

# Latent Factors of Language Disturbance and Relationships to Quantitative Speech Features

## SUPPLEMENTAL MATERIALS

### Methods:

#### *Participants*

Of the 343 speech samples analyzed in this study:

- 47 samples of free speech were collected from publicly-available educational videos of psychiatric interviews on YouTube, including 28 samples from people designated as having a psychotic disorder, and 19 samples from people reported to have a non-psychotic psychiatric condition (further detailed by Krell et al.<sup>1</sup>). Audio samples were cut from posted interviews to standardize length (1.5-2min) and isolate open-ended narratives, consistent with the other speech samples. Because psychotic disorder diagnosis could not be confirmed, samples from individuals identified as having psychotic disorders were designated as unconfirmed but probably psychosis (PSY). Analysis of these samples were done on publicly available data, so this portion of the study was not considered human subjects research and did not undergo IRB review.
- The remaining 296 speech samples were collected from participants enrolled in ongoing studies of speech biomarkers in psychiatric disorders. Participants were recruited from inpatient and outpatient programs at the Zucker Hillside Hospital, rosters of past research participants, and posts on the internet. Free speech was recorded in response to two open-ended questions asking for a self-description and report on recent experiences. Of the 296 speech samples, 153 were collected in-person via a proprietary iOS app developed by Winterlight Labs, 17 were collected in-person via a standard digital voice recorder, and 126 were collected virtually over Microsoft Teams. These samples were collected as a part of larger studies which included additional language tasks (picture descriptions, fluency tasks) and clinical assessments. TLC and SANS ratings were given holistically for the overall assessment.

#### *Speech Features*

Processing of all 343 samples occurred through an automated pipeline. We initially selected 79 features for evaluation. 52 features were omitted after VIF analysis due to redundancy with other features. A cutoff of  $VIF < 5$  was used for all categories except for cosine embedding distances, where we used  $VIF < 10$  because no features had a  $VIF < 5$ . See Supplemental Table 1 for final feature table ( $n=27$ ).

- Transcription: Human annotators produced verbatim transcriptions, including speaker labels and timestamps. Utterance boundaries were defined intuitively based on pauses and syntax. Dysfluencies and speech errors such as filled pauses, partial words and repeated words were tagged.
- Prosody and voice quality features were extracted from OpenSMILE<sup>2</sup> using the interspeech13 configuration.<sup>3</sup> Recordings were diarized for participant vs. interviewer speech and only participant segments were used to calculate prosodic and voice quality features. Starting features = 6.
- Speaking tempo and pauses were calculated by first aligning transcripts with audio recordings using the Montreal forced aligner.<sup>4</sup> Segments were labeled as voiced vs. unvoiced, participant vs. interviewer. Pause lengths and rate of speech were quantified for the participant. Starting features = 7.
- Semantic cosine embedding distances were calculated by first removing filled pauses, partial words, repetitions, and NLTK stop words.<sup>5</sup> Then, word embeddings were calculated using Latent Semantic Analysis (LSA),<sup>6</sup> Word2Vec,<sup>7</sup> and Glove.<sup>8</sup> Utterance embeddings were calculated by taking the simple mean of the word embeddings, or using TF-IDF weighting.<sup>9</sup> Starting features = 6.
- Speech errors and dysfluencies were tagged during transcription, counted, and standardized by the total word count. Starting features = 3.
- Speech graph features were generated using sequential, cooccurrence and action-predication methods further detailed by Nikzad et al.<sup>10</sup> Briefly, sequential graphs were created by connecting each word to the next, cooccurrence graphs were made by connecting all words that appear in each sentence together, and action-predication graphs were produced by connecting entities that act upon each other and the content of an activity to its participants. The graph features describing network size (number of nodes, number of edges, average shortest path length, and network diameter) and network connectedness (average degree, average weighted degree, graph density, average clustering coefficient, size of the largest clique, size of the largest strongly connected component and number of triangles) were calculated for each graph type Starting features = 31.
- Lexical characteristics for age of acquisition,<sup>11</sup> prevalence,<sup>11</sup> and semantic diversity<sup>12</sup> were calculated based on published norms. Starting features = 3.

- Sentiment features examined valence, arousal, and dominance characteristics of words based on published norms.<sup>13</sup> Starting features = 5.
- Parts-of-speech were counts of POS tags generated using spAcy,<sup>14</sup> standardized by the total word count in the sample. Starting features = 14.
- Speech quantity measures such as word count, utterance count, average utterance length, were directly calculated from the transcripts. Starting features = 4.

