Voice Command Object Localization with Spatial Audio and IoT Devices

Lucy Song, Computer Science
Mentor: Dr. Robert LiKamWa, Assistant Professor
School of Computing and Augmented Intelligence, School of Arts, Media, and Engineering

ABSTRACT

- **Spatial audio** can be especially useful for directing human attention.
- **Internet of Things (IoT)** offers more connectivity and interactions with items in an environment.
- Implementing spatial audio through speakers is difficult due to **crosstalk** issue.
- We have created an algorithm, **Xblock**, that implements crosstalk cancelation for spatial audio.
- We expand upon our existing spatial audio IoT infrastructure with **voice command features** for object finding.

MOTIVATION

Navigation	Guiding users to a destination
Memory Recall	Helping users remember where they placed items
Internet of Things	More connectivity and interactions between users and objects in environment

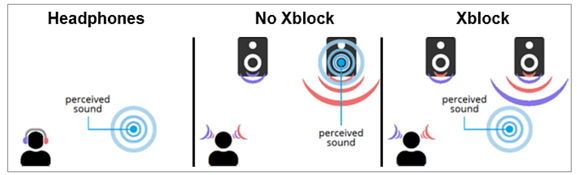
PROBLEM STATEMENT

How can we integrate a **voice command** feature into an IoT smart speaker infrastructure to use spatial audio to guide users to items in an environment?

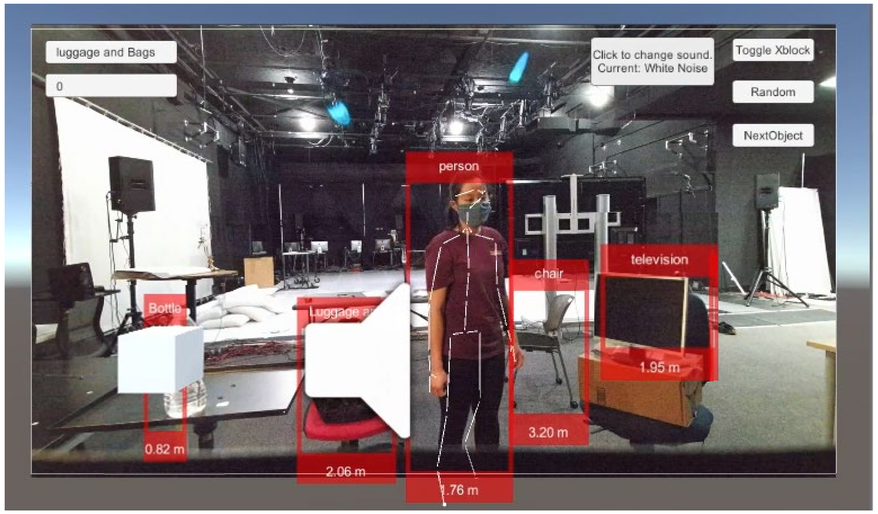
METHODS

Xblock

Algorithm that implements crosstalk cancellation technology for spatial audio.



Object Detection + Voice Commands



Example Command: "Find bag"

METHODS: MATERIALS

- Unity 3D
- Microsoft Kinect Sensor
- Raspberry Pi
- 2 Stereo Speakers
- 4 Objects (Chair, TV, Bag, Bottle)
- KeywordRecognizer

FUTURE WORK

- Integrate voice commands into user testing to validate Xblock's effectiveness.
- Explore spatial audio for narrative storytelling.

REFERENCES

[1] Gröhn, M., Lokki, T., & Takala, T. (2005). Comparison of auditory, visual, and audiovisual navigation in a 3D space. ACM Transactions on Applied Perception (TAP), 2(4), 564-570.

[2] Jing Yang and Gábor Sörös. 2018. Spatial Audio for Human-Object Interactions in Small AR Workspaces. In Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '18). Association for Computing Machinery, New York, NY, USA, 518. DOI:https://doi-org.ezproxy1.lib.asu.edu/10.1145/3210240.3210811

[3] Hanes, D., Salgueiro, G., Grossetete, P., Barton, R., & Henry, J. (2017). IoT fundamentals: Networking technologies, protocols, and use cases for the internet of things. Cisco Press.

[4] Frank Liu and Robert LiKamWa. 2019. Demo: A Spatial Audio System for the Internet-of-Things. In Proceedings of the 20th International Workshop on Mobile Computing Systems and Applications (HotMobile '19). Association for Computing Machinery, New York, NY, USA, 183. DOI:https://doi-org.ezproxy1.lib.asu.edu/10.1145/3301293.3309567

[5] "Azure Kinect body tracking joints," Microsoft Docs. [Online]. Available: https://docs.microsoft.com/en-us/azure/kinect-dk/body-joints. [Accessed: 07-Mar-2022].

ACKNOWLEDGEMENTS

This work was made possible with funding from the Fulton Undergraduate Research Initiative (**FURI**) and the **ASU Meteor Studio** research lab.

Thank you to my faculty mentor, Dr. Robert LiKamWa, and my graduate student mentor, Frank Liu, for their support and guidance.

FURI

Ira A. Fulton Schools of Engineering
Arizona State University