

SonifyAR: Context-Aware Sound Generation in Augmented Reality



Xia Su



Jon E. Froehlich



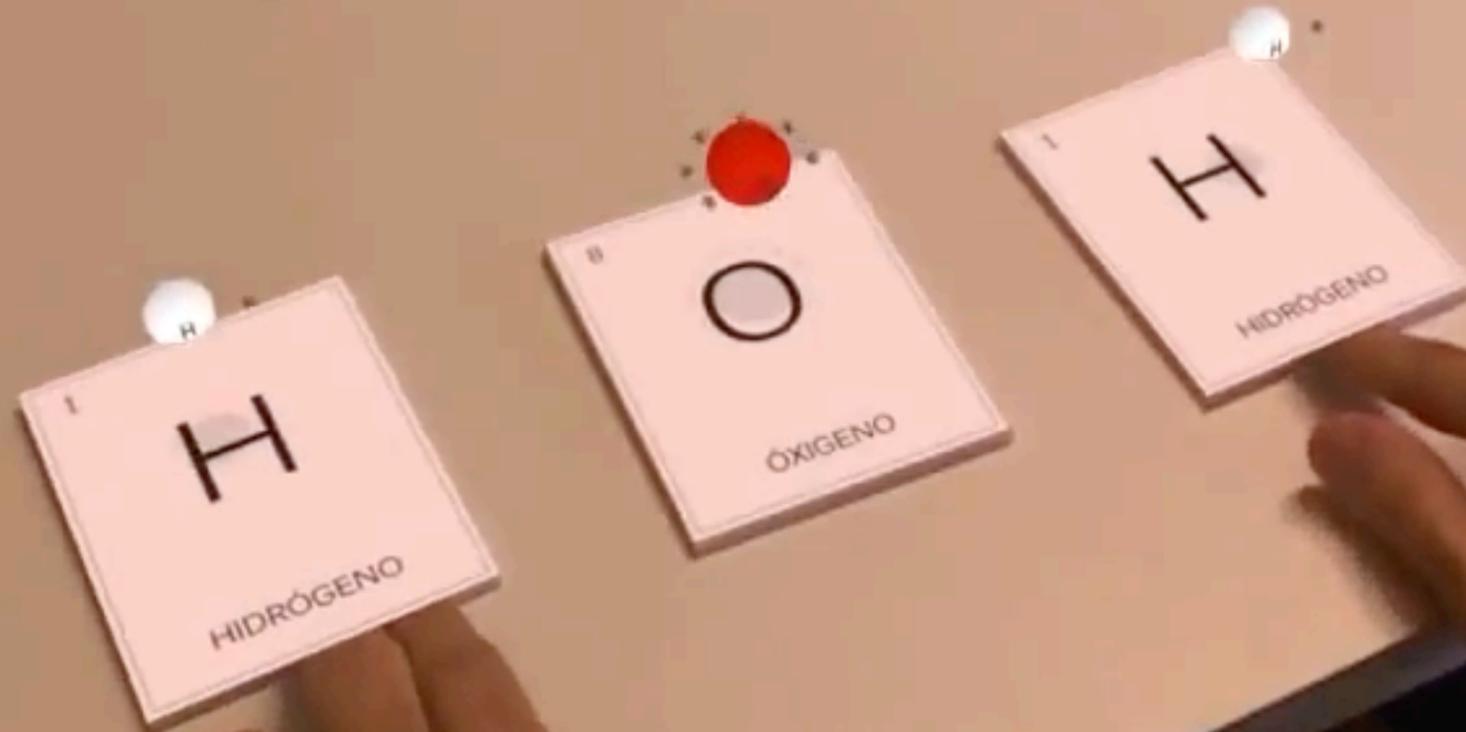
Eunyeek Koh



Chang Xiao

Intro

Why AR sound matters



Intro

Why AR sound matters



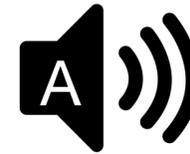
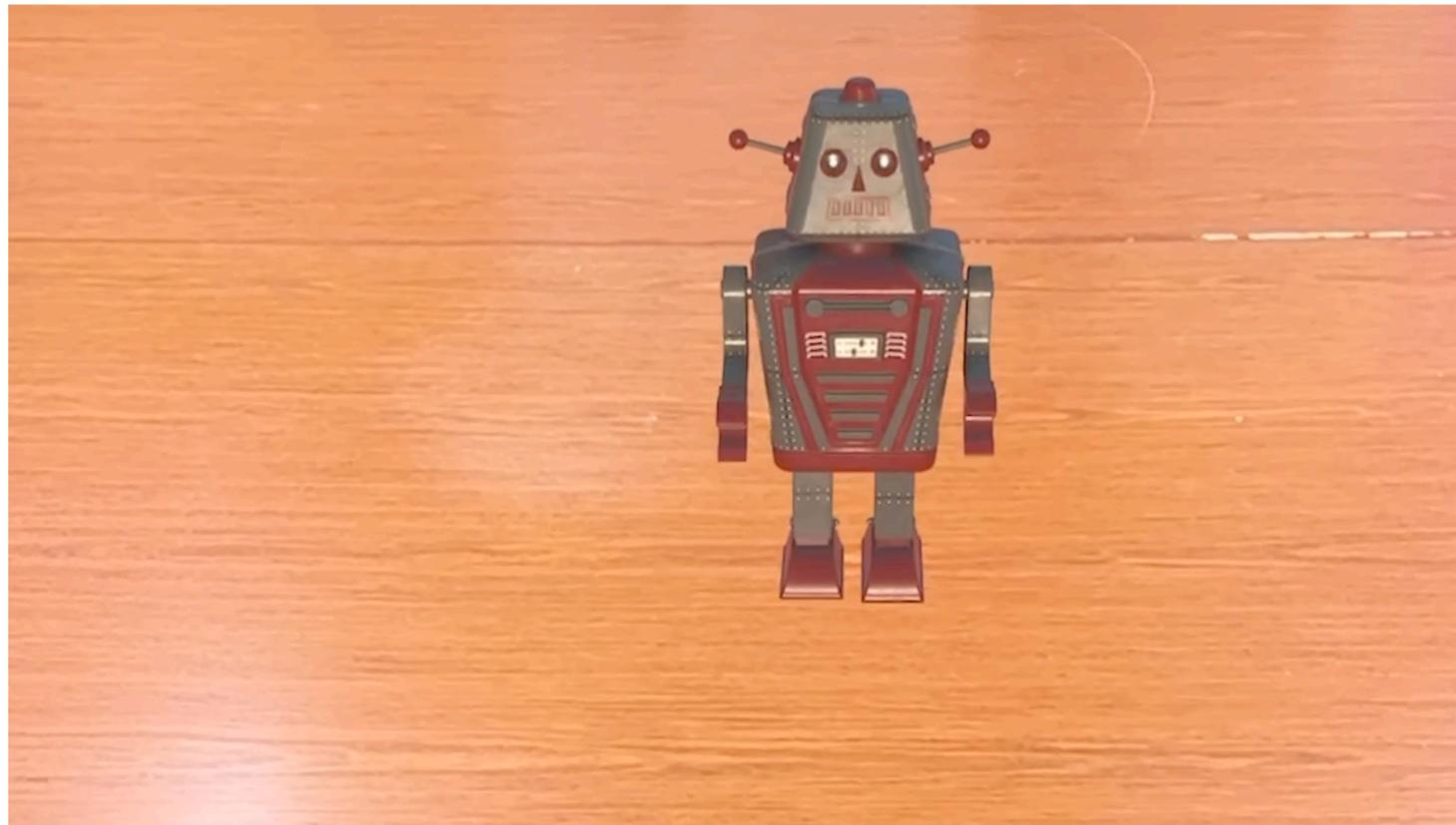
Intro

Why AR sound matters



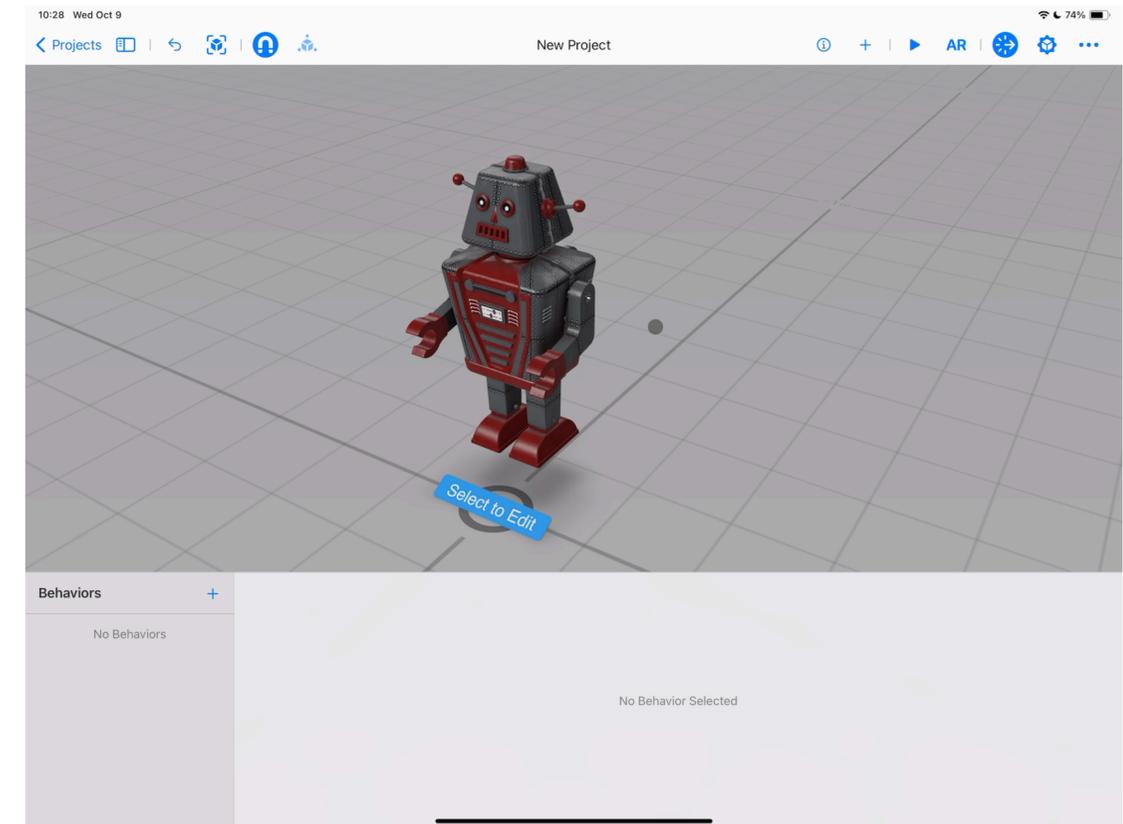
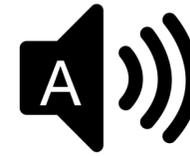
Challenges

