VocalLock: Sensing Vocal Tract for Passphrase-Independent User Authentication Leveraging Acoustic Signals on Smartphones

Presenter: Li Lu

Li Lu[1], Jiadi Yu[1], Yingying Chen[2], Yan Wang[3]

Shanghai Jiao Tong University[1]

Rutgers University[2] Temple University[3]

Background

Background

Background

Background

Fingerprint

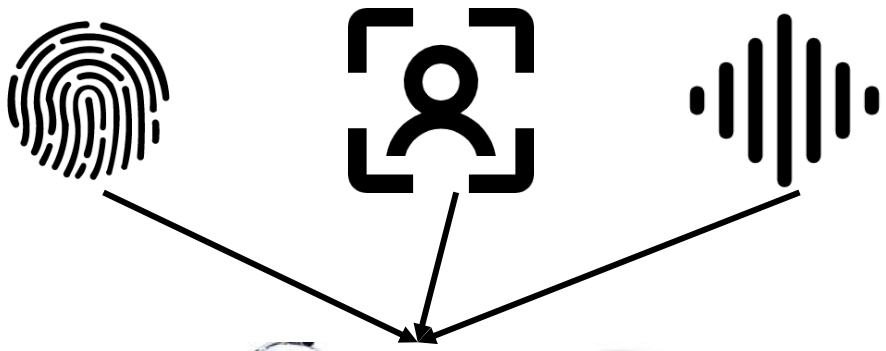
Face

Voiceprint

Existing Approaches

A legitimate user?
A spoofing AI? → Liveness Verification!

Existing Approaches

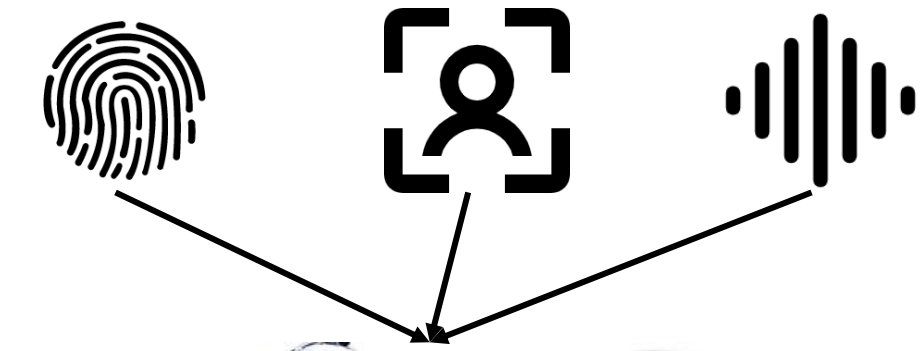


A legitimate user?
A spoofing AI?

Liveness Verification!

Behavior!

