# PRIVACY-PRESERVING SOUND TO DEGRADE AUTOMATIC SPEAKER VERIFICATION PERFORMANCE

Kei Hashimoto [1], Junichi Yamagishi [2,3], and Isao Echizen [2]
([1]Nagoya Institute of Technology, [2]National Institute of Informatics, [3]University of Edinburgh)

## 1. Introduction

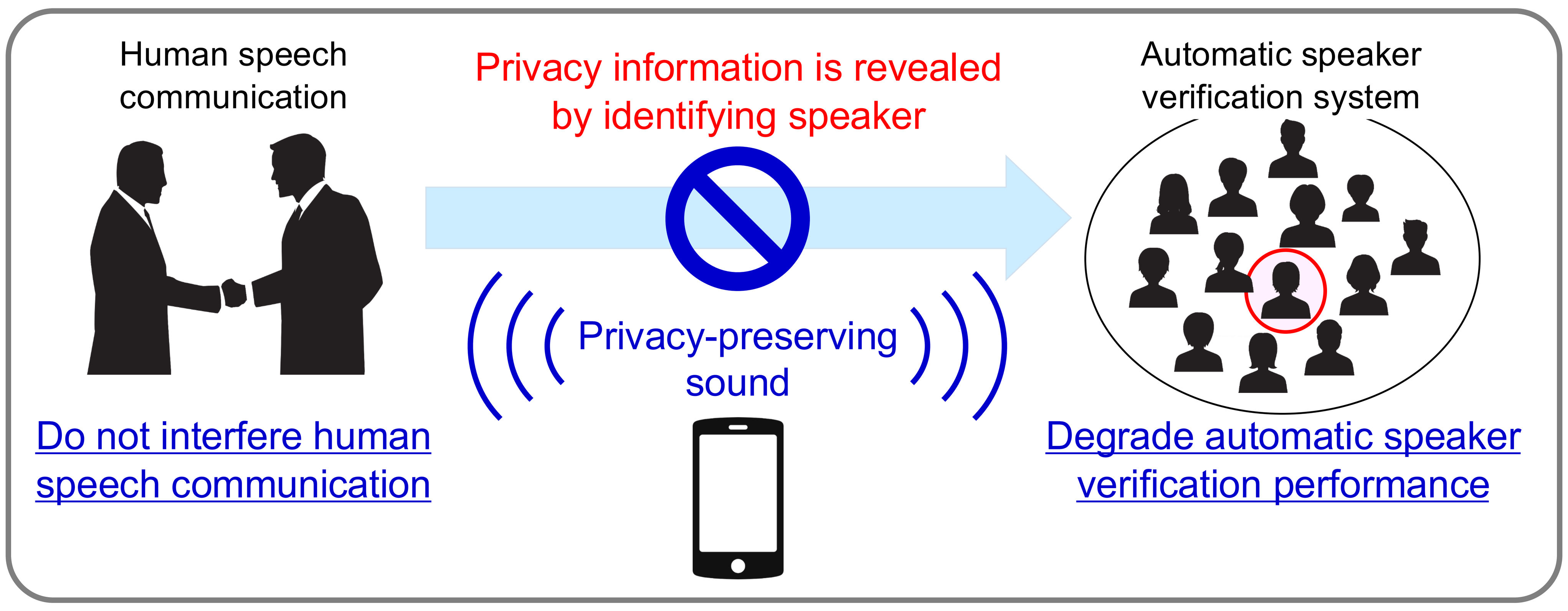
- Privacy problem
  - Multimedia information recorded by someone else is shared on the Internet without the person's permission
  - Private information is revealed by analyzing recorded data
  - Privacy problem becomes more serious if a person's identity can be obtained from speech
- Speaker recognition systems
  - Show higher recognition accuracy than human listeners
  - Privacy protection techniques to prevent speaker identification are needed
- Privacy-preserving sound
  - Degrade the performance of speaker verification systems without interring with human speech communication

## 2. Related work

- Speech includes private information
  - Non-linguistic private information (speaker's identity, gender, etc.)
  - Linguistic private information (person's name, contents, etc.)
- Many privacy protection techniques for speech
  - Speaker de-identification using voice conversion [Jin et al.; '09]
    - Protect non-linguistic private information (speaker's identity)
    - Converted speech still sounds natural and intelligible
  - Sound masking [Ueno et al.; '08] [Akagi and Irie; '12]
    - Protect linguistic private information contained in private conversations in open areas (banks, medical examination rooms, etc.)
    - Speech intelligibility for people in the area may be affected
  - Target private information and situation are different

## 3. Privacy-preserving sound

- Target problem
  - Speech is recorded by someone else and posted on the Internet without permission
  - Private information is revealed by identifying speaker with speaker recognition systems
- Privacy-preserving sound
  - Do not require any processes after recording
    ⇒ Can use in physical spaces
  - Degrade the speaker verification performance
  - Do not interfere human speech communication
- Overview of proposed privacy-preserving sound




