Optimizing automated objective speech technologies for measuring negative symptoms

ALEX S COHEN
Louisiana State University
Dept. of Psychology
Center for Computation & Technology
Quantic Innovations

Mark Opler
WCG Clinical
Quantic Innovations

Brian Kirkpatrick
Psychiatric Research Institute
University of Arkansas for Medical Sciences
Quantic Innovations

Gregory Strauss
University of Georgia, Department of Psychology
Quantic Innovations

Seth Hopkins
Sumitomo Pharma America

INTRODUCTION

Automated and objective speech analysis is a promising approach to measuring negative symptoms that can complement traditional clinical ratings. With respect to clinical trials, it can potentially be used as an exploratory endpoint and can screen participants for study suitability.

We evaluated an automated, objective, face-valid and computationally transparent tool for measuring negative symptoms. We evaluated

- the reliability of key speech features with each other [Internal consistency],
- the reliability of speech across different speaking tasks [Between task reliability],
- convergence with clinical ratings.

METHODS OVERVIEW

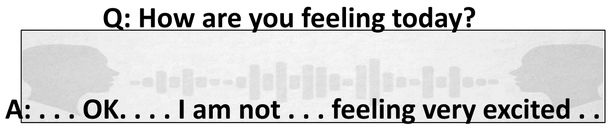
Archived audio recordings from nine separate data collections for 19 different speaking tasks were processed using transcription, speaker recognition and timestamping procedures. There were 6 different speaking tasks.

Patients were transdiagnostic: with mostly schizophrenia, and some unipolar and bipolar diagnoses, and some substance use diagnoses.

Patients were binarized based on expressive negative symptoms being present (i.e., mild or greater; n = 259 people; k = 871 recordings for analysis) or absent (i.e., absent or questionable; n = 473 people; k = 2296 recordings for analysis).

SPEECH FEATURES

Q: How are you feeling today?



A: . . . OK. . . . I am not . . . feeling very excited . . . about much.

1. Initial Turn Latency
2. Within Turn Latency
3. Between Word Latency
4. Turn Length
5. Words per Turn
6. Articulation Rate

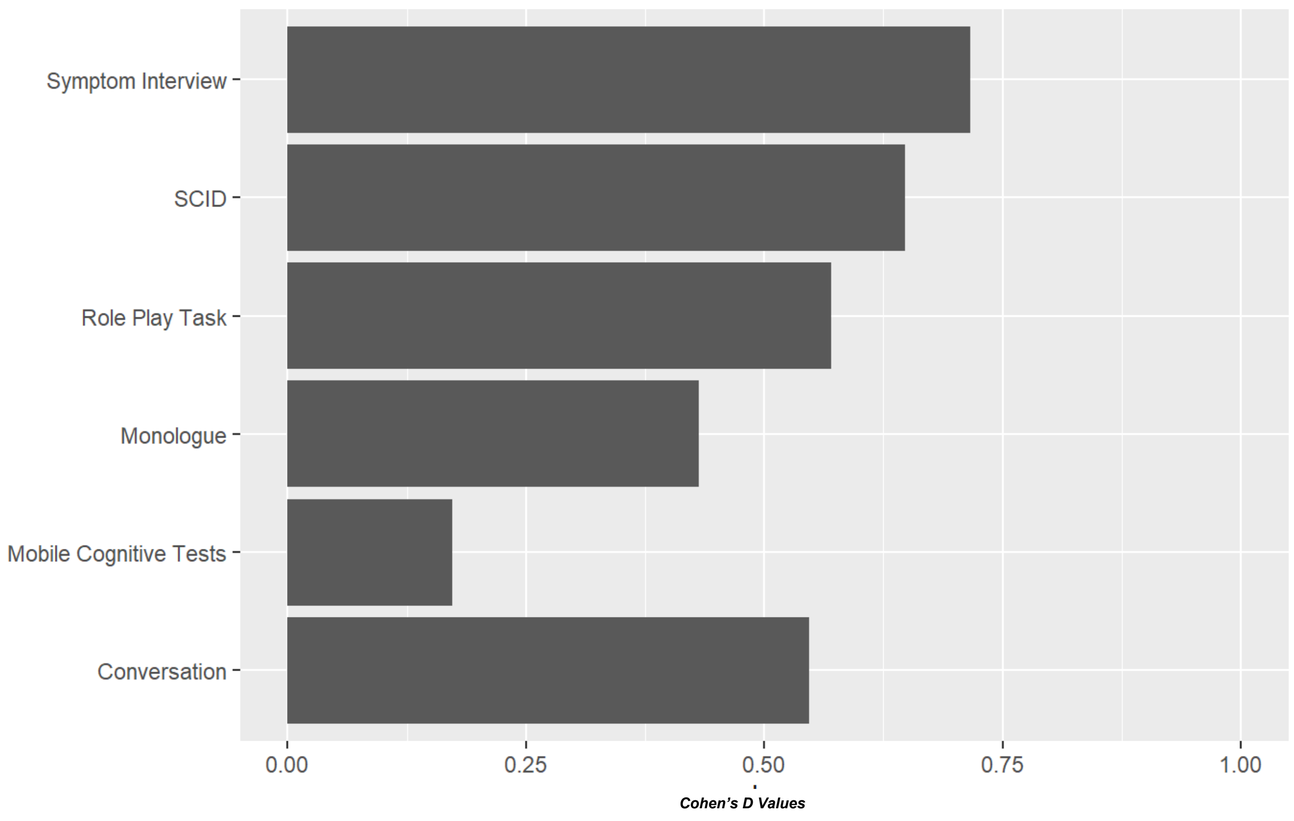
RESULTS

How reliable were the speech features with each other?
The six features were modestly inter-correlated (α= 0.52).