Existing Approaches

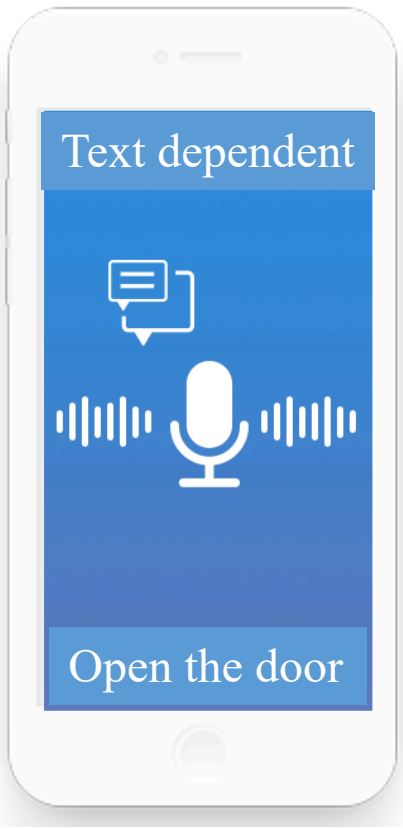


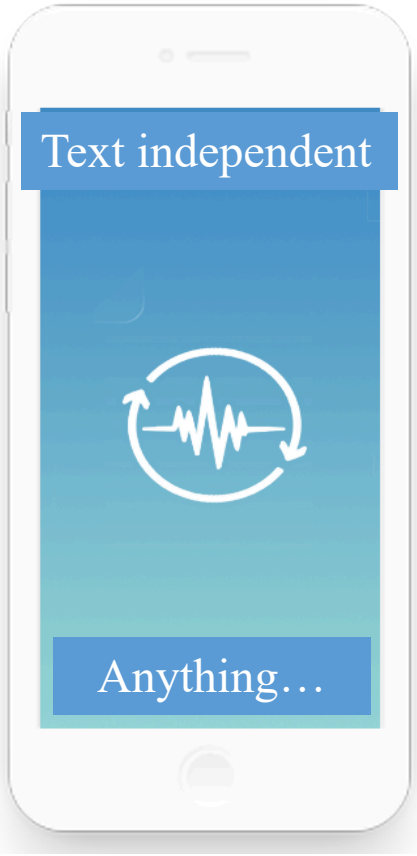
A legitimate user?
A spoofing AI?

Liveness Verification!

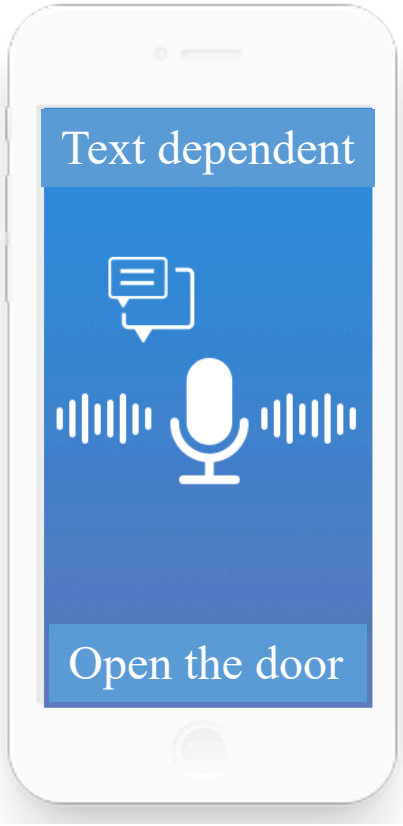
Behavior!

