heavily relied on it, which may be partially due to the fact that apathy becomes more common with increased cognitive decline [45]. These results may have important clinical implications since for each subscale and gender, a model could be trained that outperforms this baseline by at least 0.3 points on the NPI scale.

The positive and negative storytelling were chosen as semi-standardised free speech tasks, that still allow for production of natural speech, while limiting the scope/time and induce an emotional reaction in participants. This in contrast to related work measuring cognitive function from similar populations, that relied on more standardised tasks, such as describing an image or completing a verbal cognitive task.

Similar studies have shown as well that computed acoustic features can eventually be used to track changes in mental health states [46]. Indeed, speech analytics have been tested successfully in psychiatry for its use to predict treatment response in depressed patients [47] or changes in mood states in bipolar disorder [48]. As certain feature types seem to be associated with NPS, the outcome of this study may encourage a potential broader use of automatic speech analysis in other neurocognitive disorders for NPS detection. This could particularly improve the objectivity of first first front-line assessment procedures and help overcome barriers that prevent timely access diagnosis and treatment.

In parallel to classical methods, such technologies can be used to establish objective, personalized baseline reference standards to design innovative clinical trials that assess the effectiveness of new treatments [49]. For instance, the use of wearable devices for additional measurements such as heart rate variability or locomotor activity are increasingly investigated...
for detecting behavioural abnormalities [50]. Combined with speech all these data could help reduce evaluation biases and provide richer understanding of variations in NPS on a day-to-day basis even before severity reaches a level of requiring intervention [51,52]. In regards to the new rise of decentralized clinical trials, digital technologies will play a major role to complement traditional data acquisition enabling the redesign and adaptation of trials while still ongoing [53].

One limitation of our study is that patients in the sample showed a lot of variability in the MMSE score even if they all met the diagnosis criteria for mild cognitive disorder which could have affected the results. In addition, patients had no severe NPS - rarely they scored above 4 which is the original cut-off. However, taking into account the predictive value of behavioural symptoms [54], it is therefore important to investigate if the technology is able to detect subtle signs of symptoms since this may be even more of interest for early prevention strategies.

Another limitation is that we employed only the NPI scale for the assessment of NPS. Despite the fact that it is widely used and considered as a gold standard for neuropsychiatric assessment, the NPI is probably not the ideal scale for early screening, and to quantify fine changes. Recently, the Mild behavioural Inventory [55] has been proposed, that could better capture early signs of NPS, and thus may show stronger correlations with speech features in patients with mild NPS symptoms, as it is often the case for people with mild neurocognitive disorders. In addition, a larger study including more patients and more severe NPS symptoms should be replicated in order to validate the found speech markers.
Acknowledgement

The authors would like to thank all participants of this study. This research is part of the MNC3 program of the University Cote d’Azur IDEX Jedi.

Financial Support

It was partially funded by the MEPHESTO project (Grant DRI-0120105/291), the EIT Digital Well-being Activity (Grant nr. 17074), ELEMENT, the University Côte d’Azur, and by the IA association and supported by the Edmond & Lily Safra Foundation and Institute Claude Pompidou.

Conflicts of Interest

EM, JT and NL are employees and shareholders of ki elements UG.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.
References

[1] Satizabal C, Beiser AS, Seshadri S. Incidence of Dementia over Three Decades in the Framingham Heart Study. N Engl J Med 2016;375:93–4.

[2] Phan SV, Osae S, Morgan JC, Inyang M, Fagan SC. Neuropsychiatric Symptoms in Dementia: Considerations for Pharmacotherapy in the USA. Drugs R D 2019;19:93–115.

[3] Gallagher D, Fischer CE, Iaboni A. Neuropsychiatric Symptoms in Mild Cognitive Impairment. Can J Psychiatry 2017;62:161–9.

[4] Iaboni A, Rapoport MJ. Detecting and Managing Neuropsychiatric Symptoms in Dementia. Can J Psychiatry 2017;62:158–60.

[5] Livingston G, Sommerlad A, Orgeta V, Costafreda SG, Huntley J, Ames D, et al. Dementia prevention, intervention, and care. Lancet 2017;390:2673–734.

[6] Liew TM. Neuropsychiatric symptoms in cognitively normal older persons, and the association with Alzheimer’s and non-Alzheimer’s dementia. Alzheimer’s Research & Therapy 2020;12. https://doi.org/10.1186/s13195-020-00604-7.

[7] Radue R, Walaszek A, Asthana S. Neuropsychiatric symptoms in dementia. Handb Clin Neurol 2019;167:437–54.

[8] Hu M, Shu X, Wu X, Chen F, Hu H, Zhang J, et al. Neuropsychiatric symptoms as prognostic makers for the elderly with mild cognitive impairment: a meta-analysis. J Affect Disord 2020;271:185–92.

[9] Gitlin LN, Marx KA, Stanley IH, Hansen BR, Van Haitsma KS. Assessing neuropsychiatric symptoms in people with dementia: a systematic review of measures. Int Psychogeriatr 2014;26:1805–48.

