

Task Matters: A Methodological Comparison of Speech Elicitation Types for Machine Learning Classification of Psychiatric Disorders and Controls

Submitter Felix Menne

Affiliation ki:elements GmbH, Saarbrücken, Germany

SUBMISSION DETAILS

I agree to provide poster pdf for attendee download. Yes

I have used the poster abstract template to develop my abstract. Yes

Methodological Issue Being Addressed Digital speech biomarkers are increasingly investigated as objective measures of psychiatric disorders such as major depressive disorder (MDD) and schizophrenia (SZ). However, which types of speech elicitation tasks best capture the signal necessary for diagnostic differentiation remains to be fully established. Determining the most informative task type is essential for optimizing data collection and minimizing participant burden in digital assessments.

Introduction Advances in computational linguistics and acoustic signal analysis have made it possible to quantify subtle speech abnormalities associated with psychiatric disorders. Building on these developments, this study compares four types of speech tasks: positive, neutral, and negative autobiographical recall, and a structured picture description (Boston Cookie Theft). The aim is to determine which task provides the strongest discriminative signal for classifying individuals with MDD, SZ, and healthy controls (HC) to inform future methodological standards in speech-based biomarker research.

Methods A total of 66 participants were included: 22 with MDD, 22 with SZ, and 22 healthy controls recruited from the Karl-Jaspers Clinic of Psychiatry, University Hospital Oldenburg, Germany. Each participant completed four elicitation tasks: description of a positive, a neutral, and a negative autobiographical event, and a description of the Boston Cookie Theft picture. Speech recordings were processed to extract 92 acoustic (e.g., prosody, formant, spectral) and linguistic (e.g., syntactic complexity, lexical richness, sentiment) features. Multiple machine learning classifiers (Decision Trees, Extra Trees, Support Vector Machines, Linear Models) were trained in pairwise diagnostic comparisons. Classification performance was evaluated via cross-validation, and receiver operating characteristic (ROC) curves were generated for each task to compare discriminative ability across conditions.

Results In the HC vs. MDD comparison, all four tasks produced identical outcomes, with a Decision Tree classifier achieving an AUC of 0.96, sensitivity of 0.91, and specificity of 1.00. Each model used ten features.

A similar pattern was observed in the SZ vs. MDD comparison, where Decision Trees again performed best for the autobiographical recall tasks (AUC = 0.96, sensitivity = 0.91, specificity = 1.00). Performance for the picture description task was slightly lower (AUC = 0.91, sensitivity =

0.77, specificity = 0.96) and best modeled with a Random Forest.

In the HC vs. SZ comparison, Linear Models performed best, with AUCs ranging from 0.74 (picture description) to 0.83 (negative recall). Sensitivity ranged from 0.59 to 0.77, specificity from 0.64 to 0.82, and the number of selected features varied between 10 and 44.

Conclusion For MDD-related contrasts, consistent classification performance across autobiographical recall tasks indicates a stable, largely task-invariant diagnostic speech signal. In contrast, lower and more variable performance in the schizophrenia comparisons suggests greater task dependence. Methodologically, these results show that high discriminative accuracy can be achieved with minimal task standardization for depression, while task optimization may be more critical for schizophrenia and picture description tasks. These findings refine methodological approaches for speech-based digital phenotyping in psychiatric research.

Co-Authors

Felix Menne¹, Felix Dörr¹, Johannes Tröger¹, Alexandra König¹,
Diana Immel², Simon Barton², René Hurlemann²

¹ ki:elements GmbH, Saarbrücken, Germany

² Dept. of Psychiatry at Karl-Jaspers Clinic, School of Medicine & Health Sciences, Carl von Ossietzky University of Oldenburg, Germany

Keywords

Keywords

schizophrenia

machine learning

task design

speech

major depressive disorder

Guidelines I have read and understand the Poster Guidelines

Disclosures FM, FD, JT, AK are employees of ki:elements GmbH, Germany. JT is also a shareholder of the company.