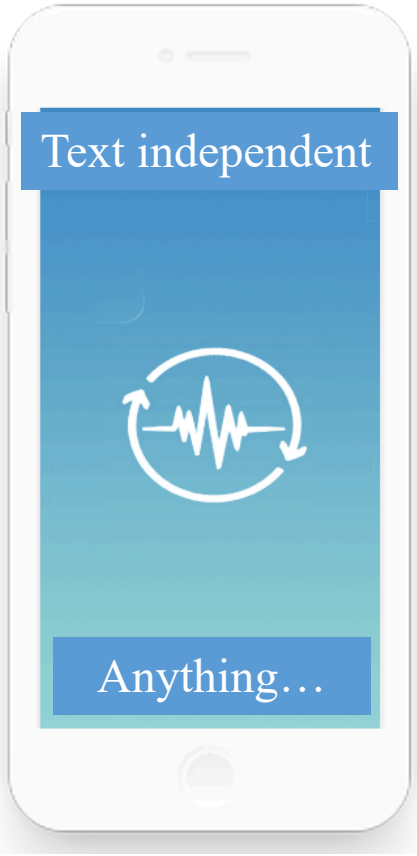
Passphrase-independent Authentication





Passphrase-independent Authentication

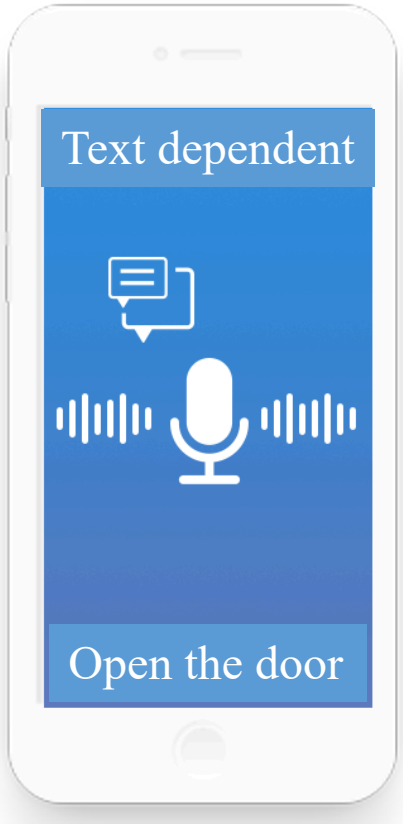


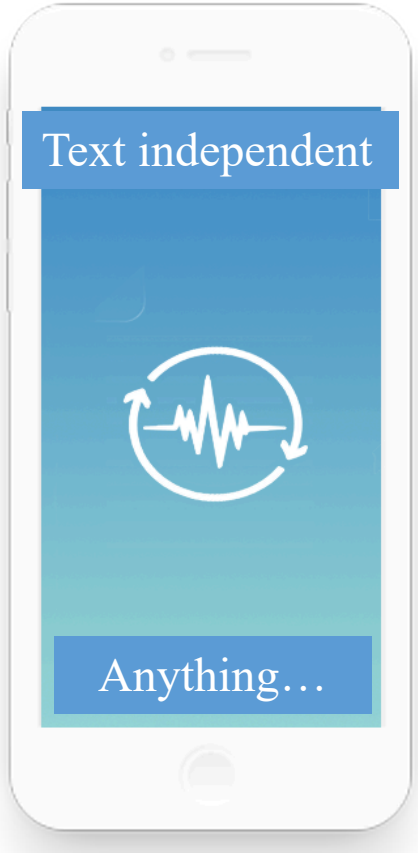


For Unlock

Memory Required

Passphrase-independent Authentication





For Unlock

Memory Required

Flexible

More scenarios supported

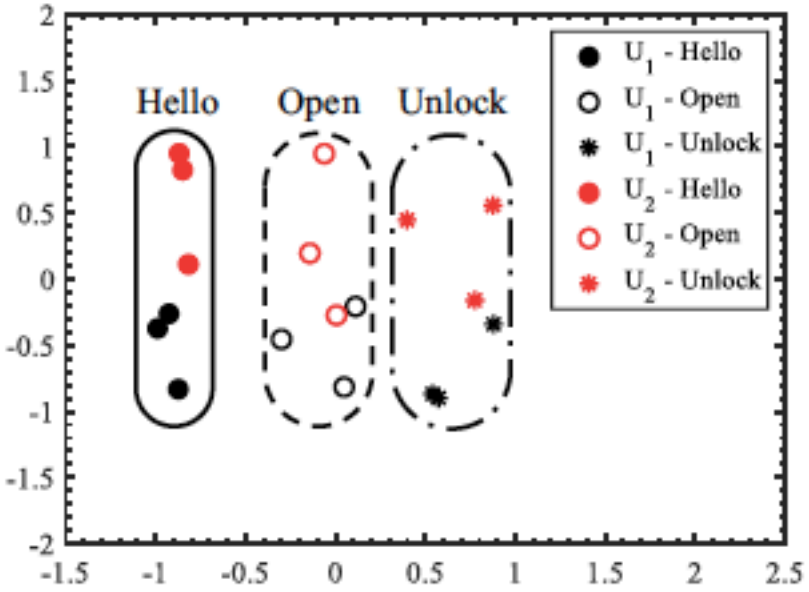
