A psychometric evaluation of digital speech-based markers of negative symptom severity in schizophrenia spectrum disorders

Michael J. Spilka[1], Jessica Robin[1], Amir H. Nikzad[2], Leily Behbehani[2], Sarah Berretta[2], Mengdan Xu[1], John M. Kane[2], & Sunny X. Tang[2]

[1]Cambridge Cognition, Cambridge, UK
[2]Institute of Behavioral Science, Feinstein Institutes for Medical Research, Zucker Hillside Hospital, Northwell Health, NY, USA

Contact: michael.spilka@camcog.com

Background

Digital speech assessment and analysis have the potential to enhance the assessment of negative symptoms in schizophrenia spectrum disorders (SSD).

However, the large number of available speech features and variability in results across studies make it difficult to determine the features with the greatest clinical utility.

Objective: to identify the most robust speech-based markers of negative symptom severity by evaluating their reliability and validity in a sample of participants with SSD.

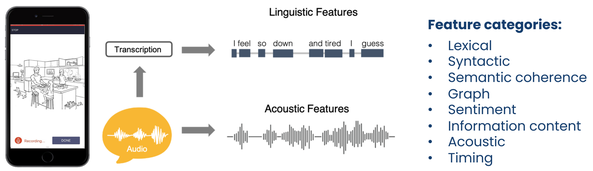
Methods

Participants: 62 inpatients with an SSD diagnosis from a longitudinal study of acute psychosis (baseline visit at inpatient admission and follow-up visit upon discharge ≈1-3 weeks later).

Clinical assessments: Scale for the Assessment of Negative Symptoms (SANS), Brief Psychiatric Rating Scale (BPRS), Extrapyramidal Symptom Rating Scale (ESRS).

Speech assessment: 2 picture description and 2 journaling tasks at each visit (using the Winterlight iOS app).

Quantitative speech features: 38 relevant features (based on prior findings and theory) quantifying acoustic and linguistic properties were extracted for each participant from their speech recordings and accompanying transcripts using acoustic signal processing and Natural Language Processing (NLP) tools.



Feature categories:
- Lexical
- Syntactic
- Semantic coherence
- Graph
- Sentiment
- Information content
- Acoustic
- Timing

Analyses: Speech features were evaluated for:

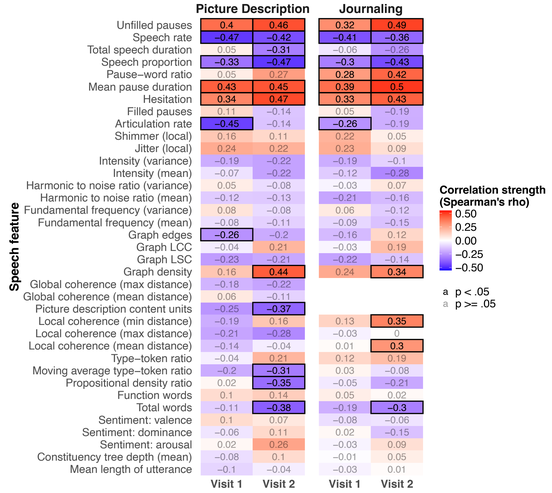
1) Associations with negative symptom severity (SANS Total) at baseline (Spearman correlations adjusting for age and sex).
2) Test-retest reliability when comparing stimuli within tasks (intraclass correlations: mean-rating consistency ICC).
3) Replicability of significant negative symptom correlations and between-stimulus test-retest reliability at the follow-up visit.
4) Convergent validity (associations with BPRS Negative), discriminant validity (no association with BPRS Positive), and specificity for negative symptom severity (no association with BPRS Total).
5) Relationships with clinical confounds (antipsychotic dose, extrapyramidal symptoms [ESRS Akathisia & Parkinsonism]) and demographics (age, sex, education).
6) Interrelationships among identified speech features.

