Poster: Natural Voice Interface for the Next Generation of Smart Spaces

Yang Bai, Nakul Garg, Harshvardhan Takawale, Anupam Das†, Nirupam Roy
{yangbai8,nakul22,htakawal}@umd.edu, †anupam.das@ncsu.edu, niruroy@umd.edu
University of Maryland College Park, †North Carolina State University

ABSTRACT

One of the popular trends in the human-machine interface is voice activation, and it is going beyond voice assistants (VAs) like Amazon Echo and Google Home. The latest smart TVs from Samsung and Honeywell's thermostat include a voice interface to understand voice commands for hands-free operations. However, while leading toward the smart home future, misactivations can also happen when multiple VAs share the same space. Such misactivations lead to the unintended uploading of sensitive user data. Thus, it is critical for voice-enabled devices to sense their physical surroundings and better understand the context in which they should interact.

CCS CONCEPTS

• **Human-centered computing → Interaction design process and methods**.

KEYWORDS

Direction of voice; Acoustic sensing; Embedded intelligence

Therefore, the goal of this work is to develop a device-free, non-obtrusive acoustic sensing technique to infer the direction of the voice and thereby associate addressability with voice commands. We also capture signal-level physical cues during conversation to implicitly determine the different contexts in which VAs should activate, allowing VAs to record and transmit audio data only when expected. In other words, we will develop signal processing building blocks that can extract the facing direction of the speaker when issuing a voice command. The notion of the physical context of voice commands through acoustic signal processing is shown in Figure 1.

Moreover, to provide an approach that offers better assurance for privacy, we embrace the embedded intelligence approach that makes resource-constrained embedded systems capable of processing and inference on-device without sharing raw data to external servers. New algorithms are proposed to enable intermittently powered, ultra-low-power computing devices capable of executing complex acoustic sensing and perception tasks. Our proposed research is comprised of two major tasks: i) Device Addressability, and ii) Conversational Interface. **Device addressbility.** The main objective of device addressability is to determine if a human speaker is facing the device. Our proposed head orientation detection approach is inspired by interpersonal communication cues humans exhibit while speaking to one another. We, therefore, propose building an acoustic sensing technique to find the multipath profile of the room and then combine it with the typical directionality pattern of voice signal around the human head to estimate if the speaker is facing toward a device. We break away from the traditional approach of covering the surface with an array of sensors and develop a single-sensor technique leveraging a novel coded signal projection method through passive microstructures at the physical layer [1, 2].

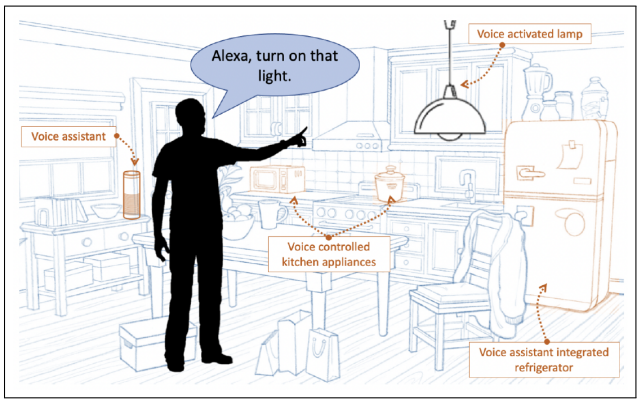

Figure 1: An application of the natural voice-based interaction with the smart spaces.

Conversational Interface. One bottleneck in combining multiple command exchanges into a logical block of conversation is the difficulty in identifying the start and end of the conversation. We see that it is possible to relate a physical context, like the facing direction, with the speech signal to identify such conversation boundaries. We propose a Generative Adversarial Network (GAN) [3] that considers the facing direction and speech together to identify the context of a phrase and relate them to the interaction with the VA. The model is structurally similar to the boundary-seeking GAN models, but it introduces the facing direction through a fusion model.

REFERENCES

[1] Yang Bai, Nakul Garg, and Nirupam Roy. Spidr: Ultra-low-power acoustic spatial sensing for micro-robot navigation. In *The 20th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '22)*. ACM, 2022.
[2] Nakul Garg, Yang Bai, and Nirupam Roy. Owlet: Enabling spatial information in ubiquitous acoustic devices. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, pages 255–268, 2021.
[3] Irtaza Shahid, Yang Bai, Nakul Garg, and Nirupam Roy. Voicefind: Noise-resilient speech recovery in commodity headphones. In *Proceedings of the 1st ACM International Workshop on Intelligent Acoustic Systems and Applications*, pages 13–18, 2022.

HotMobile '23, February 22–23, 2023, Newport Beach, CA, USA

ACM ISBN 979-8-4007-0017-0/23/02…$15.00
https://doi.org/10.1145/3572864.3581587