[10] Jang JY, Ho JK, Blanken AE, Dutt S, Nation DA; Alzheimer’s Disease Neuroimaging Initiative. Affective Neuropsychiatric Symptoms as Early Signs of Dementia Risk in Older Adults. J Alzheimers Dis. 2020;77(3):1195-1207.

[11] Padala PR, Padala KP, Lensing SY, Jackson AN, Hunter CR, Parkes CM, et al. Repetitive transcranial magnetic stimulation for apathy in mild cognitive impairment: A double-blind, randomized, sham-controlled, cross-over pilot study. Psychiatry Res 2018;261:312–8.

[12] Goukasian N, Hwang KS, Romero T, Grotts J, Do TM, Groh JR, Bateman DR, Apostolova LG. Association of brain amyloidosis with the incidence and frequency of neuropsychiatric symptoms in ADNI: a multisite observational cohort study. BMJ Open. 2019 Dec 18;9(12):e031947.
[13] Liew, T.M. Neuropsychiatric symptoms in cognitively normal older persons, and the association with Alzheimer’s and non-Alzheimer’s dementia. Alz Res Therapy 12, 35 (2020).

[14] Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J. The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia. Neurology 1994;44:2308–14.

[15] Stella F. Assessment of neuropsychiatric symptoms in dementia: toward improving accuracy. Dement Neuropsychol 2013;7:244–51.

[16] König A, Aalten P, Verhey F, Bensadoun G, Petit P-D, Robert P, et al. A review of current information and communication technologies: can they be used to assess apathy? Int J Geriatr Psychiatry 2014;29:345–58.

[17] Gros A, Bensamoun D, Manera V, Fabre R, Zacconi-Cauvin A-M, Thummler S, et al. Recommendations for the Use of ICT in Elderly Populations with Affective Disorders. Front Aging Neurosci 2016;8:269.

[18] Kourtis LC, Regele OB, Wright JM, Jones GB. Digital biomarkers for Alzheimer’s disease: the mobile/wearable devices opportunity. Npj Digital Medicine 2019;2.

[19] Piau A, Wild K, Mattek N, Kaye J. Current State of Digital Biomarker Technologies for Real-Life, Home-Based Monitoring of Cognitive Function for Mild Cognitive Impairment to Mild Alzheimer Disease and Implications for Clinical Care: Systematic Review. J Med Internet Res 2019;21:e12785.

[20] 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:e0248842.

[21] Robin J, Harrison JE, Kaufman LD, Rudzicz F, Simpson W, Yancheva M. Evaluation of Speech-Based Digital Biomarkers: Review and Recommendations. Digit Biomark 2020;4:99–108.

[22] Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. Speech Communication 2015;71:10–49. https://doi.org/10.1016/j.specom.2015.03.004.

[23] Lanctot KL, et al. Neuropsychiatric signs and symptoms of Alzheimer’s disease: new treatment paradigms. Alzheimers Dement (N Y). 2017;3(3):440–9.
Low DM, Bentley KH, Ghosh SS. Automated assessment of psychiatric disorders using speech: A systematic review. Laryngoscope Investig Otolaryngol 2020;5:96–116.

Özseven T, Düğenci M, Doruk A, Kahraman Hİ. Voice traces of anxiety: acoustic parameters affected by anxiety disorder. Arch Acoust 2018;43(4):625-636.

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 2016. https://doi.org/10.21437/speechprosody.2016-249.

König A, Riviere K, Linz N, Lindsay H, Elbaum J, Fabre R, et al. Measuring Stress in Health Professionals Over the Phone Using Automatic Speech Analysis During the COVID-19 Pandemic: Observational Pilot Study. J Med Internet Res 2021;23:e24191.

König A, Linz N, Zeghari R, Klinge X, Tröger J, Alexandersson J, et al. Detecting Apathy in Older Adults with Cognitive Disorders Using Automatic Speech Analysis. Journal of Alzheimer’s Disease 2019;69:1183–93. https://doi.org/10.3233/jad-181033.

Cummings J, Ritter A, Rothenberg K. Advances in Management of Neuropsychiatric Syndromes in Neurodegenerative Diseases. Curr Psychiatry Rep 2019;21:79.

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). American Psychiatric Pub; 2013.

Folstein MF, Folstein SE, McHugh PR. Mini-mental State: A Practical Method for Grading the Cognitive State of Patients for the Clinician. 1975.

Childers DG, Wu K. Gender recognition from speech. Part II: Fine analysis. J Acoust Soc Am 1991;90:1841–56.

Heffernan K. Evidence from HNR that /s/ is a social marker of gender. Toronto Working Papers in Linguistics 2004:23.

Wu K, Childers DG. Gender recognition from speech. Part I: Coarse analysis. J Acoust Soc Am 1991;90:1828–40.

Linz N, Klinge X, Troger J, Alexandersson J, Zeghari R, Robert P, König A. Automatic detection of apathy using acoustic markers extracted from free emotional speech. Nd Workshop on AI for Ageing, Rehabilitation and Independent Assisted Living (ARIAL) 2018:17–21.