Participant characteristics

	Visit 1 (Baseline) *n* = 62	Visit 2 (Discharge) *n* = 48
Age (*M*, *SD*)	26.4 (5.3)	26.2 (4.6)
Sex (*n*)	female = 20, male = 42	female = 15, male = 33
BPRS Total (*M*, *SD*)	48.2 (10.9)	43.1 (13.1)
BPRS Positive (*M*, *SD*)	15.9 (4.4)	13.3 (5.5)
BPRS Negative (*M*, *SD*)	5.6 (3.0)	6.4 (3.3)
SANS Total (*M*, *SD*)	28.2 (12.2)	26.1 (13.3)

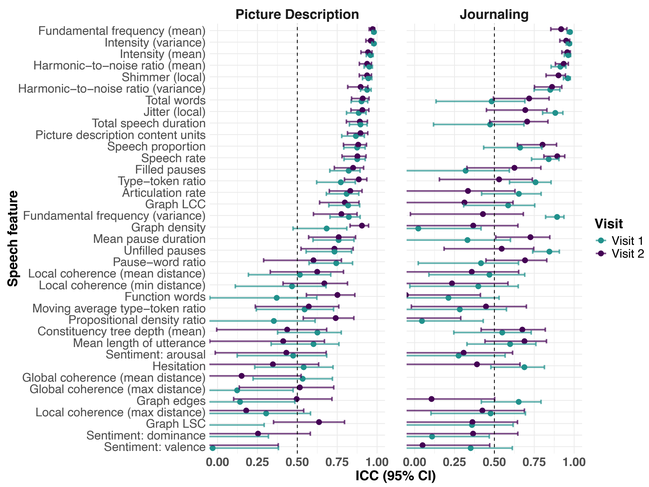
1) Speech feature correlations with negative symptoms (SANS Total)

Negative symptoms were predominantly correlated with speech timing-related features, as well as several linguistic features.



2) Test-retest reliability (within visits)

Reliability between task stimuli within visits was highest for acoustic, timing, and language production features (e.g., total words).



Conclusions

3 journaling and 4 picture description speech features consistently demonstrated significant associations with negative symptom severity and acceptable or higher test-retest reliability: ↑ **mean pause duration,** ↓ **speech proportion,** ↓ **speech rate, and** ↑ **unfilled pauses**.

Speech rate consistently demonstrated convergent and discriminant validity, and specificity for negative symptom severity. **Results suggest that speech rate may serve as a robust speech-based digital biomarker of negative symptom severity in SSD.**

3) Replicability of negative symptom correlations and test-retest reliability

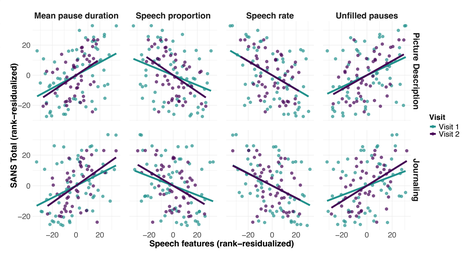
4 speech timing-related features from the Picture Description task had replicable correlations with negative symptom severity and acceptable or above (ICC > 0.5) test-retest reliability between task stimuli.

3 of these features met the same criteria for the Journaling task.

Feature	Correlation with SANS Total		Reliability (ICC)	
Picture Description				
	Visit 1	Visit 2	Visit 1	Visit 2
Mean pause duration	**0.43**	**0.45**	**0.76**	**0.76**
Speech proportion	**-0.33**	**-0.47**	**0.87**	**0.88**
Speech rate	**-0.47**	**-0.42**	**0.88**	**0.88**
Unfilled pauses	**0.40**	**0.46**	**0.73**	**0.73**
Journaling				
	Visit 1	Visit 2	Visit 1	Visit 2
Mean pause duration	0.39	0.50	0.33	0.73
Speech proportion	**-0.30**	**-0.43**	**0.66**	**0.80**
Speech rate	**-0.41**	**-0.37**	**0.84**	**0.89**
Unfilled pauses	**0.32**	**0.49**	**0.84**	**0.55**

Note. Features with significant correlations and ICC > 0.5 at both visits are indicated in **bold**. Positive correlations are in red and negative correlations are in blue.