How reliable were speech features across tasks?
In one study, participants (N = 85; K = 436 samples), provided samples for 4 tasks mobile cognitive and monologue tasks over 4 weeks. Reliability across these speaking tasks was low (range of single ICC values = -0.28 to 0.02).

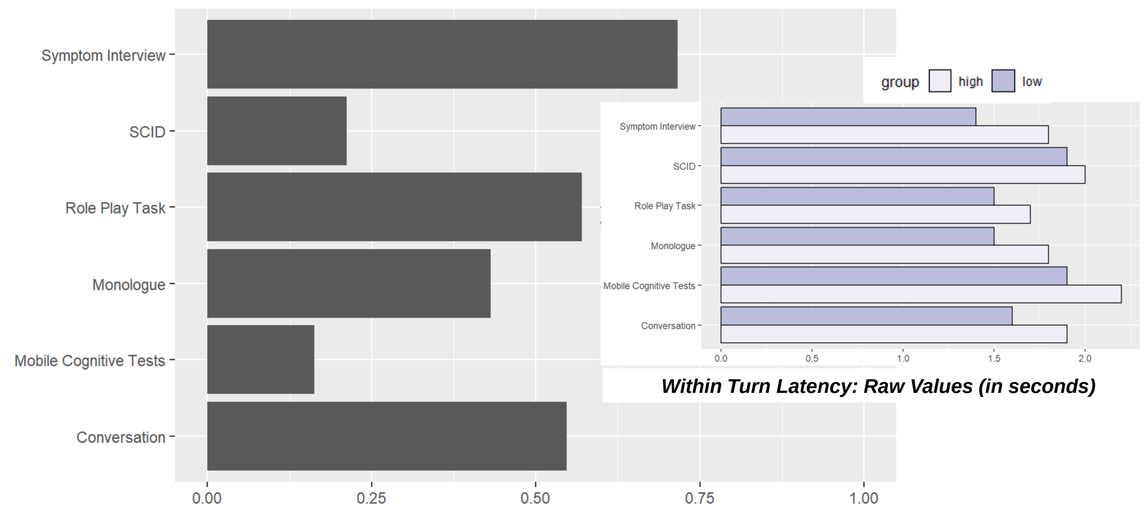
Not all speaking tasks are equal

Patients high versus low in negative symptoms
[Max Cohen's D values for all 6 features]

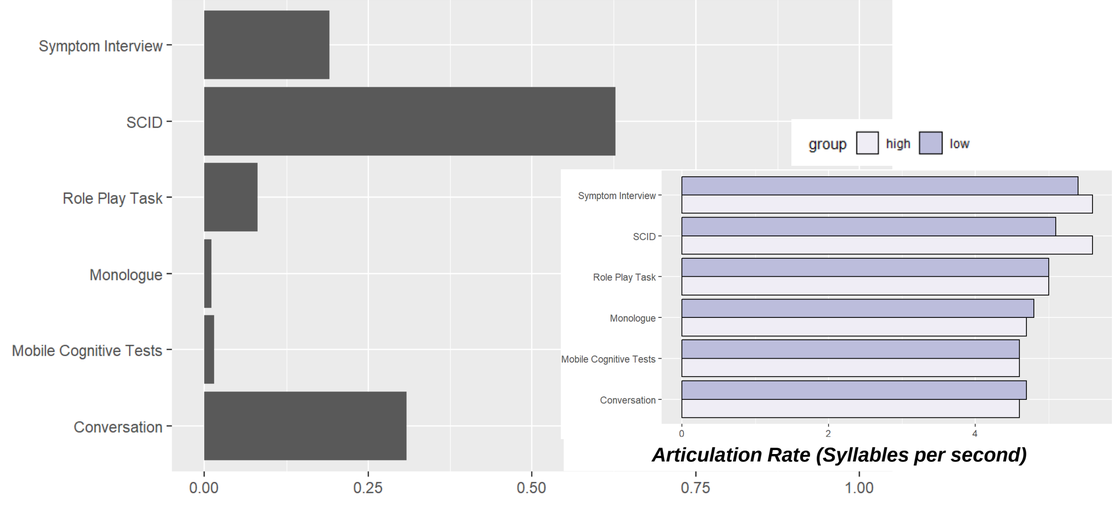


Not all speech features are equal

Within Turn Latency *is longer for negative symptom patients only in some tasks.*
[Cohen's D Values shown here].



Articulation Rate *is slower for negative symptom patients only in some tasks.*
[Cohen's D Values shown here].



CONCLUSIONS

Objective speech technologies show promise for capturing negative symptoms. However, There is no "one-size" fits all solution for measuring negative symptoms using speech analysis. Negative symptoms manifest differently across different tasks.

Clinically rated negative symptoms were best captured using turn latencies from clinical interviews. But this was not true for all clinical interviews (e.g., the SCID).

Mobile phone based cognitive tasks were not effective in capturing clinically rated negative symptoms.

Contact: Alex Cohen. acohen@lsu.edu