Limited real-world context



Challenges

Limited real-world context



Challenges

Limited interaction types



Namespace Unity.MARS.Actions

Classes

[BuildSurfaceAction](#)

[FaceAction](#)

[FaceExpressionAction](#)

[FaceExpressionEvent](#)

[MatchAction](#)

[MatchArgsEvent](#)

[MatchResultEvent](#)

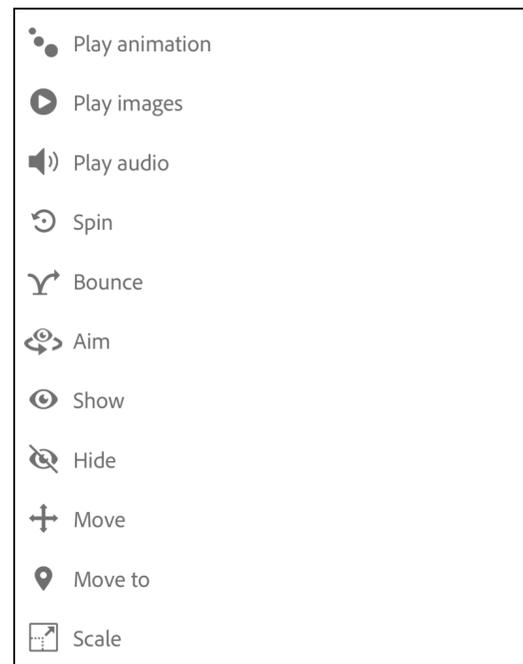
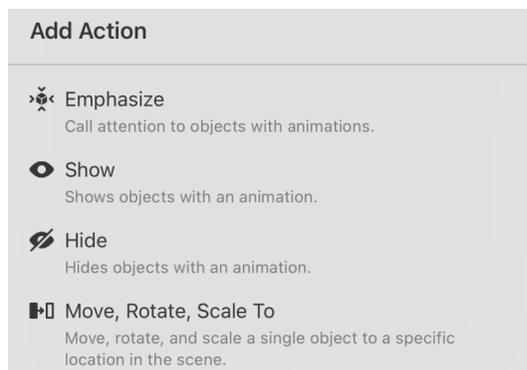
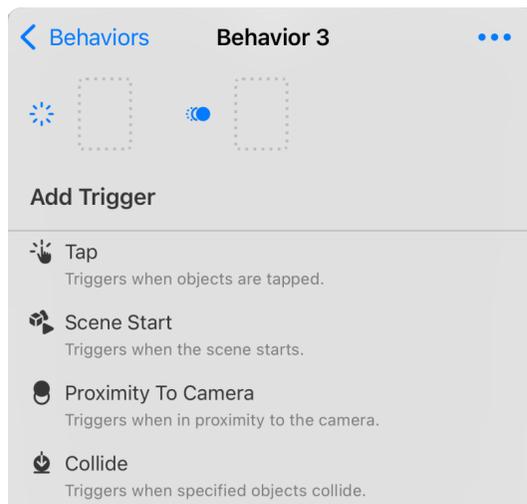
[SetAlignedPoseAction](#)

[SetPoseAction](#)

[ShowChildrenInBoundsAction](#)

[ShowChildrenOnTrackingAction](#)

[StretchToExtentsAction](#)



Challenges

Limited sound sources

Local or online sound library

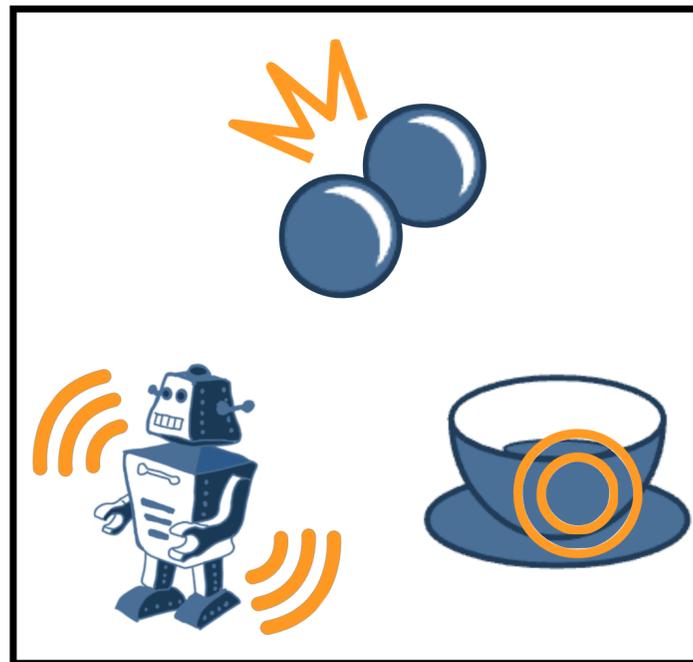
	▶ Movie Set - Digital Multitrack Sound Recorder Beep Soundkrampf	Interfaces, Devices
	▶ Detector Sounds - Clacking Sonar, Metallic Resonance Augustin Gressier	Tools, Alerts
	▶ Organic UI Sounds - Lips Pop, High Sound Ex Machina	Bodily Functions, Cartoon Percussive, Alerts, Menus
	▶ Grandpa Reactions - Laugh Out Loud, LOL Artist Foley	Men, Reactions
	▶ Metal box, moves, metallic sound Foley Walkers	Metal
	▶ Hyundai Kona Electric Car - Approaching, Virtual Engine Sound System, Stopping Alexander Gastrell	Cars, Interfaces, Alerts
	▶ Cinematic Feel - Sci Fi Sound Wave, Fluctuating, Vibrating Mechanical Wave	Sci-Fi Machines, Electronic Loops
	▶ Movie Set - Cables Connecting and Disconnecting Soundkrampf	Cameras