**Supplemental Table 1: Speech Features**

Category	Feature	Description
COS: Semantic Cosine Embedding Distances	LSAcosineDist	Utterance embeddings calculated from mean embeddings of individual words. Cosine distance calculated between adjacent sentences.
ERR: Speech Errors and Dysfluencies	FilledPauses	E.g., um, uh, er; count per 100 words.
ERR: Speech Errors and Dysfluencies	PartialWords	Dysfluent words which were interrupted by the speaker and not completely uttered. E.g., "I went to the <b>st-</b> store."
ERR: Speech Errors and Dysfluencies	RepeatedWords	Dysfluent repeated words or segments. E.g., "I I went to the store." "I went I went to the store."
GPH: Graph Features	ActionPredClustCoeff	Clustering coefficient calculated from action-predication graph with dysfluencies, repetitions removed
GPH: Graph Features	SeqClustCoeff	Clustering coefficient calculated from sequential graph with dysfluencies, repetitions removed
GPH: Graph Features	SeqLargestClique	Largest clique (interconnected component) calculated from sequential graph with dysfluencies, repetitions removed
LEX: Lexical Characteristics	AgeofAquisition	Average age at which word is acquired
POS: Parts-of-speech	Adjectives	After removing repeated segments, count of adjectives per 100 words
POS: Parts-of-speech	Adpositions	After removing repeated segments, count of prepositions per 100 words
POS: Parts-of-speech	CoorConjunc	After removing repeated segments, count of coordinating conjunctions per 100 words
POS: Parts-of-speech	Determiners	After removing repeated segments, count of determiners per 100 words
POS: Parts-of-speech	Particles	After removing repeated segments, count of particles per 100 words
POS: Parts-of-speech	SubConjunc	After removing repeated segments, count of subordinating conjunctions per 100 words
QUAN: Speech Quantity	TypeTokenRat	Ratio of unique tokens to total number of tokens
QUAN: Speech Quantity	UtterLength	Mean length (in words) of each utterance in the verbatim transcription
SENT: Sentiment	Arousal	Degree to which words reflected arousal state (+) vs. calm (-)
SENT: Sentiment	NegativeValence	Degree of negative valence expressed; mean valence score of all negative-valence-coded content words
SENT: Sentiment	PositiveValence	Degree of positive valence expressed; mean valence score of all positive-valence-coded content words
TEM: Tempo and Pauses	MeanPauseLength	Mean duration of pauses in participant speech
TEM: Tempo and Pauses	MeanSpeakRate	Words per utterance divided by utterance duration; mean across all participant speech.
TEM: Tempo and Pauses	MeanTurnLatency	Mean initial pause duration for participant turns (after interviewer speech)
TEM: Tempo and Pauses	MinSpeakRate	Words per utterance divided by utterance duration; minimum value across all participant speech.
TEM: Tempo and Pauses	PauseLengthVari	Standard deviation of pause lengths
VQ: Voice Quality and Prosody	MeanPitch	Mean of F0
VQ: Voice Quality and Prosody	PitchVari	Standard deviation of F0
VQ: Voice Quality and Prosody	ShimmerVari	Standard deviation of the local (frame-to-frame) shimmer (amplitude deviations between pitch periods)

### Clinical Assessments

A cross-diagnostic subset of participants (n=125) were assessed with the following clinical rating scales. To generate consistent directionality, the functioning and cognitive scores were inverted so that higher scores reflected greater impairment, as for the symptom ratings.

