Motivation and Research Question

- Schizophrenia is a mental disease that causes hallucinations, delusions, and disordered thinking
- Speech and oro-facial biomarkers are promising for remote assessment and monitoring

Goal: combine speech and facial biomarkers into one composite index score
- Useful as an endpoint in clinical practice & pharmaceutical trials
- Better noise robustness and statistical power than multiple individual markers
- Maintain interpretability of clinically meaningful metrics

Research Question:
Given a large, multicollinear feature set from remote audiovisual assessments, how can we determine an interpretable composite index score for remote monitoring of Schizophrenia?

Data and Feature Selection

- Multimodal dialogue platform used to collect audiovisual data (illustration Fig. 1); sessions were overseen by a psychiatrist
- Speech assessments included: diadochokinesis (DDK), reading passage (RP), picture description (PD), spontaneous speech (S)
- Clinician administered rating scales for people with Schizophrenia (pSz): PANSS, BNSS, CGI-S, AIMS, SAS, BARS
- Patient eligibility: Inpatients with diagnosis of schizophrenia, age 18-60, English speaking, WRAT-IV Reading Score ≥ 8th grade, (pSz): PANSS, BNSS, CGI-S, AIMS, SAS, BARS
- Healthy control eligibility:
  - Multimodal dialogue platform
  - Healthy controls (HC) – future work will explore broader use case

Table 1: Demographics. BNSS ranges from 0 to 78, PANSS Positive & Negative range from 7 to 100, and age 18-60, English speaking, WRAT-IV Reading Score ≥ 8th grade, (pSz): PANSS, BNSS, CGI-S, AIMS, SAS, BARS.

Table 2: Selected features. CTA: canonical timing alignment, CPP: cepstral peak prominence, JC: jaw center, LL: lower lip.

Table 3: Methods to compute an index score as (weighted) linear combination of features. Features were inverted by taking (1- scaled feature) when median value was smaller in Control cohort in the train set. LDA: Linear Discriminant Analysis

Index Score Computation

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>Linear combination with equal weights</td>
</tr>
<tr>
<td>LDA</td>
<td>Linear combination that maximizes area under ROC curve (AUC) Caveat: assumption of Gaussian distributions</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>Logistic regression coefficients as weights for linear combination, L1 regularization enforces sparse weight vector</td>
</tr>
<tr>
<td>Constrained log. reg.</td>
<td>Logistic regression coefficients as weights, constrained to be non-negative</td>
</tr>
</tbody>
</table>

Table 3: Methods to compute an index score as (weighted) linear combination of features. Features were inverted by taking (1-scaled feature) when median value was smaller in Control cohort in the train set. LDA: Linear Discriminant Analysis

Results

- Index scores yield better test results than individual metrics with reduced variability
- All methods >80% UAR
- Weak to moderate negative correlations between index scores and negative symptoms (-0.33 (p=0.001) b/w cLogReg and BNSS total, -0.37 (p=0.001) b/w cLogReg and PANSS Negative total)

Figure 1. Modality.AI dialogue platform.

Figure 2. Classification accuracy for individual features and for the different index scores, for the binary classification pSz vs. controls. Error bars represent standard deviation across validation folds for 5-fold cross validation. LogReg: logistic regression, cLogReg: constrained logistic regression.

Figure 3. Normalized feature weights (constrained log. reg.) expressed as ranges that result from the variation across validation folds. Blue dots represent the weights from one representative validation fold.

Conclusions

- Proposed a method to combine speech, oro-facial, and linguistic features into one composite index → potential endpoint for trials
- Index scores improve classification accuracy and reduce variation within cross validation
- Weighted linear combinations maintain interpretability
- For differentiating pSz from HC, speech features are most dominant

Limitations and future work:
- Variation in feature weights depending on training data
- Index tailored to specific task (classify pSz and healthy controls) – future work will explore broader use case

Key findings:
- Proposed a method to combine speech, oro-facial, and linguistic features into one composite index → potential endpoint for trials
- Index scores improve classification accuracy and reduce variation within cross validation
- Weighted linear combinations maintain interpretability
- For differentiating pSz from HC, speech features are most dominant

Limitations and future work:
- Variation in feature weights depending on training data
- Index tailored to specific task (classify pSz and healthy controls) – future work will explore broader use case

Table 2: Selected features. CTA: canonical timing alignment, CPP: cepstral peak prominence, JC: jaw center, LL: lower lip.