## 4. Experiments

- Experimental conditions

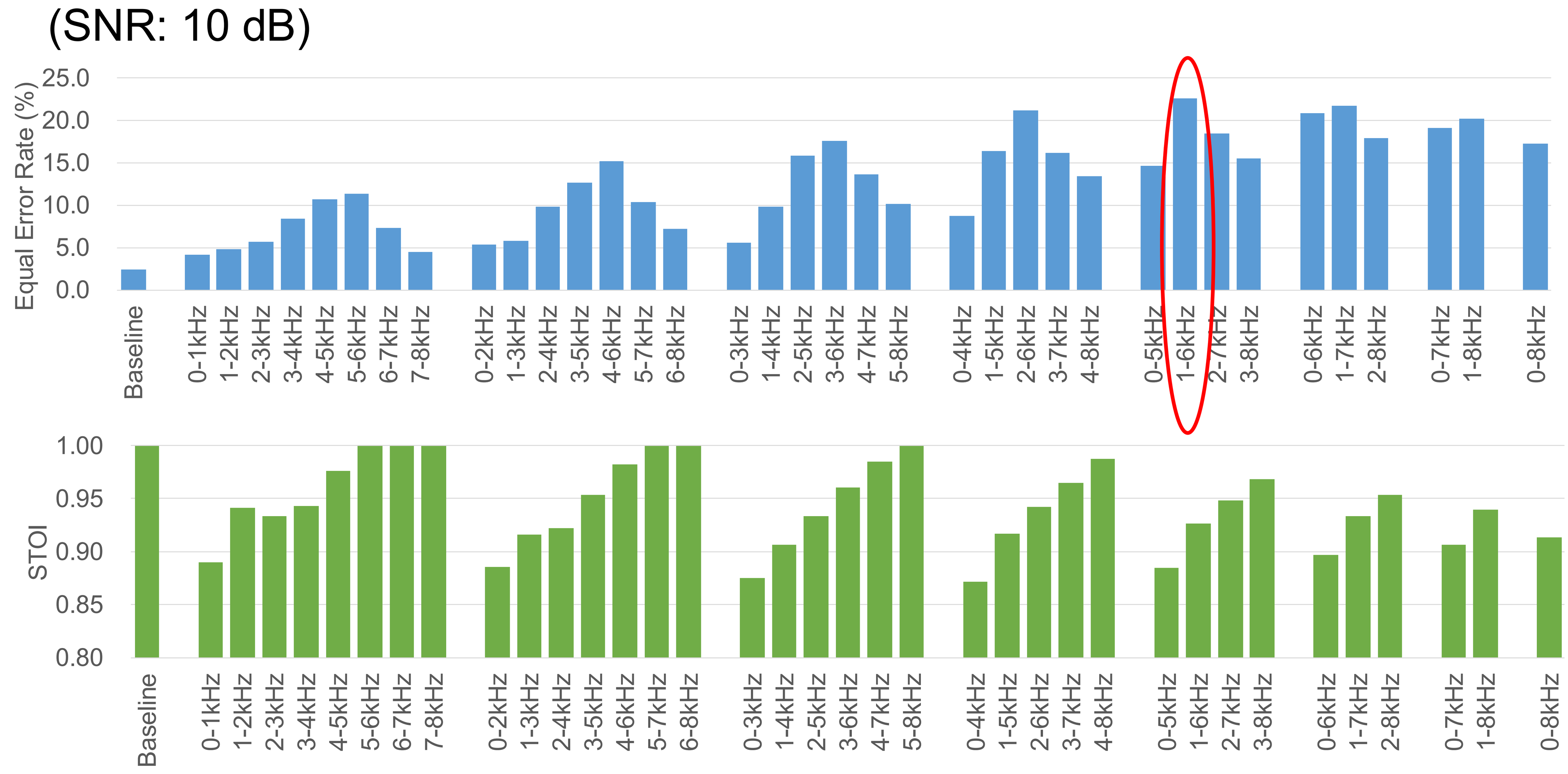
| Database | TIMIT |
|---|---|
| Enrollment speaker | 168 speakers (112 male and 56 female) |
| Training data | 8 utterances for each speaker |
| Test data | 2 utterances for each speaker |
| Sampling rate | 16 kHz |
| Frame size | 25 ms |
| Frame shift | 10 ms |
| Acoustic features | 60-dimensional MFCC vector<br>(19 MFCCs + Energy + delta + delta-delta)<br>Normalize to zero mean and unit variance |

- Automatic speaker verification system

| System | GMM-UBM<br>ALIZE 3.0 toolkit |
|---|---|
| UBM | 256 mixtures GMM<br>with a diagonal co-variance matrix |
| Training data for UBM | 462 speakers (326 male and 136 female)<br>10 utterances for each speaker<br>(total: 4620 utterances) |

- Objective measures
  - Equal error rate (EER)
    - Represents speaker verification performance
  - Short-time objective intelligibility (STOI)
    - STOI outputs a score from 0 to 1 which correlates with human speech intelligibility

- Experimental results
  - Analyze the impact of frequency on EER and STOI
  - Band-pass filters were applied to white noise and the filtered noise were added (SNR: 10 dB)




  - Frequency having the strongest impacts on EER and STOI are different
    - Around 5-6 kHz gave high EER
    - Low-frequency gave small STOI
  - Can create sound that degrades speaker verification performance without degrading human speech intelligibility by taking account of the difference

- Future work
  - Subjective evaluation for speech intelligibility and impacts on conversation in the area

| | 0-8 kHz | 1-6 kHz |
|---|---|---|
| EER | 17.26 | 22.62 |
| STOI | 0.914 | 0.926 |