Low L-SA, Maddage MC, Lech M, Sheeber LB, Allen NB. Detection of Clinical Depression in Adolescents’ Speech During Family Interactions. IEEE Transactions on Biomedical Engineering 2011;58:574–86. https://doi.org/10.1109/tbme.2010.2091640.
[37] Gupta, R., Malandrakis, N., Xiao, B., Guha, T., Van Segbroeck, M., Black, M., Potamianos, A., Narayanan, S., 2014. Multimodal prediction of affective dimensions and depression in human–computer interactions. Proceedings of the 4th ACM International Workshop on Audio/Visual Emotion Challenge (AVEC ’14). ACM, Orlando, Florida, USA, pp. 33–40.

[38] Roberto N, Portella MJ, Marquié M, Alegret M, Hernández I, Mauleón A, et al. Neuropsychiatric profiles and conversion to dementia in mild cognitive impairment, a latent class analysis. Sci Rep 2021;11:6448.

[39] Miller DS, Robert P, Ereshefsky L, Adler L, Bateman D, Cummings J, et al. Diagnostic criteria for apathy in neurocognitive disorders. Alzheimers Dement 2021. https://doi.org/10.1002/alz.12358.

[40] Robert P, Lanctôt KL, Agüera-Ortiz L, Aalten P, Bremond F, Defrancesco M, et al. Is it time to revise the diagnostic criteria for apathy in brain disorders? The 2018 international consensus group. Eur Psychiatry 2018;54:71–6.

[41] Weeks JW, Lee C-Y, Reilly AR, Howell AN, France C, Kowalsky JM, et al. “The Sound of Fear”: assessing vocal fundamental frequency as a physiological indicator of social anxiety disorder. J Anxiety Disord 2012;26:811–22.

[42] Galili L, Amir O, Gilboa-Schechtman E. Acoustic Properties of Dominance and Request Utterances in Social Anxiety. Journal of Social and Clinical Psychology 2013;32:651–73. https://doi.org/10.1521/jscp.2013.32.6.651.

[43] Weeks JW, Srivastav A, Howell AN, Menatti AR. “Speaking More than Words”: Classifying Men with Social Anxiety Disorder via Vocal Acoustic Analyses of Diagnostic Interviews. Journal of Psychopathology and Behavioral Assessment 2016;38:30–41. https://doi.org/10.1007/s10862-015-9495-9.

[44] Deng Y, Chang L, Yang M, Huo M, Zhou R. Gender Differences in Emotional Response: Inconsistency between Experience and Expressivity. PLoS One 2016;11:e0158666.

[45] Manera V, Fabre R, Stella F, Loureiro JC, Agüera-Ortiz L, López-Álvarez J, et al. A survey on the prevalence of apathy in elderly people referred to specialized memory centers. Int J Geriatr Psychiatry 2019;34:1369–77.

[46] Arevian AC, Bone D, Malandrakis N, Martinez VR, Wells KB, Miklowitz DJ, et al. Clinical state tracking in serious mental illness through computational analysis of speech. PLoS One 2020;15:e0225695.
Carrillo F, Sigman M, Fernández Slezak D, Ashton P, Fitzgerald L, Stroud J, et al. Natural speech algorithm applied to baseline interview data can predict which patients will respond to psilocybin for treatment-resistant depression. J Affect Disord 2018;230:84–6.

Gideon J, Provost EM, McInnis M. Mood state prediction from speech of varying acoustic quality for individuals with bipolar disorder. Proc IEEE Int Conf Acoust Speech Signal Process 2016;2016:2359–63.

Teipel S, König A, Hoey J, Kaye J, Krüger F, Robillard JM, et al. Use of nonintrusive sensor-based information and communication technology for real-world evidence for clinical trials in dementia. Alzheimers Dement 2018;14:1216–31.

Reinertsen E, Clifford GD. A review of physiological and behavioral monitoring with digital sensors for neuropsychiatric illnesses. Physiol Meas 2018;39:05TR01.

Karow A, Pajonk F-G, Reimer J, Hirdes F, Osterwald C, Naber D, et al. The dilemma of insight into illness in schizophrenia: self- and expert-rated insight and quality of life. Eur Arch Psychiatry Clin Neurosci 2008;258:152–9.

Copeland LA, Zeber JE, Salloum IM, Pincus HA, Fine MJ, Kilbourne AM. Treatment adherence and illness insight in veterans with bipolar disorder. J Nerv Ment Dis 2008;196:16–21.

Chow S-C. Adaptive Clinical Trial Design. Annual Review of Medicine 2014;65:405–15. https://doi.org/10.1146/annurev-med-092012-112310.

van Dalen JW, van Wanrooj LL, Moll van Charante EP, Brayne C, van Gool WA, Richard E. Association of Apathy With Risk of Incident Dementia: A Systematic Review and Meta-analysis. JAMA Psychiatry 2018;75:1012–21.