Passphrase-independent Authentication

F pported

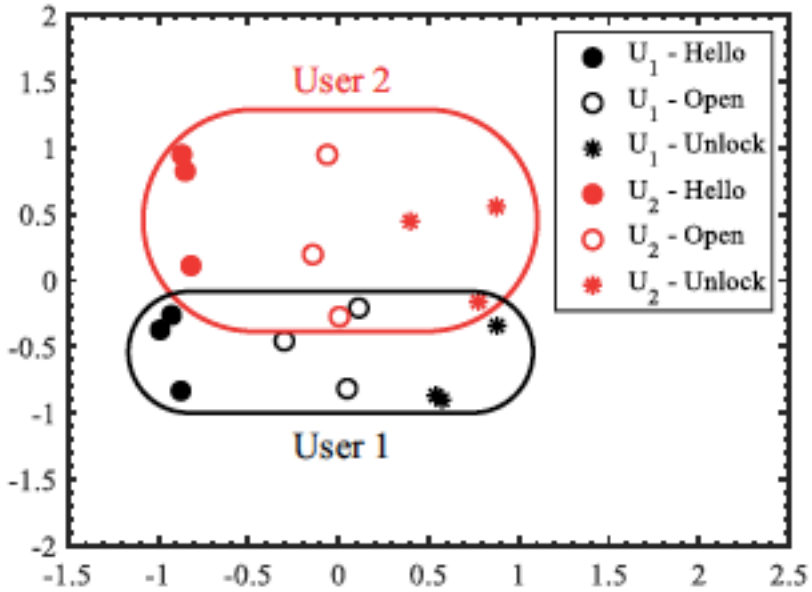
Preliminary

➢ How to implement?

- Inspired by text-independent voiceprint authentication, we employ statistical feature-based method
- Feasibility study: based on PCA



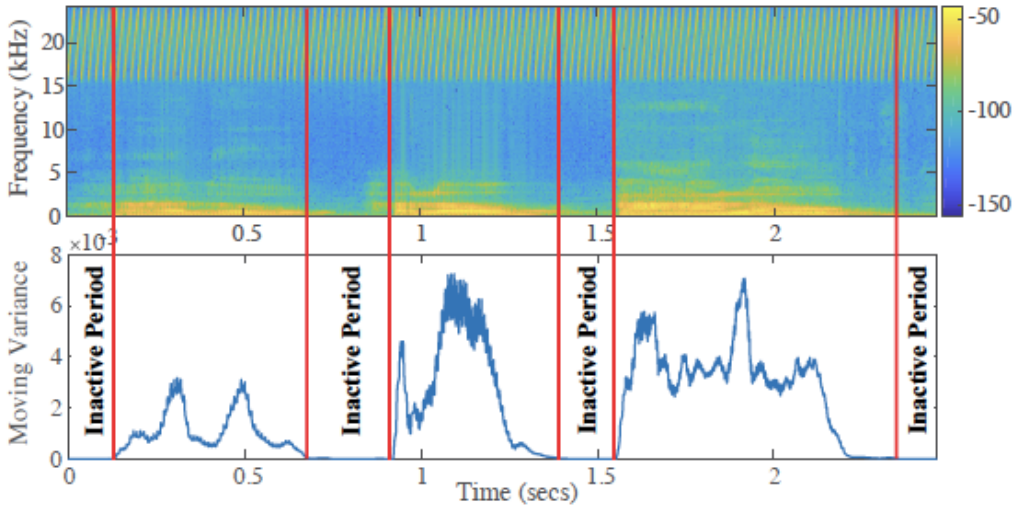
(a) Perspective of passphrases.

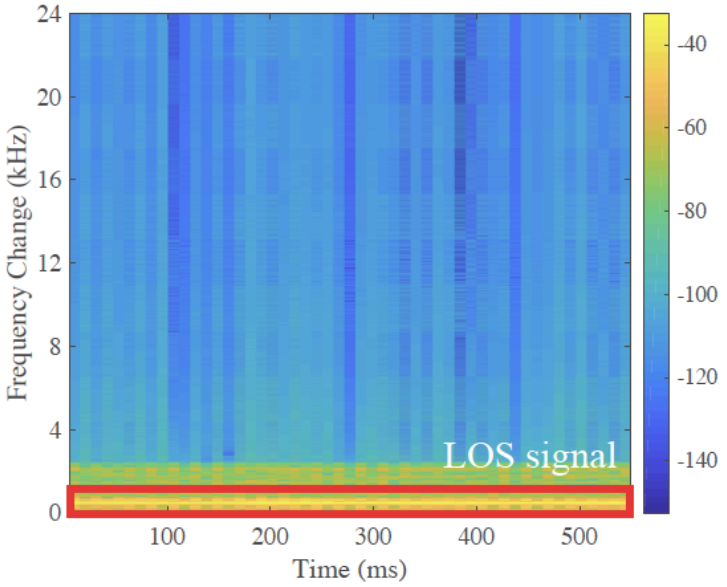


(b) Perspective of volunteers.