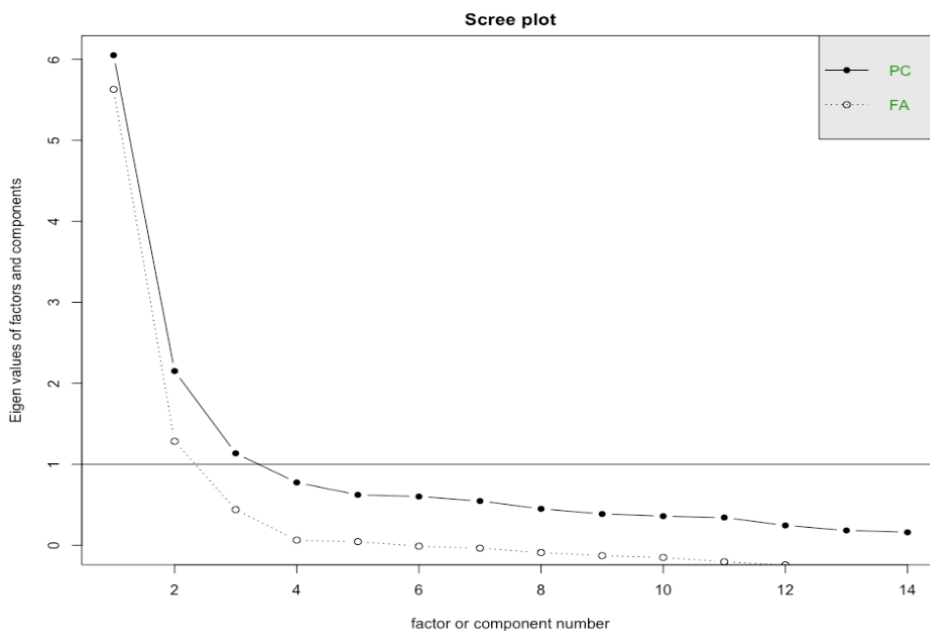
- BPRS: Brief Psychiatric Rating Scale,<sup>15</sup> with 5 factors calculated as per Overall et al.
- SANS: Scale for the Assessment of Negative Symptoms,<sup>16</sup> with global scores for 4 domains as per Andreasen.
- QLS: Heinrich's Quality of Life Scale.<sup>17</sup>
- Cognitive Impairment (COG): Social cognition was assessed with the Hinting Task,<sup>18</sup> Penn emotion recognition task,<sup>19</sup> and verbal fluency tasks (F-letter, and category).

**Supplemental Table 2: Clinical Characteristics**

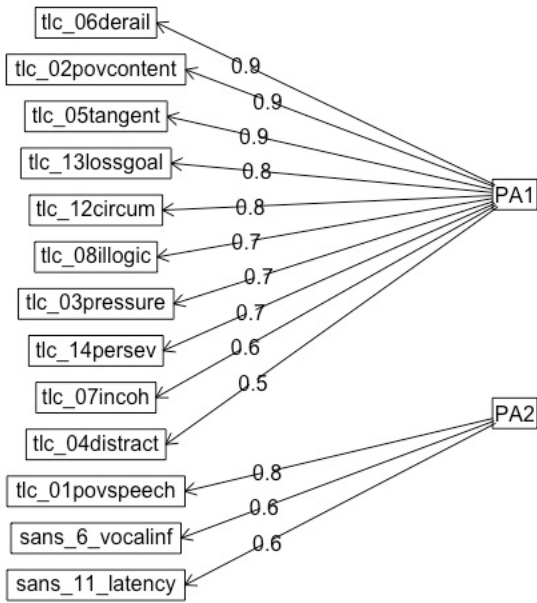
Category	Item	Description
BPRS	AnxDep	Factor score for anxious depression: Anxiety, Guilty Feelings, Depressive Mood
BPRS	HostSusp	Factor score for hostile suspiciousness: Hostility, Suspiciousness, Uncooperativeness
BPRS	ThoughtDist	Factor score for thought disturbance: Conceptual Disorganization, Hallucinatory Behavior, Unusual Thought Content
BPRS	WithdrawRet	Factor score for withdrawal retardation: Emotional withdrawal, Motor retardation, Blunted affect
SANS	AffectFlat	Global score for affective flattening
SANS	Alogia	Global score for alogia
SANS	Avolition	Global score for avolition
SANS	Anhedonia	Global score for anhedonia and asociality
QLS	RoleFx	(Inverted) Subscore for role functioning
QLS	SocialFx	(Inverted) Subscore for social functioning
COG	TOM	(Inverted) Hinting score
COG	EmoProc	(Inverted) Penn emotion recognition score
COG	LetterFl	(Inverted) Performance on F-letter fluency task
COG	CategoryFl	(Inverted) Performance on category (animals) fluency task

### Factor Analyses

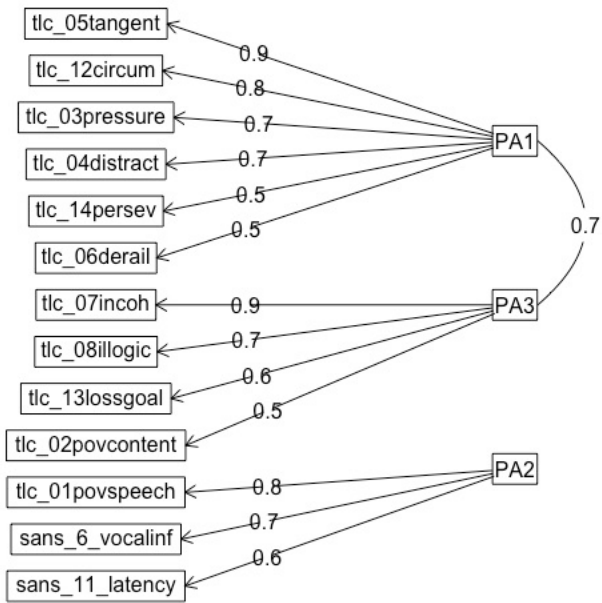
**Supplemental Figure 1: Scree Plot for Exploratory Factor Analysis**