Record

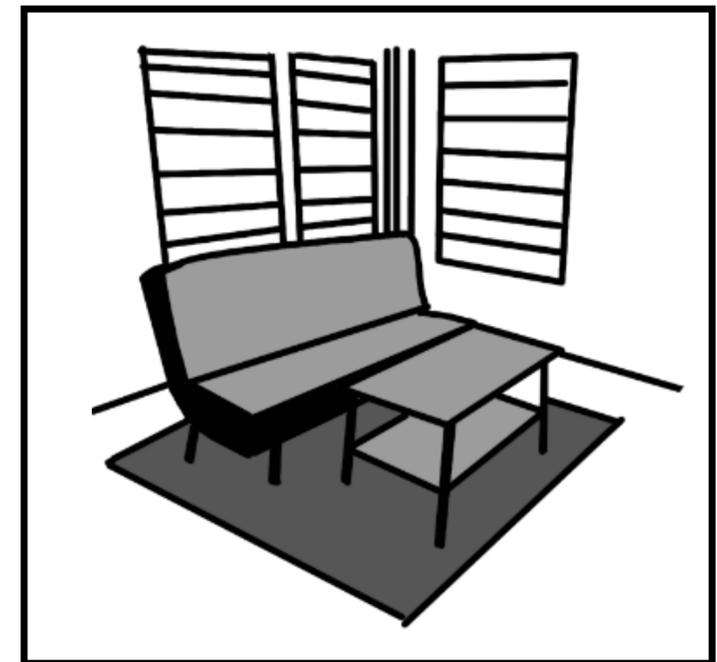




User

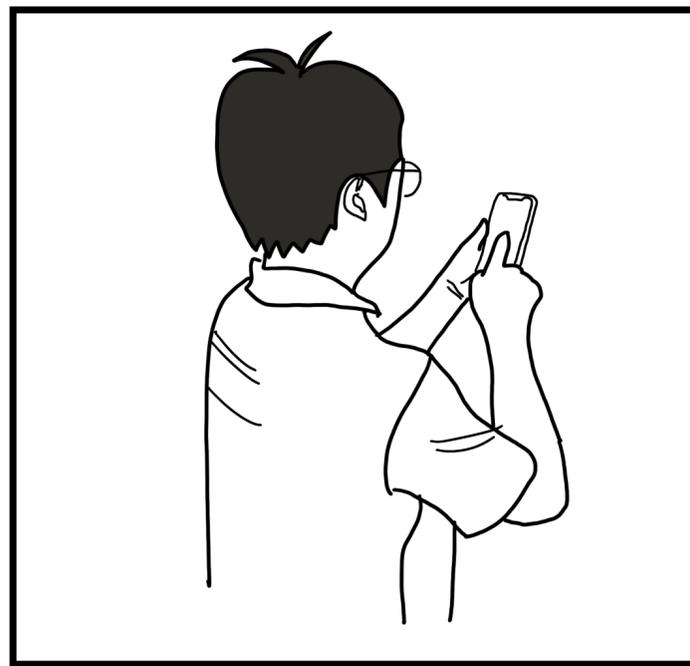


Virtuality



Reality

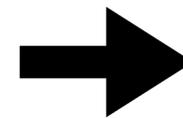
Current Tools



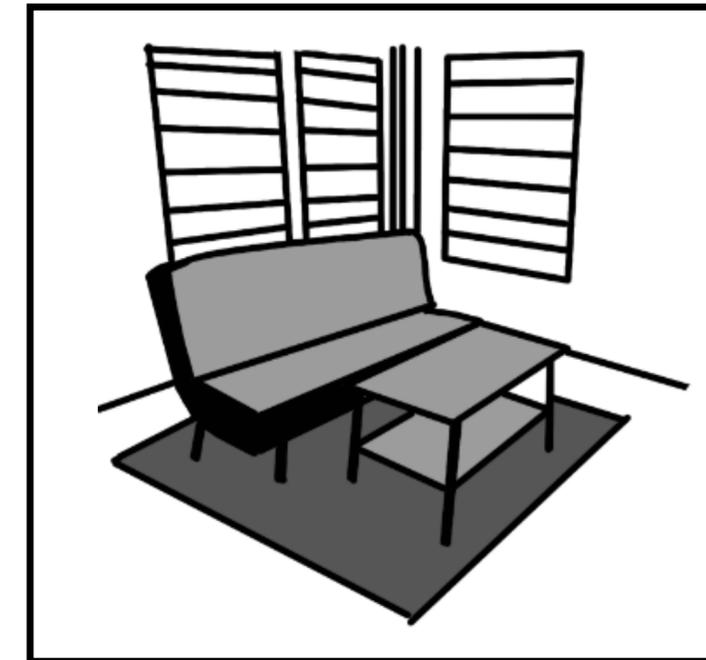
User



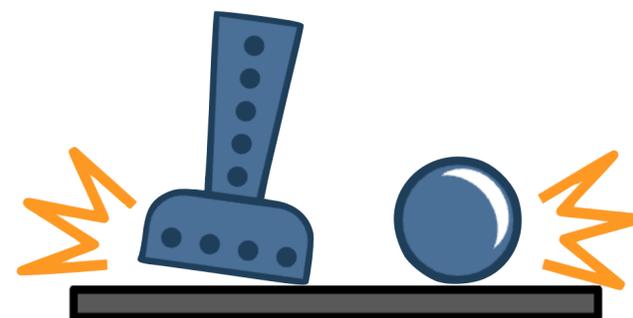
User-Reality



User-Virtuality-Reality



Reality

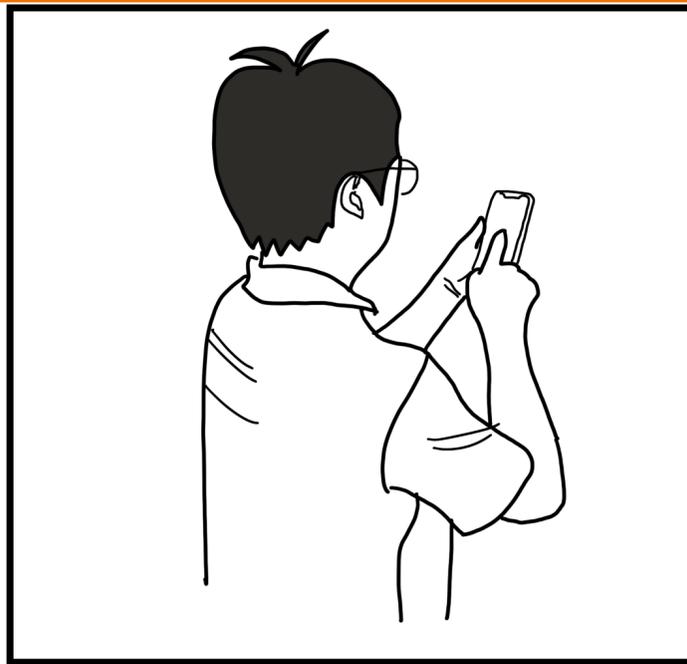


Virtuality-Reality

SonifyAR



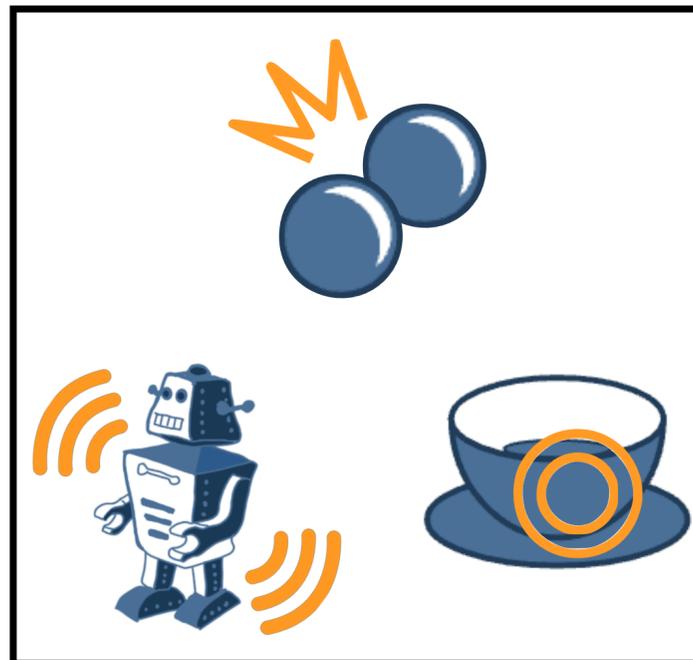
User-Virtuality



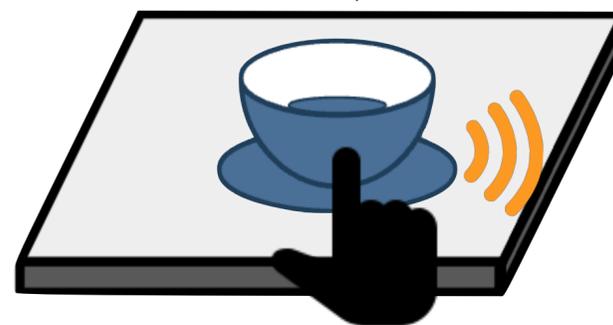
User



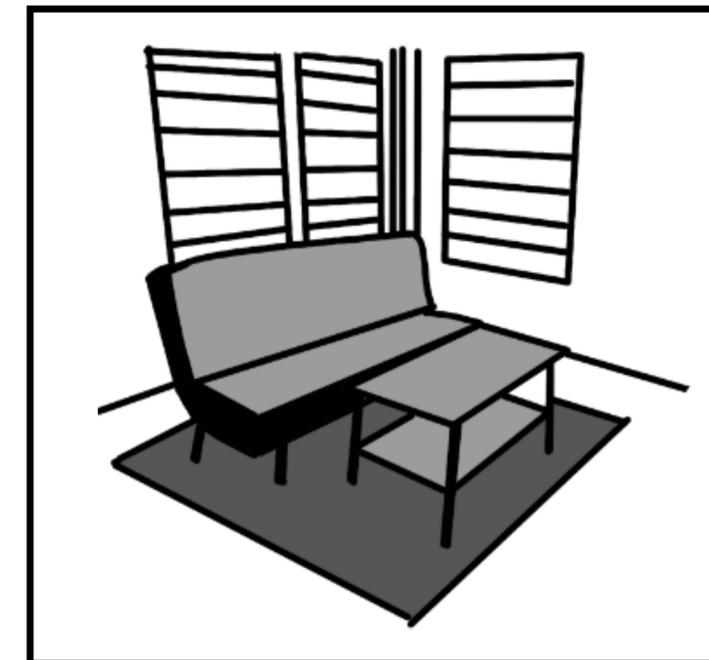
User-Reality



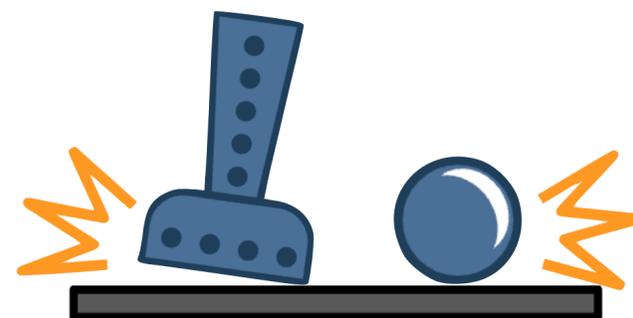
