

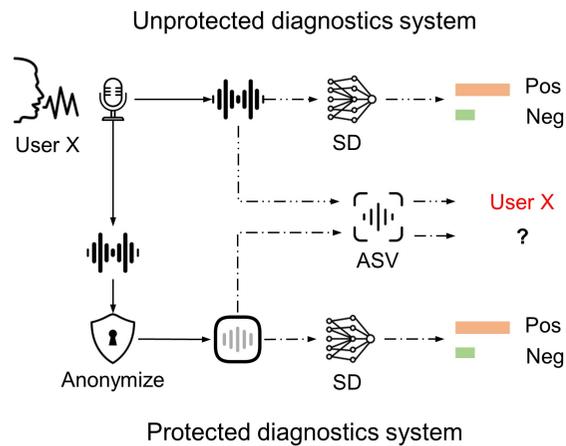
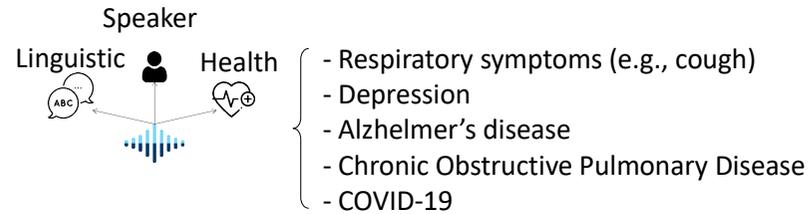
Investigating Biases in COVID-19 Diagnostic Systems Processed with Automated Speech Anonymization Algorithms

Yi Zhu | Multisensory Signal Analysis&Enhancement Lab (MuSAE), Institut national de la recherche scientifique (INRS)

E-mail: Yi.Zhu@inrs.ca



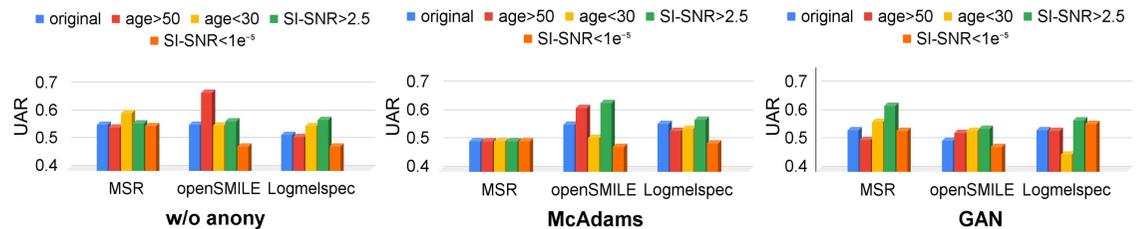
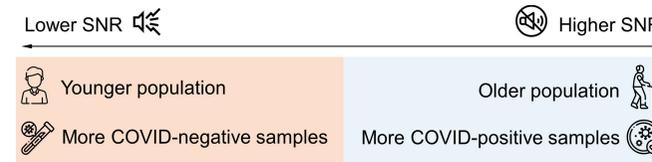
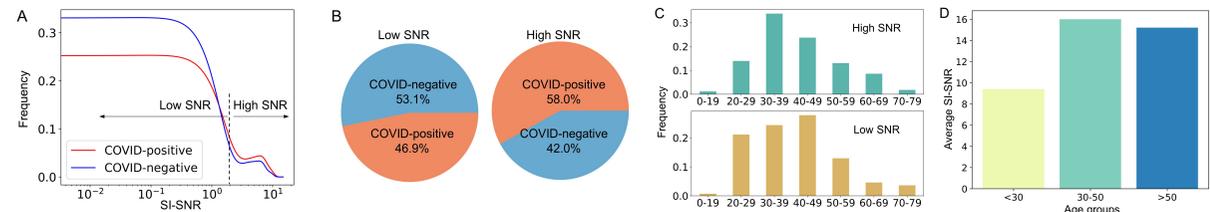
Background



A case study on COVID-19 speech data

- Cambridge COVID-19 Sounds Database - speech
- Utterance of a sentence with COVID labels
- Binary classification: Positive v.s. Negative
- Equally distributed class labels and genders

Results



Take-away message

- Diagnostic performance is found biased by environmental noise and age
- The diagnostic accuracy is potentially lower for individuals with lower socio-economic status
- The effects of biases can be accentuated by voice anonymization
- Data augmentation helps with diagnostic performance in some cases, but may accentuate biases for over- and under-represented groups

Research question

- What are the biases?
- How do they impact an anonymized diagnostics system?
- Countermeasures?