Taragano FE, Allegri RF, Heisecke SL, Martelli MI, Feldman ML, Sánchez V, et al. Risk of Conversion to Dementia in a Mild Behavioral Impairment Group Compared to a Psychiatric Group and to a Mild Cognitive Impairment Group. J Alzheimers Dis 2018;62:227–38.
### TABLE 1. Overview and explanation of extracted speech features.

| Speech features          | Explanation                                                                                     | Category     |
|--------------------------|-------------------------------------------------------------------------------------------------|--------------|
| Speech ratio             | Prosodic feature for verbal fluency from speech to non-speech proportion in the audio          | Temporal     |
| Speech interval          | Speech segments uninterrupted by pauses between syllables to measure speech production efficiency | Temporal     |
| Total phonation time     | Total time duration of all words across all sentences                                             | Temporal     |
| Number of pauses         | Total number of pauses longer than a time threshold between the syllables                         | Temporal     |
| Harmonics to noise ratio | Ratio between periodic and non-periodic components of speech which reflects voice quality       | Source       |
| Sound to noise ratio     | Ratio between the power of speech signal and the power of background noise that reflects voice quality | Source       |
| Jitter                   | Jitter is a measure of random perturbation in signal periodicity More representative when examining long vowels. Jitter calculation is based on relative jitter. | Source       |
| Espinola zero crossing metric | Measure for the rate at which speech signal crosses the zero reference and its deviation from the reference | Source       |
| Average amplitude change | Variation of the signal amplitude over time                                                      | Source       |
| Amplitude mean absolute value | Variation of the signal amplitude over time without change direction                           | Source       |
| Parameter                          | Description                                                                                                                                                                                                 | Source                  |
|-----------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------|
| Amplitude third moment            | Skewness of the signal amplitude over time                                                                                                                                                                  | Source                  |
| Amplitude fourth moment           | Kurtosis of the signal amplitude over time                                                                                                                                                                  | Source                  |
| Max amplitude                     | Measure for the maximum disturbance of the air caused by the speech signal                                                                                                                                   | Source                  |
| Mean power                        | Mean power transmitted by the signal                                                                                                                                                                         | Source                  |
| Total power                       | Total power transmitted by the signal                                                                                                                                                                         | Source                  |
| F₀ mean                           | Mean value of the fundamental frequency in voiced parts of the audio. Fundamental frequency or pitch quantifies speech signal’s periodic components for speech production assessment | Prosodic                |
| F₀ standard deviation             | Standard deviation of fundamental frequency in voiced parts of the audio                                                                                                                                      | Prosodic                |
| Average Mel-frequency cepstrals (MFCC) | Decomposition of Mel-frequency cepstral (MFCC) into a range of spectrum coefficients 1 through 20 MFC represents the short-term power spectrum of a sound                                                                 | Spectral                |
| Deltas                            | First derivative of the average MFCC values that presents the change in the power spectrum                                                                                                                   | Spectral                |
| Delta deltas                      | First derivative of the average MFCC values that presents the rate of change in the power spectrum                                                                                                           | Spectral                |
| Peak frequency                    | Frequency of the maximum power value in signal frequency spectrum                                                                                                                                              | Spectral                |
| Power spectrum ratio              | Power of the most powerful frequency relative to all other frequencies                                                                                                                                         | Spectral                |
TABLE 2. Demographic data for included participants, split by gender. Mean and standard deviation are reported.

|                  | Females            | Males             |
|------------------|--------------------|-------------------|
| N                | 92                 | 49                |
| Age              | 75.01 (7.88)       | 76.08 (6.75)      |
| NPI anxiety      | 1.84 (2.64)        | 1.86 (2.67)       |
| NPI depression   | 0.85 (1.90)        | 0.84 (1.69)       |
| NPI apathy       | 1.64 (2.74)        | 3.37 (3.81)       |
| NPI total        | 7.10 (9.23)        | 9.82 (8.96)       |
| MMSE             | 24.01 (4.95)       | 23.69 (4.34)      |
**TABLE 3.** Spearman rank correlations between MMSE and NPI subscales for females and males.

| NPI score | Females | | | Males | | |
|-----------|---------|---|---|------|---|---|
|           | Correlation | p value | Correlation | p value |
| Anxiety   | -0.335 | < 0.01 | -0.268 | > 0.05 |
| Depression| -0.148 | > 0.05 | -0.163 | > 0.05 |
| Apathy    | -0.584 | < 0.001 | -0.486 | < 0.001 |
TABLE 4. Mean Absolute Error of regression methods (Linear Regression L1 penalization and SVM) and of the baseline, for males and females separately. Results better than baseline are marked in bold.

|                | Male                      | Female                     |
|----------------|---------------------------|----------------------------|
|                | Baseline | Linear Regression L1 penalization | SVM | Baseline | Linear Regression L1 penalization | SVM |
| NPI Anxiety    | 2.05232 | 2.05232 | **1.68202** | 2.17857 | 2.22331 | **1.73591** |
| NPI Depression | 1.26708 | 1.26708 | **0.89701** | 1.08078 | 1.08078 | **0.86314** |
| NPI Apathy     | 2.13903 | 1.84431 | **1.66386** | 3.22024 | **2.84111** | 3.37315 |

**Figures**

FIGURE 1. Plot between (left) spearman rank correlations and (right) spearman rank partial correlations corrected for MMSE, between audio features and NPI subscales, separated by gender and voice task. Only significant correlations are reported. Absolute value of correlation is reflected in the size and color (positive correlations in blue; negative correlations in red) of the dot.
TABLE S1. Top 10 highest spearman rank correlations between audio features calculated per speech task and NPI subscales. Results are split up by gender.

