Using Speech to Detect Mental Health Symptoms: Benefits and Risks

Daniel M. Low
Harvard University & MIT

ADSA 27, Northeastern University
11/15/2023
dlow@g.harvard.edu

Detecting mental health symptoms from speech

- **Advantages over self-report rating scale assessments:** more objective, ecological, passive (journaling, phone calls), not retrospective, not constrained, and also remote and low-cost.
- **Results:** hundreds of studies of detected symptoms with low-to-high performance [1]
- **Limitations of most studies:** small sample sizes; cross-sectional; do not test generalizability across datasets, accents, time; do not control for all confounders and biases
- **TRL:** 5-7 (outside the lab with real patients)
- **Low risks:** send irrelevant resources or follow-up assessments to false positives
- **High/unacceptable risks:**
  - Hospitalizing false positives, unconsented assessments, rejecting a candidate for a job or insurance
  - European Union AI Act suggests many uses of biometrics in law enforcement (e.g., face or voice identity recognition) currently have an unacceptable risk and proposes to forbid its practice with some exceptions
- **Path forward:**
  - Large representative samples, standardization of feature extraction, and an understanding of errors, biases, ethics, and privacy.
  - NIH Common Funds Bridge2AI Voice as a Biomarker is a consortium that is tackling precisely this for mood disorders, vocal pathologies, dementia, respiratory disorders, and developmental disorders.

1. Low et al. (2020). Automated assessment of psychiatric disorders using speech: A systematic review. *Laryngoscope investigative otolaryngology.*

Speech analytics for healthcare settings

- Mental health disorders in the United States affect about 1 in 5 adults [1]
- Police officers seem to exhibit higher rates of PTSD and suicide than the general population [2]
- Mental health issues can deteriorate sleep and increase aggression and substance use.[2]

1. SAMHSA. (2022). Results from the 2021 National Survey on Drug Use and Health.
2. Otto & Gatens (2022). Understanding police officer stress: A review of the literature. Illinois Criminal Justice Information Authority.

Disorder	Articles % (N)	Median sample size (range)
Depression	49.6 (63)	123 (11-1688)
PTSD	7.9 (10)	41 (10-253)
Schizophrenia	18.1 (23)	44 (18-195)
Anxiety	4.7 (6)	45 (20-104)
Bipolar	16.5 (21)	39 (5-89)
Bulimia	0.8 (1)	22 (-)
Anorexia	1.6 (2)	107 (66-148)
OCD	0.8 (1)	35 (-)

Low et al. (2020). Automated assessment of psychiatric disorders using speech: A systematic review. *Laryngoscope investigative otolaryngology.*
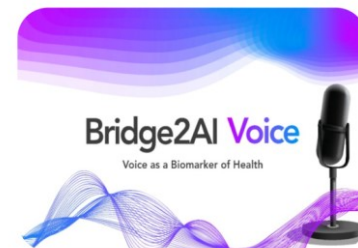
For speech-based assessments, we need:
- larger samples
- representative samples
- public datasets
- standardization (recording, protocols, feature extraction, validation)
- Longitudinal data
- Ethics and privacy
- To understand confounders and bias

Consider partnering with

NIH
The Common Fund

Bridge2AI Voice
Voice as a Biomarker of Health

b2ai-voice.org

Speech analytics for law enforcement

There is increasing interest in using speech to detect malicious intent, aggression, and antisocial behavior.

However, all artificial intelligence models make mistakes:

- Should we trust a model that has a 75% accuracy? What about 95%? Does it have bias? Will it generalize to new samples over time?
- Errors, biases, and confounders of specific models are still not well understood by researchers.
- Verifying malicious *intent* (cognition, e.g., lying) is very different than detecting *actual* malicious speech (behavior)

Many use cases in law enforcement seem to have an unacceptable risk:

- Erroneously confusing healthy anger for violence which could lead to conflict escalation, a risk for suspects and officers
- We need to distinguish between *individual* assessments (with or without court orders) vs. *public* surveillance
- Monitoring for *public* violence or screams of enable data privacy leaks and totalitarian practices: harmful surveillance, dissuading freedom of speech and assembly, predictive policing which may violate the presumption of innocence
- Geo-fence warrants or dragnets could violate the 4th amendment in the US against unreasonable search and seizure

A commission of experts released the European Union AI Act that suggests many uses in law enforcement, especially real-time remote biometric identification, have an unacceptable risk and proposes to forbid its practice for now

More research is needed on lower risk scenarios like screening tools for healthcare → Bridge2AI Voice as a Biomarker is doing just this.

Thank you!

Any questions?

dlow@g.harvard.edu

@danielmlow