Sensing Vocal Tract with FMCW

- Signal Design & Processing
  - Period: time-invariant behavior within 20ms
  - Bandwidth: sensing resolution + device limitation and human auditory → 16kHz～24kHz
  - Segmentation: voice detection → Moving variance on 300Hz~5kHz signal band
  - Dechirping: multiply operation on both signals

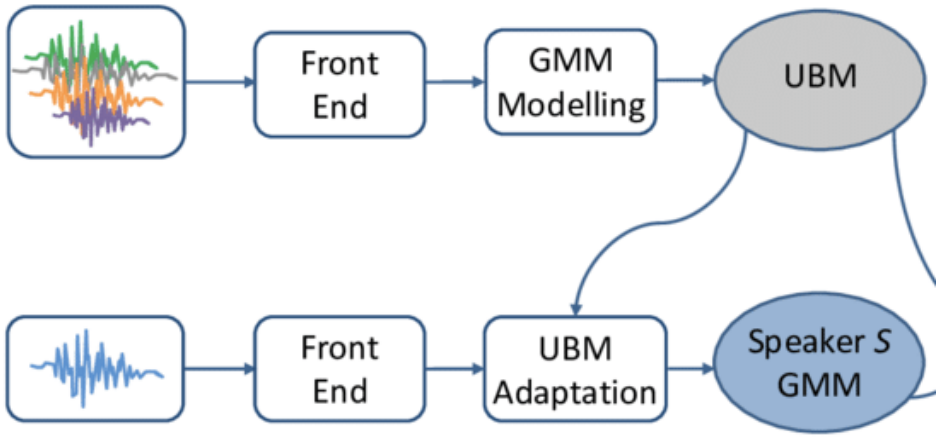


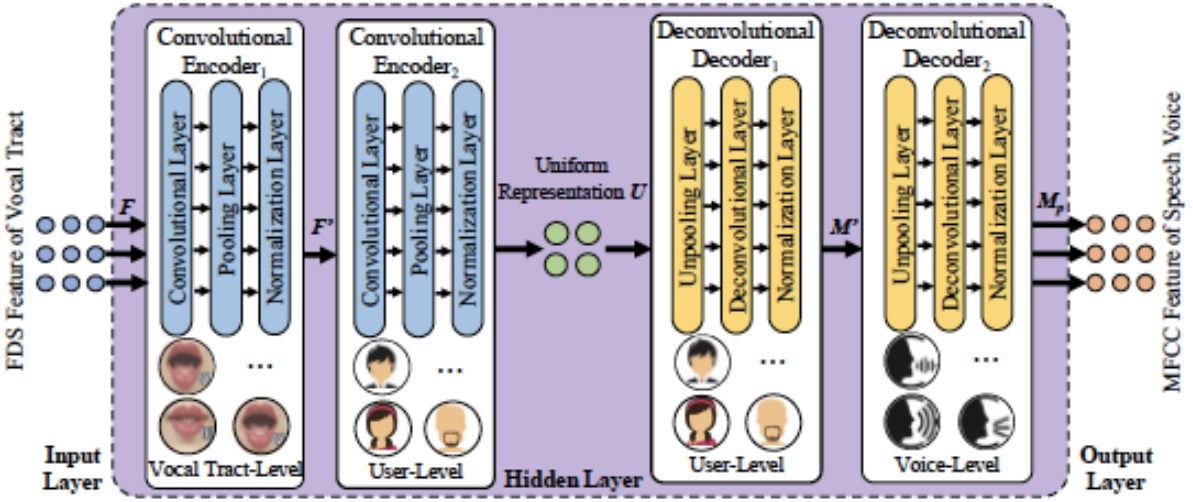


- Extract Features of Vocal Tract
  - Signal components: targeted signal + LOS + other reflections
  - LOS: significantly strong, covering other signals
    - Elimination: STFT-based heuristic method → Search n-th maximum frequency response and eliminate the first m-th ones by empirical studies
  - Other reflections elimination: threshold on ToF

