The always-on nature of VAs triggers misactivation!

Enhancing Security and Privacy Control for Voice Assistants Using Speaker Orientation

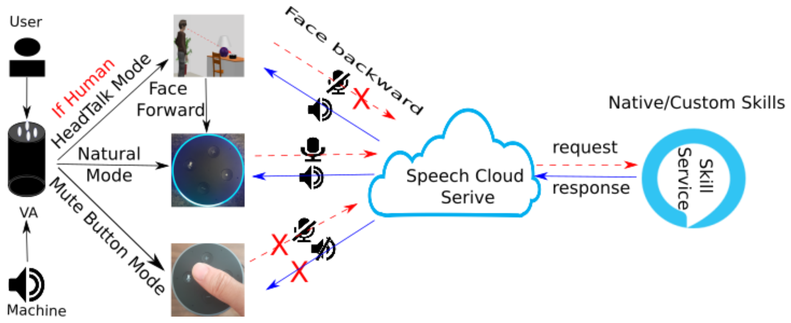
Shaohu Zhang szhang42@ncsu.edu

Aafaq Sabir asabir2@ncsu.edu

Anupam Das anupam.das@ncsu.edu

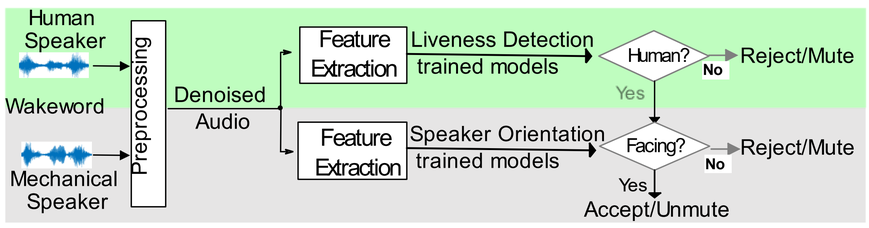
Motivation

- VAs can be misactivated by suboptimal wakeword and background chat.
- Visual gaze plays an important role in interpersonal communication. People tend to look at each other when chatting.
- We propose a device-free and non-obtrusive acoustic sensing system called ***HeadTalk***, a speaker orientation–aware privacy control approach.



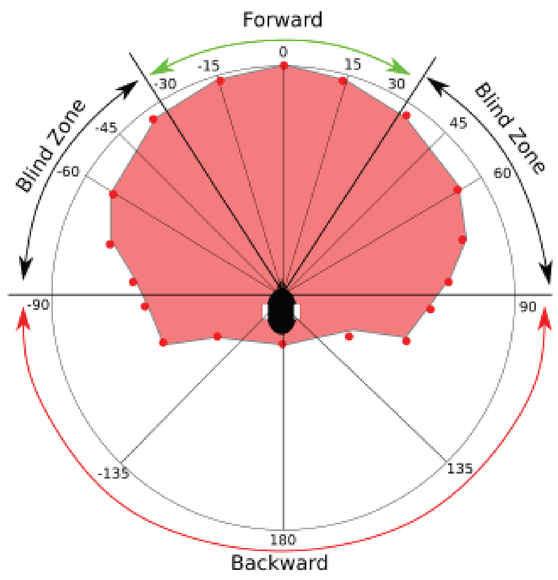
HeadTalk privacy control vs. Natural mode vs. Mute button mode. HeadTalk mode detects the presence of the wake word, and if spoken while facing the VA, it continues to operate in normal mode, otherwise it mutes the microphones

System Design



HeadTalk consists of two main components including liveness detection and speaker orientation recognition. HeadTalk only accepts the wakeword when the user is facing the device.

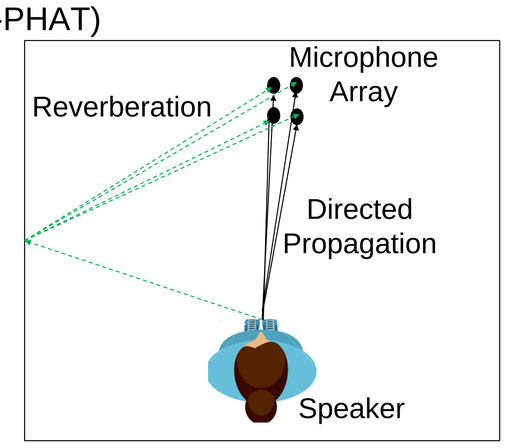
Human Field-of-View (FoV)

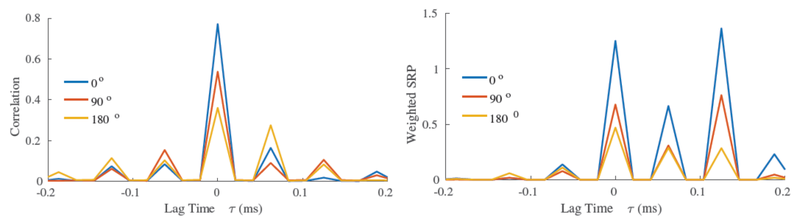


Speech propagation FoV in the horizontal plane. 35 degree on both sides of the centerline is referred to as the immediate FoV

Speech Propagation

- Generalized Cross Correlation with Phase Transform (GCC-PHAT)
- Steered Response Power with Phase Transform (SRP-PHAT)





(a) GCC-PHAT (b) SRP-PHAT

Higher magnitude in the forward direction compared to the backward direction

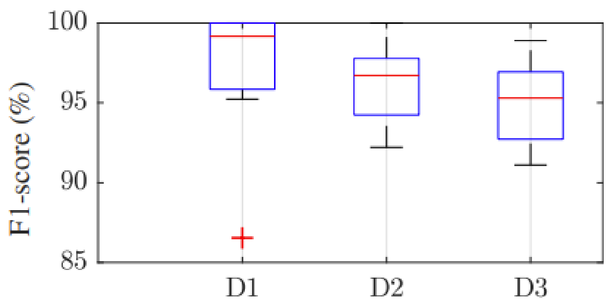
Evaluation

Experiment across three prototypes

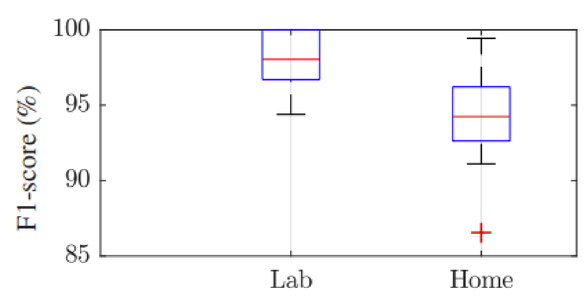
Lab and home setup

- 3 devices: UMA-8 USB mic array, ReSpeaker Core v2.0, and ReSpeaker USB mic array;
- 3 wake words: "Hey Assistant!", "Computer" and "Amazon";
- 3 different time frame: day, week and month;
- 2 rooms: lab and home;
- 3 distances: 1 meter, 3 meters and 5 meters;
- 14 angles: 0◦, +15◦, -15◦, +30◦, -30◦, +45◦, -45◦, +60◦, -60◦, +90◦, -90◦, +135◦, -135◦, 180◦.

Overall Accuracy



Performance of different devices



Performance of different environments

Takeaway

- Voice Command can be used to sense head orientation for privacy control
- Potential to make distributed voice interactions more practical and secure

NC STATE UNIVERSITY

More detail to download the full draft of the paper

WSPR