Virtuality



User-Virtuality-Reality



Reality

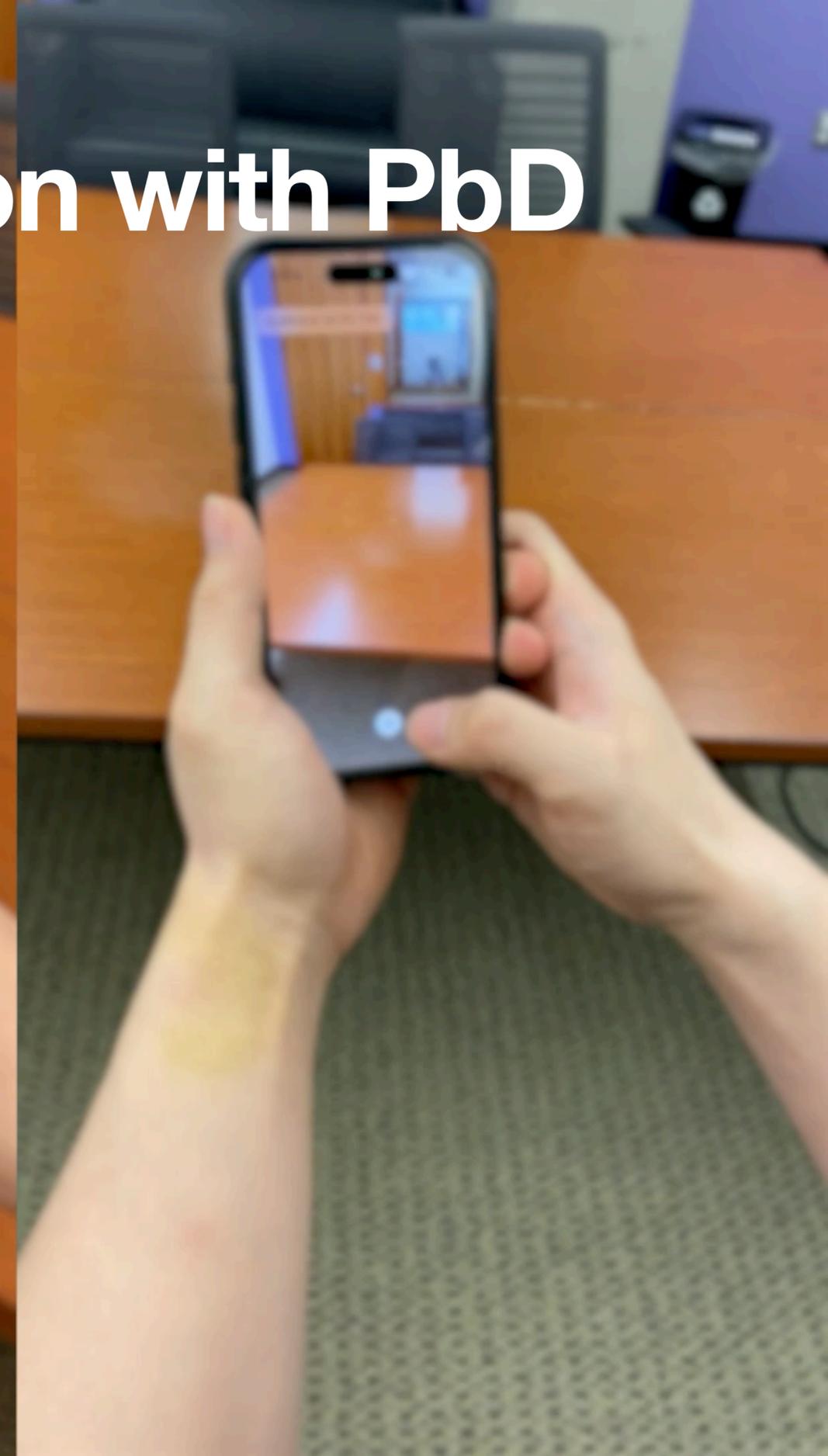
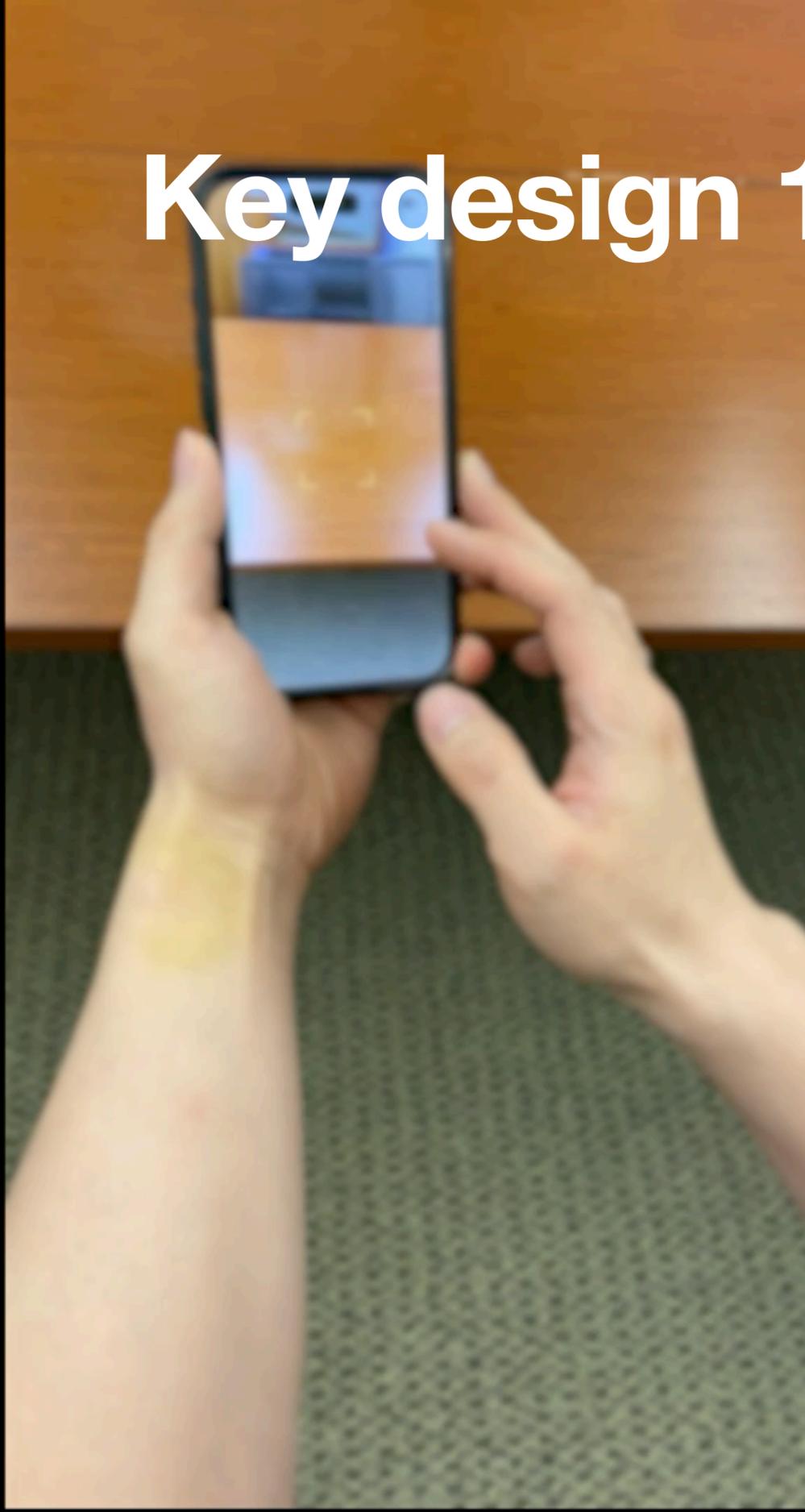


Virtuality-Reality

*Context-
aware
AR sound
authoring*



Key design 1: Context Collection with PbD



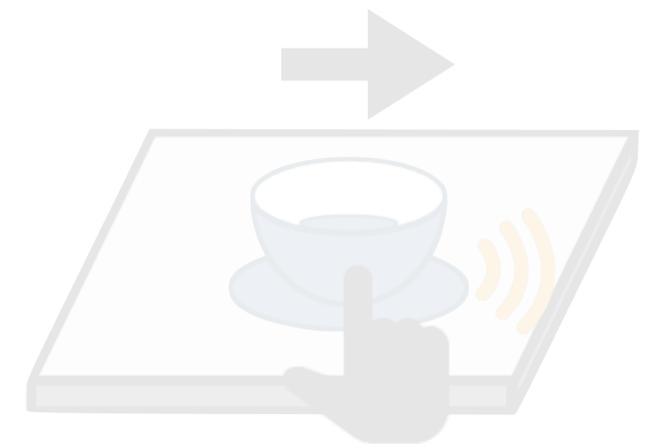
Key design 1: Context Collection with PbD



Reality



Upchurch, Paul, and Ransen Niu. "A dense material segmentation dataset for indoor and outdoor scene parsing." *European conference on computer vision*. Cham: Springer Nature Switzerland, 2022.



Event Type

How to holistically incorporate these context information?