Speech correlations with SANS Total for identified features



4) Convergent validity, discriminant validity, and specificity for negative symptoms

The 4 features were not correlated with positive symptom severity (discriminant validity) and were not correlated with overall clinical severity (specificity for negative symptoms).

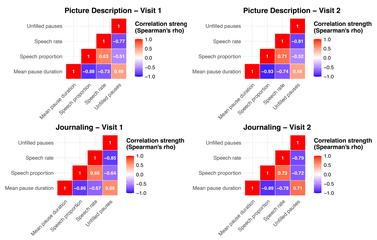
Speech rate most consistently demonstrated convergent validity (correlations with BPRS Negative score) across tasks and visits.

Feature	Convergent Validity (BPRS Negative)		Discriminant validity (BPRS Positive)		Specificity (BPRS Total)	
Picture Description						
	Visit 1	Visit 2	Visit 1	Visit 2	Visit 1	Visit 2
Mean pause duration	ns	0.48	ns	ns	ns	ns
Speech proportion	ns	-0.46	ns	ns	ns	ns
Speech rate	**-0.37**	**-0.38**	**ns**	**ns**	**ns**	**ns**
Unfilled pauses	0.32	0.29	ns	ns	ns	ns
Journaling						
	Visit 1	Visit 2	Visit 1	Visit 2	Visit 1	Visit 2
Mean pause duration	ns	0.53	ns	ns	ns	ns
Speech proportion	ns	-0.41	ns	ns	ns	ns
Speech rate	**-0.37**	**-0.51**	**ns**	**ns**	**ns**	**ns**
Unfilled pauses	ns	0.44	ns	ns	ns	ns

Note. Features demonstrating significant convergent validity at both visits are indicated in **bold**. Positive correlations are in red and negative correlations are in blue.

6) Speech feature intercorrelations

Mean pause duration and speech proportion appeared redundant (*rho* > 0.8).



5) Relationships with clinical confounds & demographic characteristics

The 4 features were not correlated with antipsychotic dose, extrapyramidal symptoms, or age. Features were most consistently correlated with sex and inconsistently correlated with education level.

Feature	Antipsychotic dose		ESRS Akathisia		ESRS Parkinsonism		Age		Sex		Education	
Picture Description												
	Visit 1	Visit 2	Visit 1	Visit 2	Visit 1	Visit 2	Visit 1	Visit 2	Visit 1	Visit 2	Visit 1	Visit 2
Mean pause duration	ns	ns	ns	ns	ns	ns	ns	ns	ns	-0.52	ns	ns
Speech proportion	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.49	ns	ns
Speech rate	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.42	ns	ns
Unfilled pauses	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
Journaling												
	Visit 1	Visit 2	Visit 1	Visit 2	Visit 1	Visit 2	Visit 1	Visit 2	Visit 1	Visit 2	Visit 1	Visit 2
Mean pause duration	ns	ns	ns	ns	ns	ns	ns	ns	ns	-0.38	ns	-0.32
Speech proportion	ns	ns	ns	ns	ns	ns	ns	ns	0.26	0.35	ns	0.27
Speech rate	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.42	ns	0.46
Unfilled pauses	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	-0.40

Note. Positive correlation values for sex indicates that higher features values are associated with female sex.

Disclosures: MJS, JR, & MX are employees of Cambridge Cognition. SXT owns equity and/or has consulted for North Shore Therapeutics, Psyrin, and Winterlight Labs. This work was funded by Winterlight Labs (acquired by Cambridge Cognition).

Feinstein Institute for Medical Research
Northwell Health

CAMBRIDGE COGNITION