(a) (Female; Anxiety)

| Feature                          | Correlation | p value | Feature                      | Correlation | p value |
|----------------------------------|-------------|---------|------------------------------|-------------|---------|
| sound_to_noise_ratio             | 0.3469641   | 0.0007031 | peak_frequency               | -0.4433038  | 9.622e-06 |
| peak_frequency                   | -0.3012164  | 0.003526  | amplitude_mav                | -0.4230408  | 2.655e-05 |
| amplitude_kurtosis               | 0.2851256   | 0.005872  | total_power                  | -0.4223768  | 2.742e-05 |
| amplitude_mav                    | -0.2809288  | 0.006676  | mean_power                   | -0.4223768  | 2.742e-05 |
| total_power                      | -0.2690185  | 0.009513  | amplitude_fourth_moment      | -0.3832605  | 0.0001627 |
| mean_power                       | -0.2690185  | 0.009513  | max_amplitude                | -0.3171703  | 0.002066 |
| amplitude_fourth_moment          | -0.2468373  | 0.01769   | sound_to_noise_ratio         | 0.2830272   | 0.006262 |
| speech_ratio                     | -0.2232307  | 0.03244   | mean_f0                      | -0.2781462  | 0.007262 |
| max_amplitude                    | -0.1895974  | 0.07027   | speech_ratio                 | -0.273535   | 0.008332 |
| total_phonation_time             | -0.1268491  | 0.2282   | amplitude_kurtosis           | 0.258846    | 0.01272  |
| Feature                | Correlation | p value | Feature                | Correlation | p value |
|-----------------------|-------------|---------|-----------------------|-------------|---------|
| sound_to_noise_ratio  | 0.3396247   | 0.01697 | amplitude_kurtosis    | 0.3433922   | 0.01571 |
| espinola_zero_crossing_metric | -0.1831235 | 0.2079  | sound_to_noise_ratio  | 0.3172379   | 0.02635 |
| total_phonation_time  | -0.1097501  | 0.4528  | amplitude_mav         | -0.2490399  | 0.08442 |
| power_spectrum_ratio  | 0.1057096   | 0.4697  | speech_ratio          | -0.2163333  | 0.1354  |
| total_power           | -0.09555357 | 0.5137  | number_of_pauses      | -0.2162034  | 0.1357  |
| mean_power            | -0.09555357 | 0.5137  | mean_f0               | -0.1889312  | 0.1936  |
| speech_ratio          | 0.09511676  | 0.5156  | peak_frequency        | -0.1708498  | 0.2405  |
| mean_f0               | -0.07978293 | 0.5858  | total_power           | -0.164516   | 0.2586  |
| local_absolute_jitter | 0.06655989  | 0.6495  | mean_power            | -0.164516   | 0.2586  |
| average_amplitude_change | -0.06585006 | 0.653   | total_phonation_time  | -0.159711   | 0.273   |
| Feature                  | Correlation | p value | Feature                  | Correlation | p value |
|-------------------------|-------------|---------|-------------------------|-------------|---------|
| peak_frequency          | -0.2291068  | 0.02804 | max_amplitude           | -0.292862   | 0.004611|
| amplitude_fourth_moment | -0.2163193  | 0.03835 | peak_frequency          | -0.2623844  | 0.01151 |
| total_power             | -0.2116438  | 0.04284 | amplitude_fourth_moment | -0.260786   | 0.01205 |
| mean_power              | -0.2116438  | 0.04284 | total_power             | -0.2544122  | 0.01439 |
| amplitude_mav           | -0.1882766  | 0.07229 | mean_power              | -0.2544122  | 0.01439 |
| max_amplitude           | -0.1748197  | 0.09556 | amplitude_mav           | -0.2281278  | 0.02873 |
| amplitude_third_moment  | -0.1557183  | 0.1383  | local_jitter            | 0.1580161   | 0.1325  |
| sound_to_noise_ratio    | 0.1425711   | 0.1752  | ppq5_jitter             | 0.1381954   | 0.1889  |
| mean_f0                 | -0.1225804  | 0.2444  | rap_jitter              | 0.1316518   | 0.211   |
| espinola_zero_crossing_metric | -0.0778772 | 0.4606  | ddp_jitter              | 0.1316518   | 0.211   |
(d) (Male; Depression)