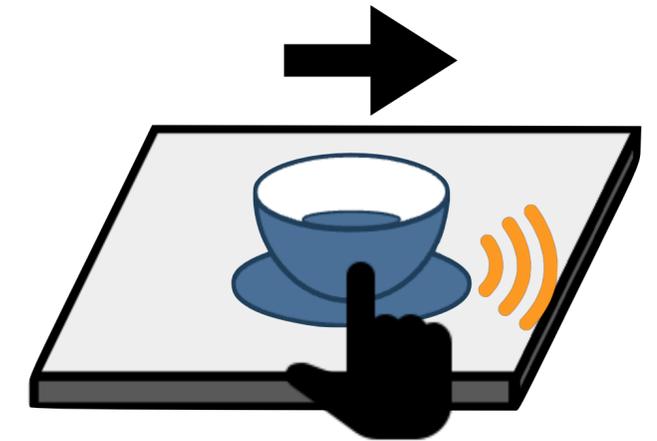
Virtual Object



Reality



User



Event Type

Key Design 2: Context as Text



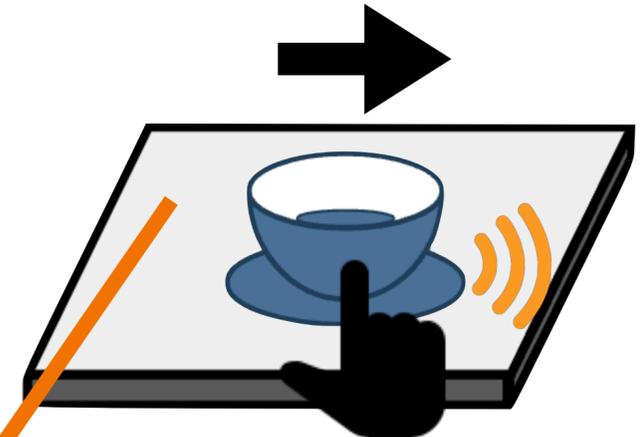
Virtual Object



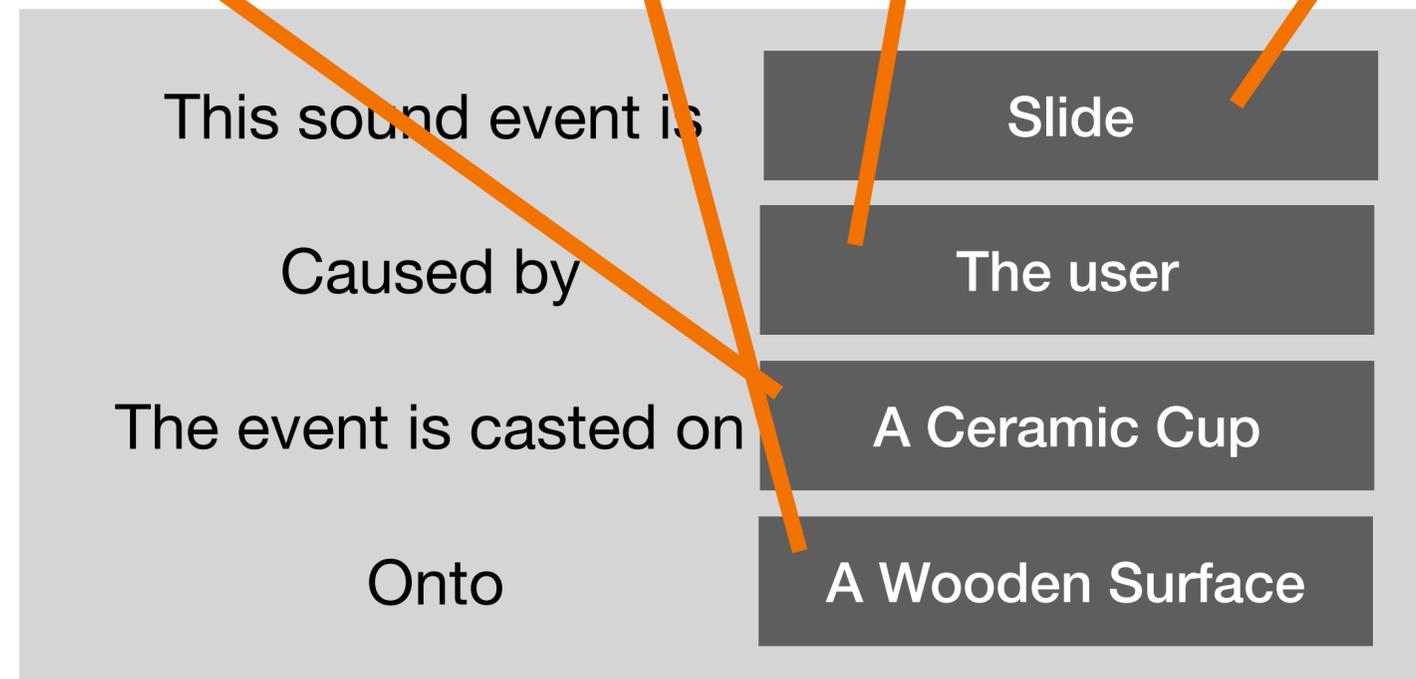
Reality



User

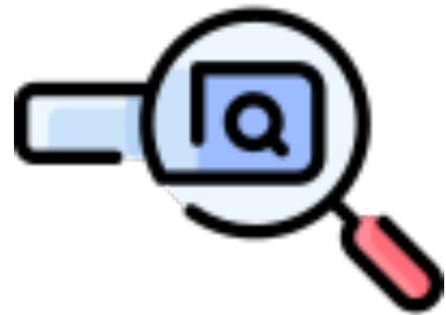


Event Type



Tradeoffs: Retrieval vs Generation

+Quality



Retrieve

-Coverage

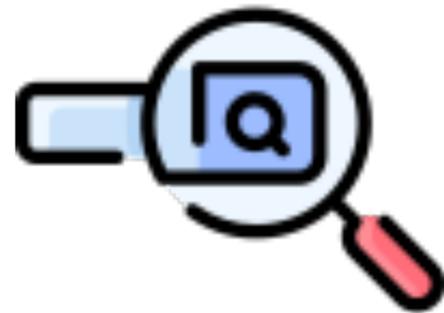
This sound event is	Slide
Caused by	The user
The event is casted on	A Ceramic Cup
Onto	A Wooden Surface

+Adaptation



Generate

-Quality



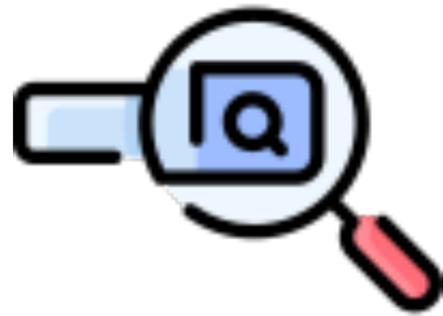
Retrieve

&



Generate

Key Design 3: LLM for generation and retrieval



Local
Recommend



Online
Retrieve



Text2Sound
Generate

Text2Sound
Transfer

Recommended



Retrieved



Generated

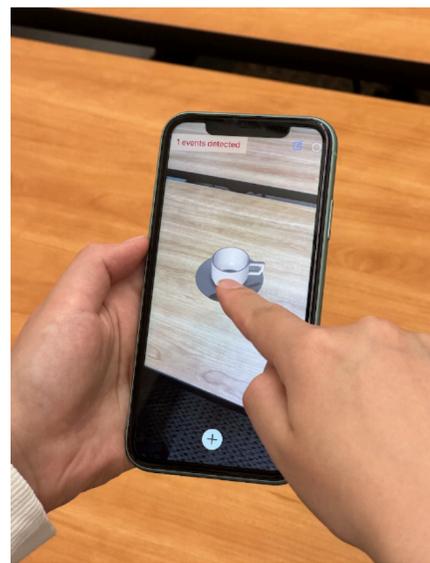


Transferred



Key Design 3: LLM for generation and retrieval

Local Recommend

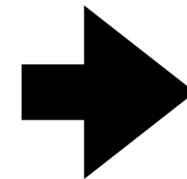
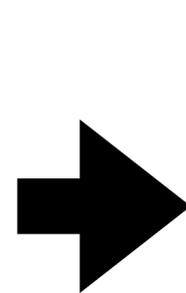


Tap Virtual Model

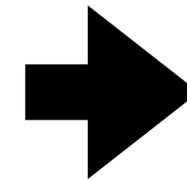
Caused by **User**

The event is casted on

a Ceramic Tea Cup



Top 5 Filenames:
Knock Surface
Crash Bulb Break
Crash Aluminum
Garbage Can Dropping
Baseball Bat Strike



Recommended

Local sound library

- Aluminum Baseball Bat Striking;
- Garbage Can Dropping On Pavement;
- Bongo Hit Drum;
- Blow Torch Gas Release;
-