**Supplemental Figure 2: 2-Factor Model**



**Supplemental Figure 3: 3-Factor Model**



## Supplemental References:

1. Krell R, Tang W, Hänsel K, Sobolev M, Cho S, Tang SX. Lexical and Acoustic Correlates of Clinical Speech Disturbance in Schizophrenia. *W3PHAI 2021 Stud Comput Intell.* 2021;1013:9. doi:[https://doi.org/10.1007/978-3-030-93080-6\\_3](https://doi.org/10.1007/978-3-030-93080-6_3)
2. Eyben F, Wöllmer M, Schuller B. Opensmile: the munich versatile and fast open-source audio feature extractor. In: *Proceedings of the 18th ACM International Conference on Multimedia.* MM '10. Association for Computing Machinery; 2010:1459-1462. doi:10.1145/1873951.1874246
3. Eyben F, Weninger F, Gross F, Schuller B. Recent developments in openSMILE, the Munich Open-Source Multimedia Feature Extractor. In: *Proceedings of ACM Multimedia.* ; 2013:835-838.
4. McAuliffe M, Socolof M, Mihuc S, Wagner M, Sonderegger M. Montreal Forced Aligner: trainable text-speech alignment using Kaldi. In: *Proceedings of the 18th Conference of the International Speech Communication Association.* ; 2017.
5. Loper E, Bird S. NLTK: the Natural Language Toolkit. In: *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics.* Vol 1. Association for Computational Linguistics (ACL); 2002:63-70. doi:10.3115/1118108.1118117
6. Landauer TK, Foltz PW, Laham D. An introduction to latent semantic analysis. *Discourse Process.* 1998;25(2-3):259-284. doi:10.1080/01638539809545028
7. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed Representations of Words and Phrases and their Compositionality. *ArXiv13104546 Cs Stat.* Published online October 16, 2013. Accessed August 29, 2021. <http://arxiv.org/abs/1310.4546>
8. Pennington J, Socher R, Manning CD. GloVe: Global Vectors for Word Representation. *Proc 2014 Conf Empir Methods Nat Lang Process.* Published online 2014:1532-1543.
9. Manning C, Raghavan P, Schütze H. *Introduction to Information Retrieval.* Cambridge University Press; 2008. doi:10.1017/CBO9780511809071
10. Nikzad A, Sharifi V, Naghavi H. Network analysis of speech in schizophrenia and bipolar disorder. *Schizophr Res.* Submitted.
11. Brysbaert M, Mandera P, Keuleers E. Word prevalence norms for 62,000 English lemmas. *Behav Res Methods.* 2018;51:467-479. doi:10.3758/s13428-018-1077-9
12. Hoffman P, Lambon Ralph MA, Rogers TT. Semantic diversity: a measure of semantic ambiguity based on variability in the contextual usage of words. *Behav Res Methods.* 2013;45:718-730. doi:10.3758/s13428-012-0278-x
13. Warriner AB, Kuperman V, Brysbaert M. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behav Res Methods.* 2013;45(4):1191-1207. doi:10.3758/s13428-012-0314-x
14. Honnibal M, Johnson M. An Improved Non-monotonic Transition System for Dependency Parsing. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing.* Association for Computational Linguistics; 2015:1373-1378.
15. Overall JE, Gorham DR. The Brief Psychiatric Rating Scale. *Psychol Rep.* 1962;10(3):799-812. doi:10.2466/pr0.1962.10.3.799
16. Andreasen NC. Scale for the Assessment of Negative Symptoms (SANS). In: University of Iowa; 1984.
17. Heinrichs DW, Hanlon TE, Carpenter WT. The Quality of Life Scale: an instrument for rating the schizophrenic deficit syndrome. *Schizophr Bull.* 1984;10(3):388-398. doi:10.1093/schbul/10.3.388

18. Corcoran R, Mercer G, Frith CD. Schizophrenia, symptomatology and social inference: Investigating “theory of mind” in people with schizophrenia. *Schizophr Res.* 1995;17(1):5-13. doi:10.1016/0920-9964(95)00024-G
19. Kohler CG, Richard JA, Brensinger CM, et al. Facial emotion perception differs in young persons at genetic and clinical high-risk for psychosis. *Psychiatry Res.* 2014;216(2):206-212. doi:10.1016/j.psychres.2014.01.023