Construct Passphrase-independent Model

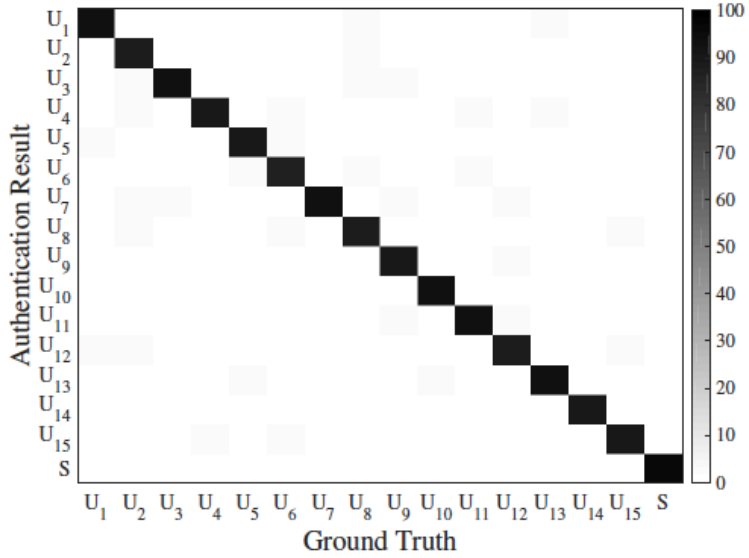
- Passphrase-independent Authentication
  - Statistical model: GMM-UBM (widely-used in text-dependent voiceprint authentication)
  - Employed features: MFCC
  - Pre-operation: frequency difference → MFCC
- Feature Transferring for Model Construction
  - Encoder-decoder structure
  - Encoder: transferring frequency difference to uniform representation, based on CNN
  - Decoder: transferring uniform representation to MFCC, based on Deconv network



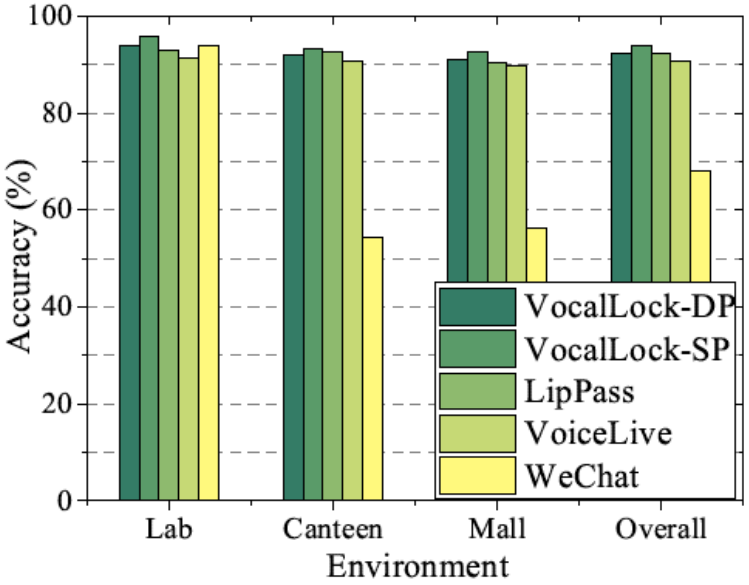


Overall Performance

- Authenticating users in passphrase-independent manner
  - Accuracy of identifying legitimate users: 90.4%
  - Accuracy of detecting spoofers: 96.7%
  - Overall accuracy: 91.0%, standard derivation: 3.1%