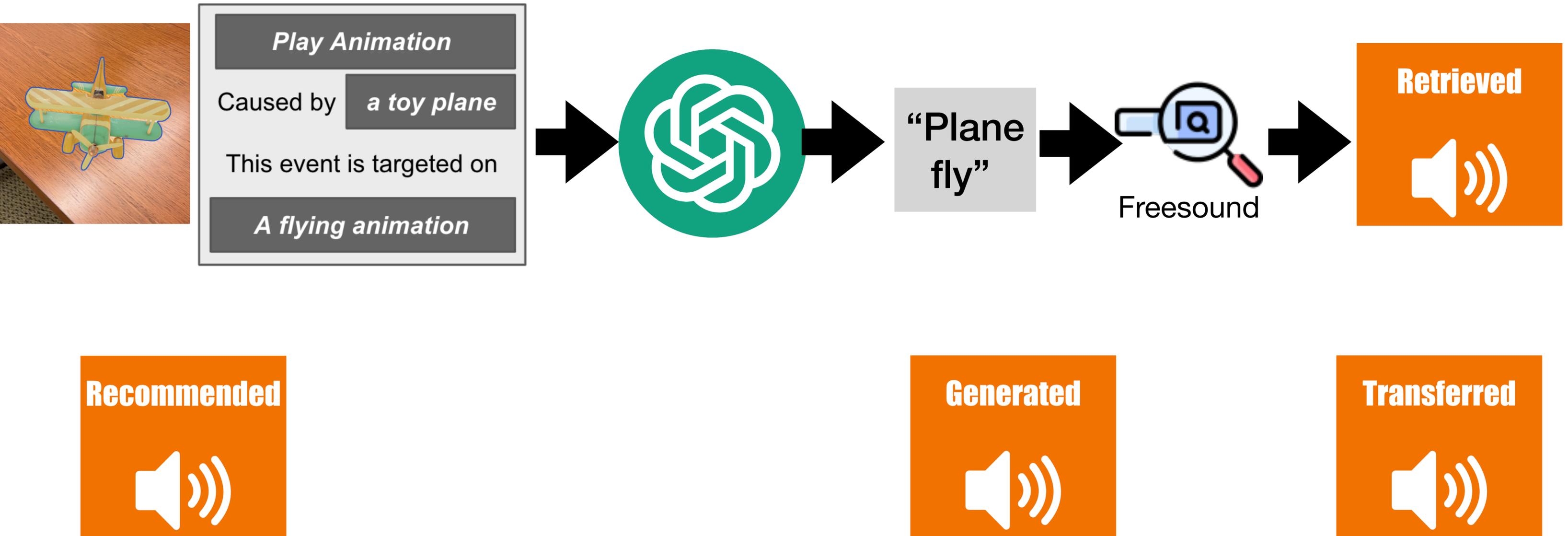
Retrieved

Generated

Transferred

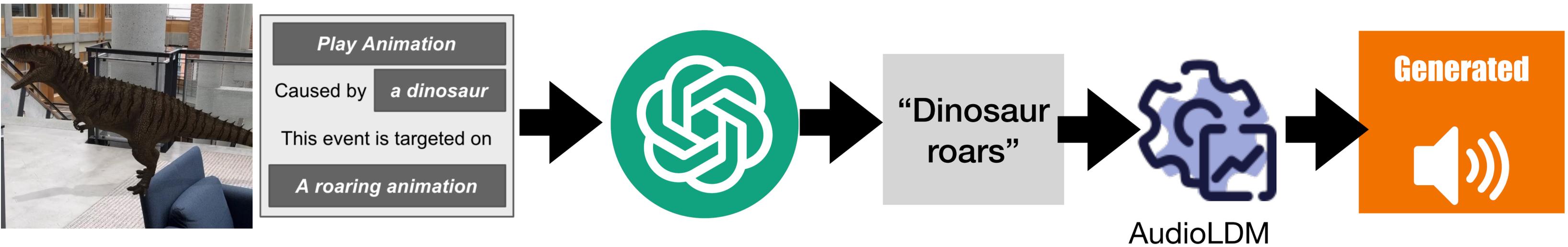
Key Design 3: LLM for generation and retrieval

Online Retrieval



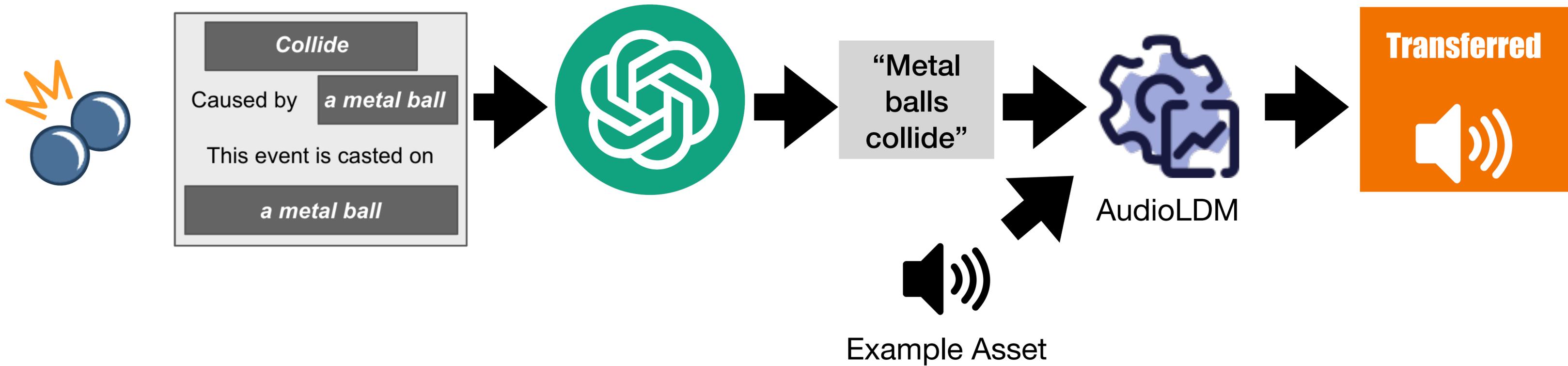
Key Design 3: LLM for generation and retrieval

Text2Sound Generate



Key Design 3: LLM for generation and retrieval

Text2Sound Transfer



Recommended



Retrieved



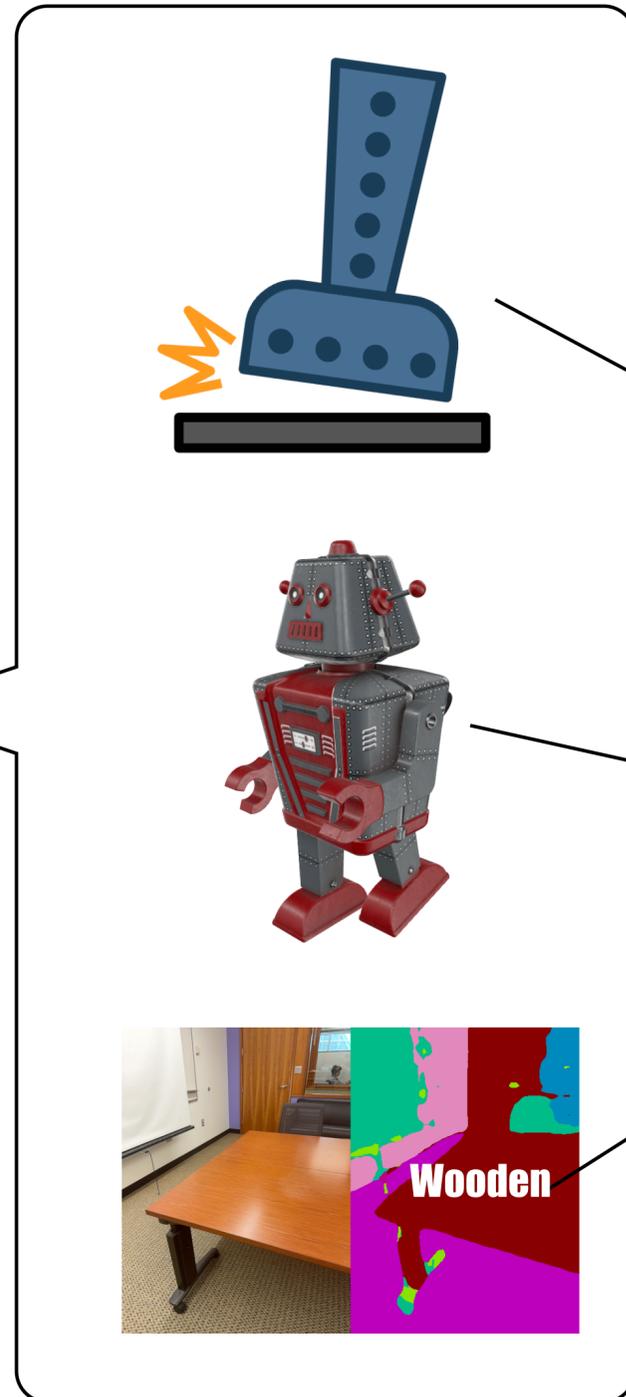
Generated



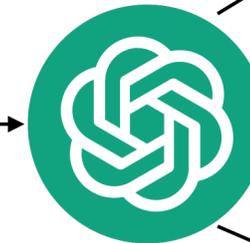
SonifyAR Pipeline



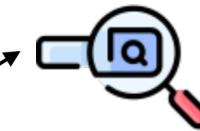
AR Event



This sound event is
Collide Caused by
Virtual model
Which is
A toy robot made of metal
This event is casted on
A wooden plane