| Positive story                              | Negative story                              |
|---------------------------------------------|---------------------------------------------|
| **Feature**                                 | **Feature**                                 |
| **Correlation**  | **p value**  | **Correlation**  | **p value** |
| average_amplitude_change  | 0.2030018  | 0.1618  | average_amplitude_change  | 0.2699959  | 0.06063 |
| harmonics_to_noise_ratio  | 0.1820364  | 0.2106  | amplitude_fourth_moment  | 0.1450946  | 0.3199 |
| local_jitter  | -0.1503211  | 0.3026  | peak_frequency  | 0.1181306  | 0.4189 |
| ppq5_jitter  | -0.1214566  | 0.4058  | local_absolute_jitter  | -0.1035796  | 0.4788 |
| local_absolute_jitter  | -0.1152798  | 0.4303  | mean_f0  | 0.09952806  | 0.4962 |
| rap_jitter  | -0.1086873  | 0.4573  | power_spectrum_ratio  | 0.0969277  | 0.5076 |
| ddp_jitter  | -0.1086873  | 0.4573  | local_jitter  | 0.08112944  | 0.5795 |
| amplitude_fourth_moment  | 0.1041735  | 0.4763  | ppq5_jitter  | 0.08053552  | 0.5823 |
| total_phonation_time  | -0.09591804  | 0.5121  | total_power  | 0.07394301  | 0.6136 |
| max_amplitude  | 0.09455202  | 0.5181  | mean_power  | 0.07394301  | 0.6136 |
(Female; Apathy)

| Feature                  | Correlation | p value   | Feature                  | Correlation | p value   |
|--------------------------|-------------|-----------|--------------------------|-------------|-----------|
| speech_ratio             | -0.4649382  | 3.024e-06 | speech_ratio             | -0.4422672  | 1.015e-05 |
| amplitude_kurtosis       | 0.4122244   | 4.447e-05 | number_of_pauses         | -0.3738093  | 0.0002421 |
| number_of_pauses         | -0.4005278  | 7.617e-05 | amplitude_mav            | -0.3416138  | 0.0008601 |
| total_phonation_time     | -0.3871873  | 0.0001374 | total_phonation_time     | -0.3212951  | 0.001791  |
| amplitude_mav            | -0.2895547  | 0.005117  | total_power              | -0.3045195  | 0.003164  |
| peak_frequency           | -0.2145304  | 0.04002   | mean_power               | -0.3045195  | 0.003164  |
| total_power              | -0.199052   | 0.05714   | harmonics_to_noise_ratio | -0.3018382  | 0.003455  |
| mean_power               | -0.199052   | 0.05714   | peak_frequency           | -0.2815369  | 0.006554  |
| harmonics_to_noise_ratio | -0.1656924  | 0.1145    | amplitude_kurtosis       | 0.2753473   | 0.007896  |
| amplitude_fourth_moment  | -0.123723   | 0.24      | amplitude_fourth_moment  | -0.250406   | 0.01606   |
| Feature               | Correlation  | p value  | Feature               | Correlation  | p value  |
|-----------------------|--------------|----------|-----------------------|--------------|----------|
| speech_ratio          | -0.5011488   | 0.0002446| speech_ratio          | -0.5768886   | 1.433e-05|
| number_of_pause       | -0.3865379   | 0.00608  | amplitude_kurtosis    | 0.455276     | 0.001013 |
| amplitude_kurtosis    | 0.3169775    | 0.02648  | number_of_pause       | -0.4269726   | 0.002216 |
| total_phonation_time  | -0.3050516   | 0.03306  | total_phonation_time  | -0.3781761   | 0.00738  |
| amplitude_mav         | -0.2994025   | 0.03662  | amplitude_mav         | -0.2946426   | 0.03987  |
| peak_frequency        | -0.2489267   | 0.08457  | peak_frequency        | -0.2666586   | 0.06401  |
| total_power           | -0.2485083   | 0.08511  | total_power           | -0.2493975   | 0.08396  |
| mean_power            | -0.2485083   | 0.08511  | mean_power            | -0.2493975   | 0.08396  |
| espinola_zero_crossing_metric | -0.1929972 | 0.184    | local_jitter         | 0.1978233    | 0.173    |
| mean_f0               | 0.1854857    | 0.202    | mean_f0              | 0.181538     | 0.2119   |
TABLE S2. Top 10 highest spearman rank partial correlations between audio features calculated per speech task and NPI subscales, corrected for MMSE. Results are split up by gender.

(a) (Female; Anxiety)

| Feature                   | Correlation | p value | Feature                   | Correlation | p value |
|---------------------------|-------------|---------|---------------------------|-------------|---------|
| sound_to_noise_ratio      | 0.33073     | 0.00136 | mean_f0                  | -0.36523    | 0.00037 |
| peak_frequency            | -0.20172    | 0.05518 | peak_frequency            | -0.35242    | 0.00061 |
| amplitude_fourth_moment   | -0.17364    | 0.09975 | total_power               | -0.32664    | 0.00158 |
| total_power               | -0.17349    | 0.10004 | mean_power                | -0.32664    | 0.00158 |
| mean_power                | -0.17349    | 0.10004 | amplitude_mav             | -0.32015    | 0.00198 |
| amplitude_kurtosis        | 0.16728     | 0.11298 | amplitude_fourth_moment   | -0.29205    | 0.00497 |
| amplitude_mav             | -0.16361    | 0.12124 | local_absolute_jitter     | 0.28813     | 0.00561 |
| mean_f0                   | -0.12298    | 0.24548 | sound_to_noise_ratio      | 0.27374     | 0.00865 |
| local_absolute_jitter     | 0.11589     | 0.27400 | max_amplitude             | -0.23753    | 0.02339 |
| max_amplitude             | -0.11202    | 0.29043 | ppq5_jitter               | 0.23121     | 0.02744 |
(b) (Male; Anxiety)