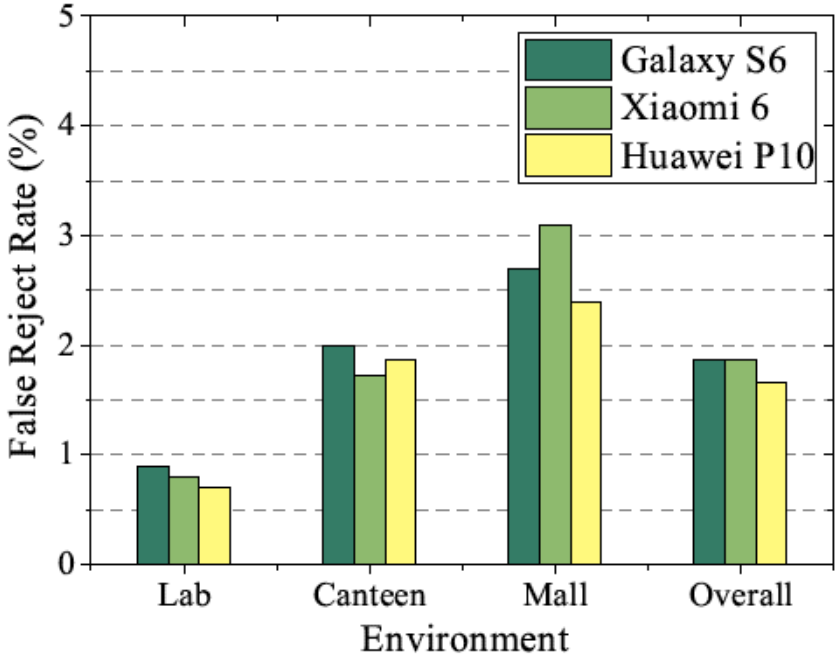


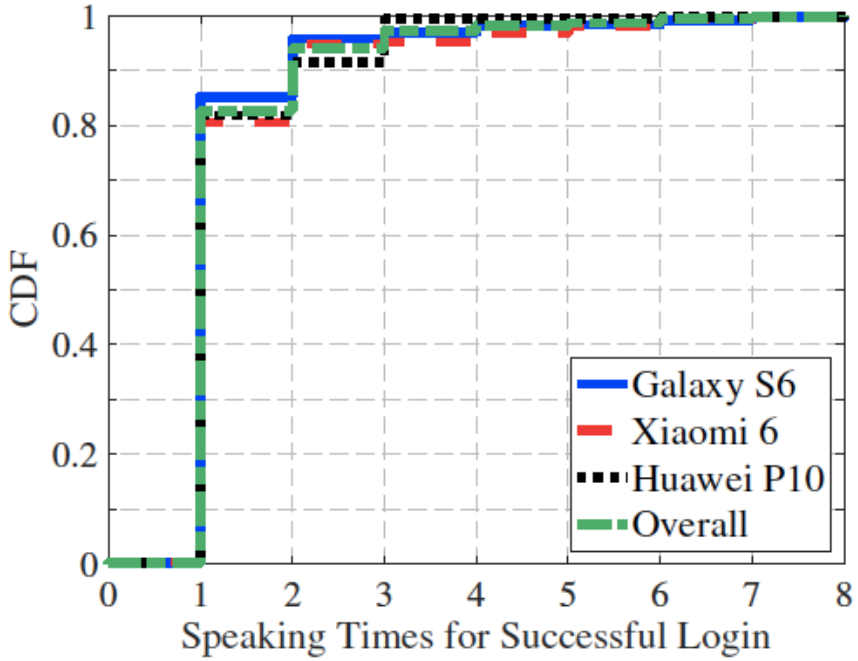
- Performance Comparison
  - VocalLock: 93.8%
  - LipPass: 92.8%
  - WeChat: 94.0%
  - VoiceLive： 90.6%



User Experience

- False Reject Rate
  - Overall: All below 2%
  - FRR under Complex Environment being a little higher (i.e., mall in our experiments)
- Speaking Times for Successful Login
  - 82.7% users could login using only 1 speaking
  - Over 95% users could login within 3 speakings





Conclusion

- Observation:
  - Investigate the feasibility of employing FMCW on acoustic signals to sense the vocal tract
  - Study the feasibility of using statistical methods to realize passphrase-independent user authentication
- Technical Contribution:
  - Propose a passphrase-independent user authentication by sensing vocal tract with FMCW on acoustic signals
  - Develop an STFT-based heuristic method to extract the reflected signals from vocal tract
  - Design transfer learning-based neural network and employ GMM-UBM to construct authentication model
- Performance Evaluation:
  - Authentication accuracy: above 90%
  - False reject rate: below 2%

Thank you!

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

RUTGERS

TEMPLE
UNIVERSITY®