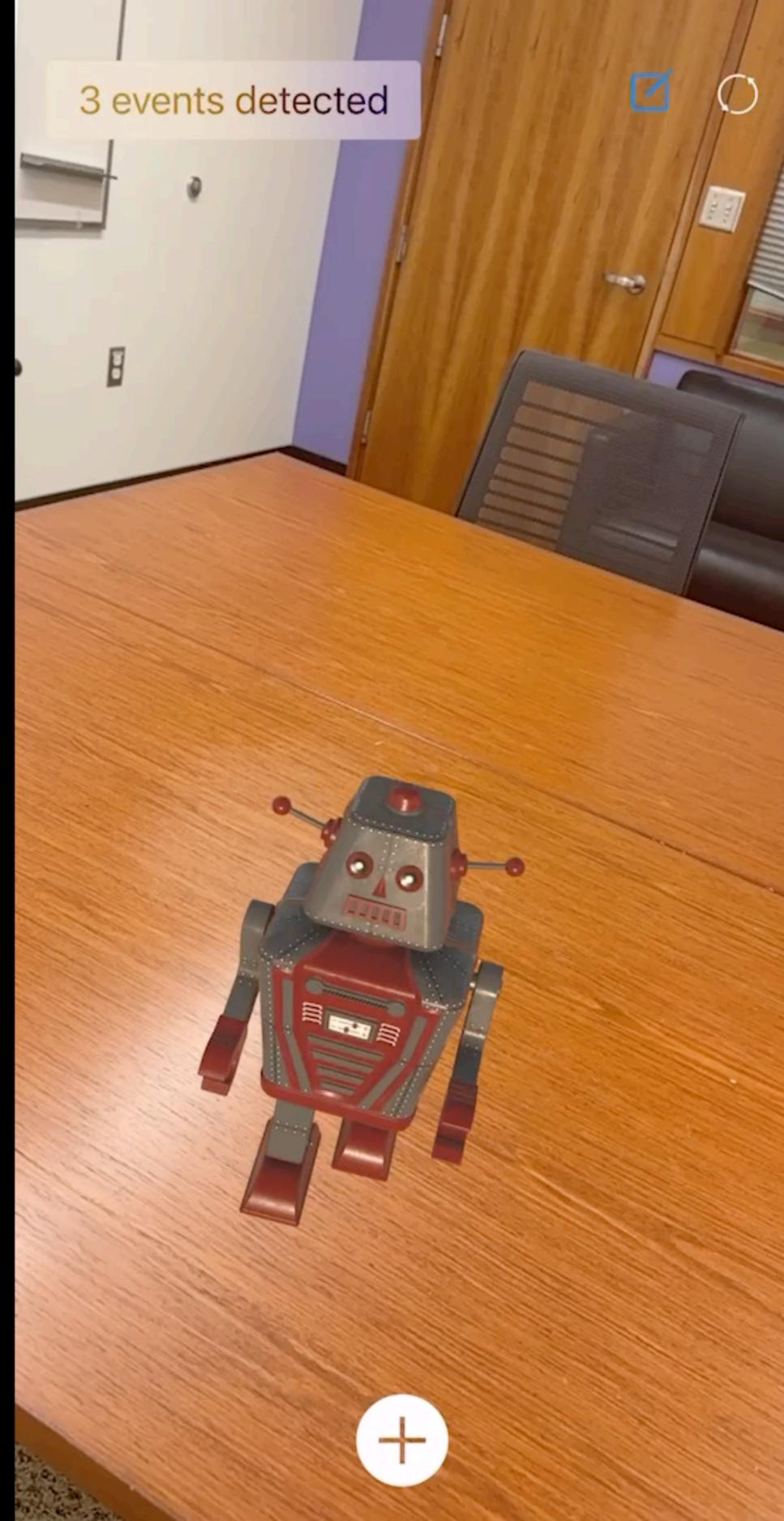
Local/online
Retrieval



Text2Sound Generation
/Transfer



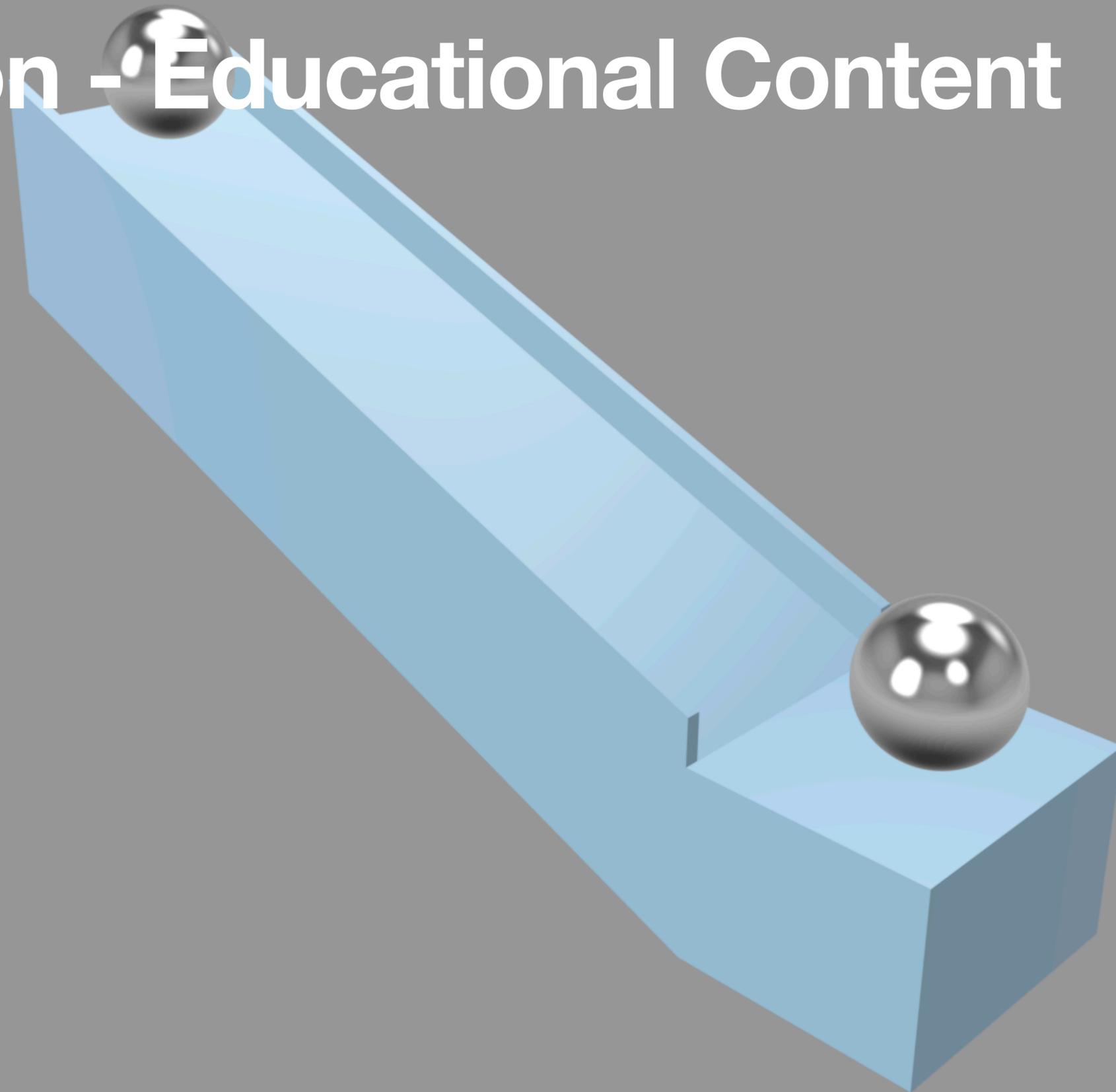
SonifyAR UI



Application - Animation Content



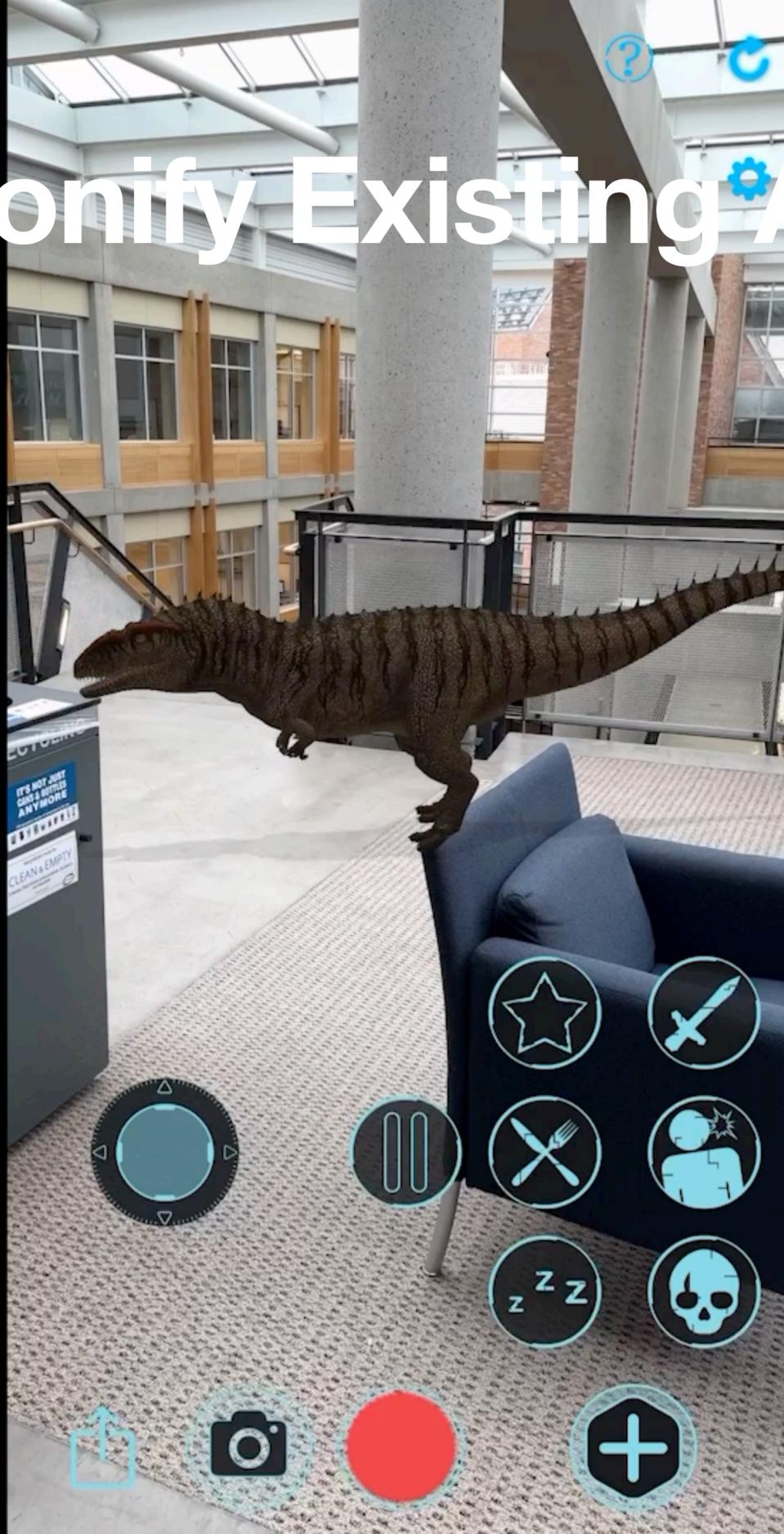
Application - Educational Content



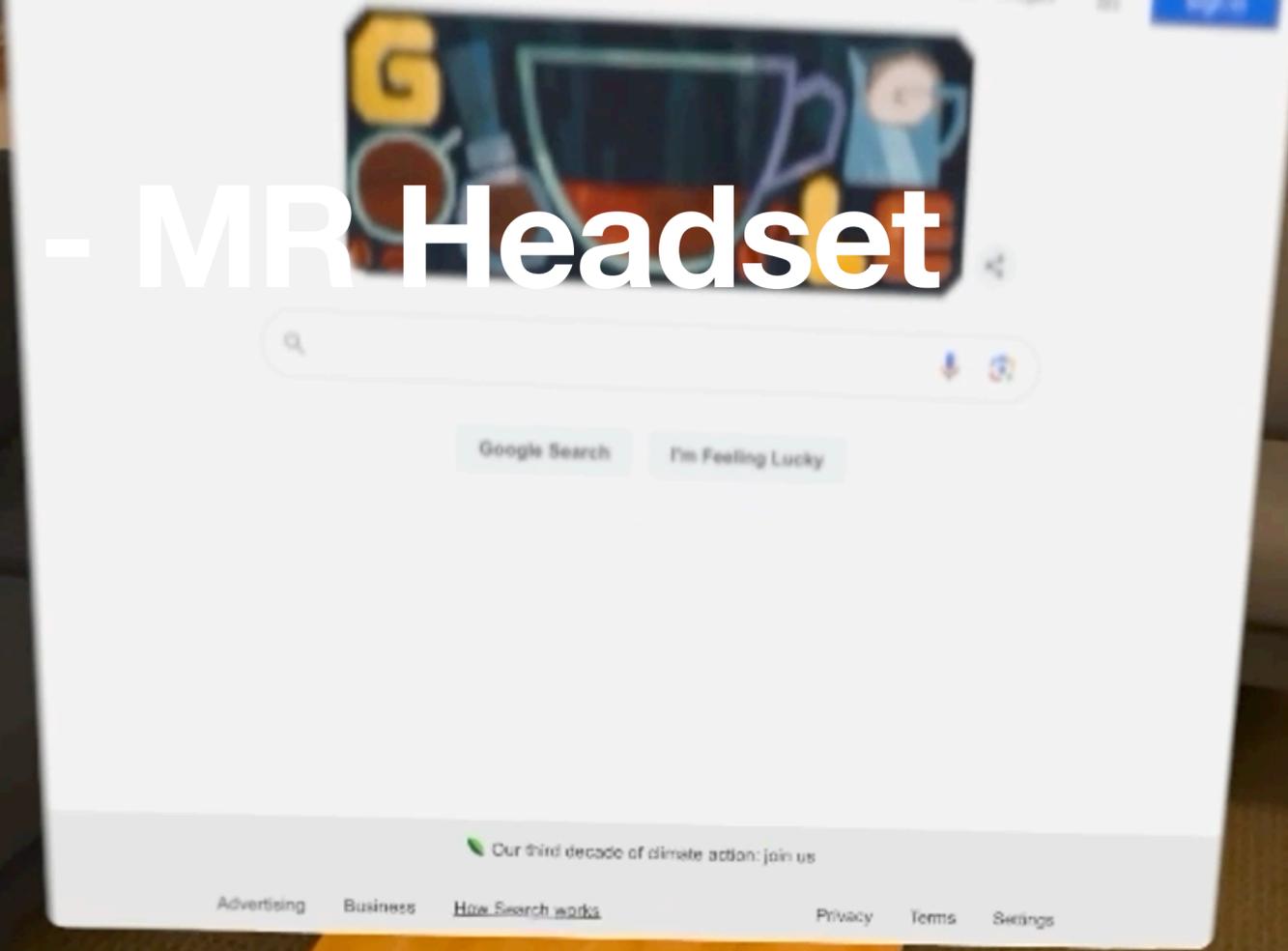
Application - Accessibility Potential



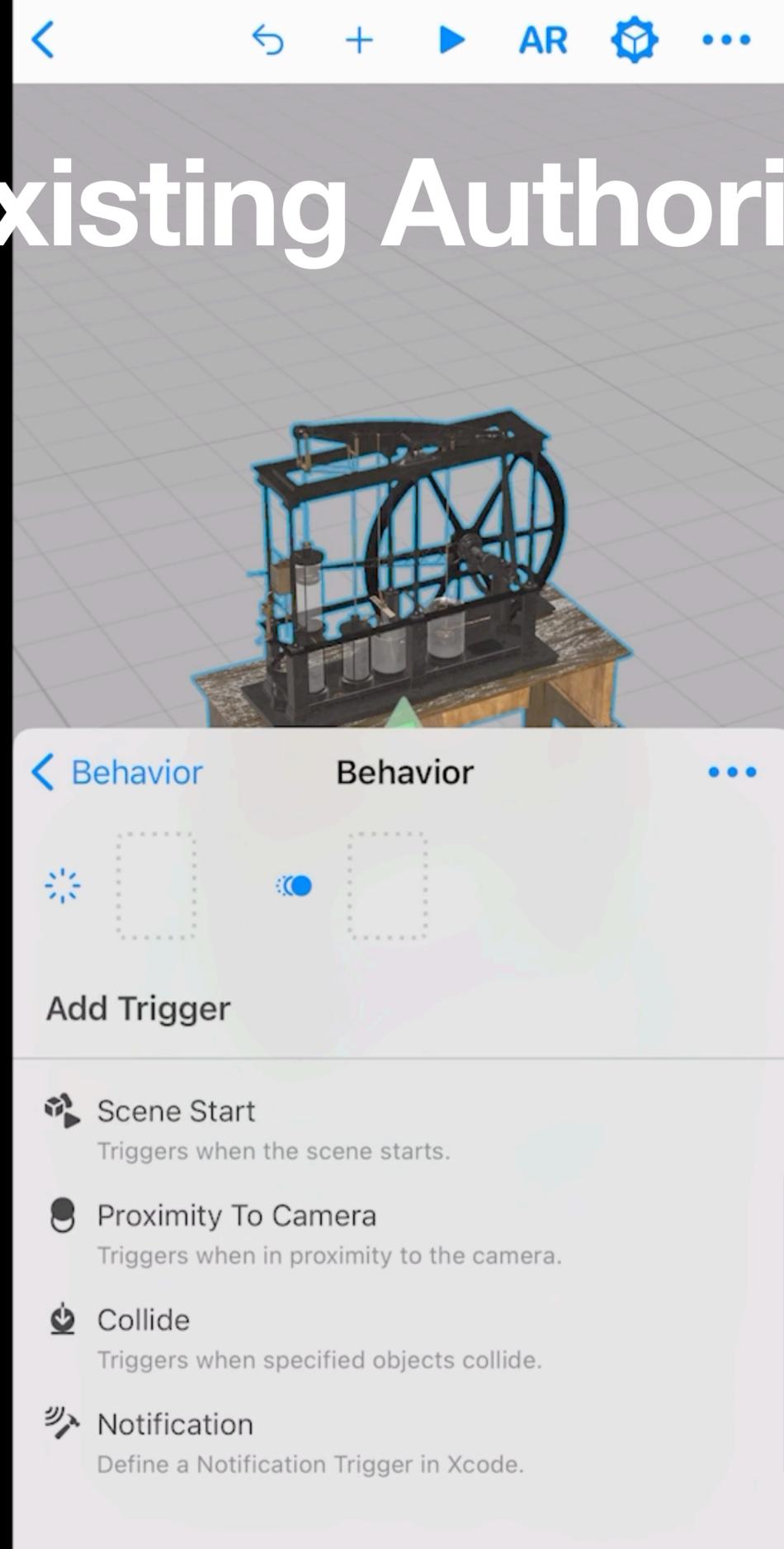
Application - Sonify Existing App



Application - MR Headset

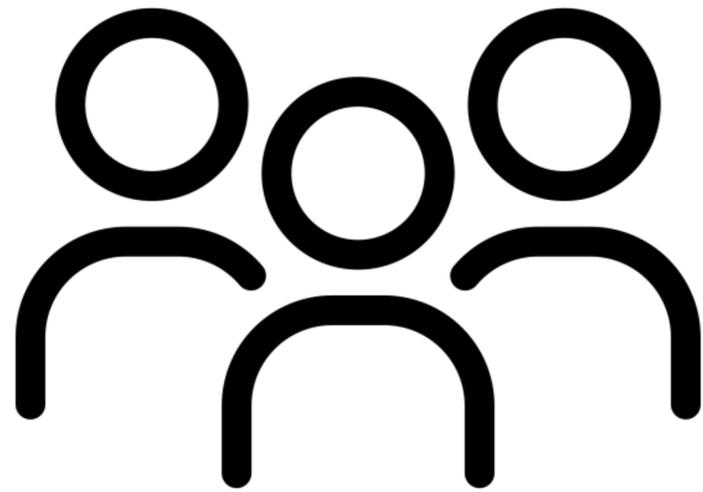


Application - Existing Authoring Tool



User Study

Stats



8 participants

Helpfulness



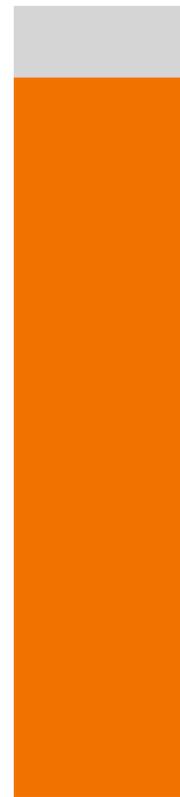
6/7

Willingness
to use



6.3/7

Improve
immersiveness



6.4/7

Preference



6.1/7

Sound
Quality



4.8/7

Limitations & Future Work

- LLM output unstable
- Sound quality
- Time sync
- Merge with existing tools
- Expand environment understanding
- Automated error handling

SonifyAR: Context-Aware Sound Generation in Augmented Reality



Xia Su



Jon E. Froehlich



Eunye Koh



Chang Xiao



Video Figure

Looking for summer 2025 internship!