| Feature                        | Correlation | p value | Feature                        | Correlation | p value |
|--------------------------------|-------------|---------|--------------------------------|-------------|---------|
| sound_to_noise_ratio           | 0.39219     | 0.00583 | sound_to_noise_ratio           | 0.40730     | 0.00406 |
| espinola_zero_crossing_metric  | -0.17880    | 0.22400 | amplitude_kurtosis             | 0.27846     | 0.05531 |
| power_spectrum_ratio           | 0.13428     | 0.36285 | mean_f0                        | -0.23137    | 0.11358 |
| speech_ratio                   | 0.13293     | 0.36776 | amplitude_mav                  | -0.22492    | 0.12430 |
| ppq5_jitter                    | -0.11695    | 0.42858 | ppq5_jitter                    | -0.17848    | 0.22484 |
| total_power                    | -0.10299    | 0.48604 | peak_frequency                 | -0.16178    | 0.27196 |
| mean_power                     | -0.10299    | 0.48604 | number_of_pauses               | -0.15929    | 0.27950 |
| mean_f0                        | -0.10192    | 0.49062 | total_power                    | -0.15637    | 0.28852 |
| average_amplitude_change       | -0.08493    | 0.56600 | mean_power                     | -0.15637    | 0.28852 |
| amplitude_fourth_moment        | -0.07194    | 0.62704 | speech_ratio                   | -0.14818    | 0.31482 |
| Feature                  | Correlation | p value | Feature                  | Correlation | p value |
|-------------------------|-------------|---------|-------------------------|-------------|---------|
| peak_frequency          | -0.18970    | 0.07170 | max_amplitude           | -0.26266    | 0.01189 |
| amplitude_fourth_moment | -0.18571    | 0.07800 | amplitude_fourth_moment | -0.22337    | 0.03331 |
| total_power             | -0.17313    | 0.10076 | peak_frequency           | -0.22293    | 0.03367 |
| mean_power              | -0.17313    | 0.10076 | total_power              | -0.21372    | 0.04193 |
| max_amplitude           | -0.14265    | 0.17737 | mean_power               | -0.21372    | 0.04193 |
| amplitude_mav           | -0.14058    | 0.18381 | amplitude_mav           | -0.18209    | 0.08408 |
| amplitude_third_moment  | -0.13400    | 0.20540 | amplitude_third_moment  | -0.17726    | 0.09277 |
| mean_f0                 | -0.12836    | 0.22530 | local_jitter            | 0.15316     | 0.14722 |
| sound_to_noise_ratio    | 0.12835     | 0.22533 | local_absolute_jitter   | 0.14442     | 0.17200 |
| average_amplitude_change| -0.09983    | 0.34644 | ppq5_jitter             | 0.13458     | 0.20342 |
| Feature               | Correlation  | p value  | Feature               | Correlation  | p value |
|-----------------------|--------------|----------|-----------------------|--------------|---------|
| harmonics_to_noise_ratio | 0.22161      | 0.13008   | average_amplitude_change | 0.28733      | 0.04768 |
| average_amplitude_change     | 0.19633      | 0.18108   | amplitude_fourth_moment | 0.14453      | 0.32705 |
| local_jitter             | -0.17628     | 0.23070   | local_absolute_jitter | -0.13237     | 0.36979 |
| ppq5_jitter             | -0.15358     | 0.29733   | peak_frequency        | 0.12938      | 0.38080 |
| rap_jitter              | -0.13196     | 0.37127   | speech_ratio         | 0.11710      | 0.42801 |
| ddp_jitter              | -0.13196     | 0.37127   | power_spectrum_ratio | 0.10835      | 0.46352 |
| local_absolute_jitter   | -0.12744     | 0.38802   | number_of.pause       | 0.10633      | 0.47196 |
| amplitude_fourth_moment | 0.09941      | 0.50142   | amplitude_kurtosis    | -0.09845     | 0.50561 |
| power_spectrum_ratio    | 0.09347      | 0.52743   | sound_to_noise_ratio  | 0.09809      | 0.50713 |
| max_amplitude           | 0.09334      | 0.52804   | total_power           | 0.08377      | 0.57134 |
| Feature                  | Correlation | p value | Feature                  | Correlation | p value |
|--------------------------|-------------|---------|--------------------------|-------------|---------|
| amplitude_kurtosis       | 0.22271     | 0.03385 | harmonics_to_noise_ratio | -0.24608    | 0.01871 |
| speech_ratio             | -0.19707    | 0.06116 | speech_ratio             | -0.22413    | 0.03270 |
| total_phonation_time     | -0.17709    | 0.09311 | ppq5_jitter              | 0.18561     | 0.07816 |
| number_of_pauses         | -0.16877    | 0.10978 | rap_jitter               | 0.17773     | 0.09191 |
| sound_to_noise_ratio     | -0.16037    | 0.12889 | ddp_jitter               | 0.17773     | 0.09191 |
| harmonics_to_noise_ratio | -0.12995    | 0.21955 | local_jitter             | 0.16416     | 0.11998 |
| mean_f0                  | 0.10509     | 0.32151 | sound_to_noise_ratio     | -0.13812    | 0.19168 |
| amplitude_mav            | -0.05878    | 0.57995 | number_of_pauses         | -0.13435    | 0.20419 |
| ppq5_jitter              | 0.05321     | 0.61643 | espinola_zero_crossing_metric | -0.11075 | 0.29597 |
| power_spectrum_ratio     | -0.05298    | 0.61795 | total_phonation_time     | -0.10475    | 0.32307 |
| Feature                  | Correlation | p value | Feature                  | Correlation | p value |
|-------------------------|-------------|---------|-------------------------|-------------|---------|
| speech_ratio            | -0.51070    | 0.00021 | speech_ratio            | -0.51766    | 0.00016 |
| amplitude_mav           | -0.33569    | 0.01967 | number_of_pauses        | -0.35981    | 0.01201 |
| number_of_pauses        | -0.30012    | 0.03821 | total_phonation_time    | -0.35587    | 0.01305 |
| total_power             | -0.29197    | 0.04405 | amplitude_kurtosis      | 0.35308     | 0.01383 |
| mean_power              | -0.29197    | 0.04405 | peak_frequency          | -0.27395    | 0.05954 |
| peak_frequency          | -0.27980    | 0.05409 | amplitude_mav           | -0.26857    | 0.06493 |
| amplitude_kurtosis      | 0.26409     | 0.06971 | total_power             | -0.25654    | 0.07840 |
| total_phonation_time    | -0.22924    | 0.11704 | mean_power              | -0.25654    | 0.07840 |
| amplitude_fourth_moment | -0.20050    | 0.17180 | amplitude_fourth_moment | -0.19747    | 0.17852 |
| average_amplitude_change| -0.19865    | 0.17588 | amplitude_third_moment  | -0.17273    | 0.24038 |
**TABLE S3.** Top 5 highest feature weights selected by SVM regression models in predicting each NPI subscale, separated for females and males.

(a) Anxiety

| Feature       | Task  | Mean  | Standard Deviation |
|---------------|-------|-------|--------------------|
| MFCC 18       | Neg   | 0.0192| 0.00076            |
| MFCC 4        | Neg   | 0.0188| 0.00061            |
| Sound-to-noise ratio | Pos | 0.0182| 0.00070            |
| Deltas 7      | Pos   | 0.0178| 0.00083            |
| Delta Deltas 17 | Neg | 0.0172| 0.00086            |

| Feature       | Task  | Mean  | Standard Deviation |
|---------------|-------|-------|--------------------|
| MFCC 9        | Neg   | 0.0170| 0.00050            |
| MFCC 4        | Pos   | 0.0154| 0.00060            |
| MFCC 10       | Pos   | 0.0146| 0.00051            |
| MFCC 16       | Neg   | 0.0143| 0.00075            |
| MFCC 6        | Neg   | 0.0141| 0.00067            |
## Depression

| Feature          | Task | Mean      | Standard Deviation |
|------------------|------|-----------|--------------------|
| Delta Deltas 11  | Neg  | 0.0014    | 5.13151e-05        |
| Deltas 8         | Pos  | 0.0013    | 6.30781e-05        |
| Deltas 9         | Pos  | 0.0011    | 5.65088e-05        |
| Deltas 16        | Neg  | 0.0009    | 5.16828e-05        |
| Deltas 17        | Neg  | 0.0009    | 6.49522e-05        |

| Feature          | Task     | Mean | Standard Deviation |
|------------------|----------|------|--------------------|
| Delta Deltas 5   | Neg      | 0.0009 | 6.46563e-05        |
| Delta Deltas 20  | Neg      | 0.0009 | 9.55432e-05        |
| Amplitude kurtosis | Neg    | 0.0008 | 3.10521e-05        |
| Deltas 1         | Neg      | 0.0007 | 7.12870e-05        |
| Delta deltas 14  | Pos      | 0.0007 | 7.22985e-05        |
| Feature                | Task | Mean   | Standard Deviation |
|------------------------|------|--------|--------------------|
| MMSE                   | 0.0035 | 0.00011 |
| Speech ratio Pos       | 0.0021   | 0.00011 |
| Speech ratio Neg       | 0.0019       | 0.00012 |
| Harmonics-to-noise ratio Neg | 0.0018       | 9.59043e-05 |
| Deltas 2 Pos           | 0.0017         | 8.55444e-05 |
| Deltas 5 Neg           | 0.1289       | 0.09128 |

| Male Feature            | Task | Mean   | Standard Deviation |
|-------------------------|------|--------|--------------------|
| Speech ratio            | Neg  | 0.2203 | 0.09936 |
| MMSE                    |      | 0.1910 | 0.09611 |
| Speech ratio Pos        | Pos  | 0.1774 | 0.04123 |
| Deltas 5 Neg            |      | 0.1289 | 0.09128 |
| Number of pauses        | Neg  | 0.1187 | 0.02705 |