

Audio Cues in Virtual Environments

f

[Community Discussion 9/22](#)

[Event Description](#)

[Resources](#)

[Participants](#)[Resources](#)

[Discussion Topics](#)

[Notes](#)

Community Discussion 9/22

Event Description

[Event Link](#)

The physical world offers us a rich tapestry of audio sounds to help us navigate our world. The tapping of footsteps, rumble of traffic, or gentle creak of a door opening give us many clues to our environment. However, virtual spaces often lack these cues, making them difficult to use for people who are blind or visually impaired.

On 9/22 The XR Access Initiative and the [XR Association](#) are hosting a 1.5 hour community discussion on audio cues in virtual environments. Together, let's explore lessons we've learned with audio cues in physical life and audio games and explore how those can be applied in virtual experiences to accelerate accessibility deployment in XR. [Register for the conversation here.](#)

Resources

Audio Technology Resources:

Spatial Sound: [Spatial Sound for app developers for Windows, Xbox, and HoloLens 2 - Win32 apps | Microsoft Learn](#)

Project Acoustics: [Project Acoustics Overview - Microsoft Game Dev | Microsoft Learn](#)

Microsoft Soundscape (audio based navigation): [Microsoft Soundscape - Microsoft Research](#)

Text to Speech: [Create Dynamic, Accessible Content with Azure Neural Voice - YouTube](#)

Example projects mentioned by Tim Stutts:

- Cognixion One assisted reality head-mounted display for communications: <https://youtu.be/TDCviys1aFg> & <https://www.cognixion.com/>
- Vuforia Capture and Vantage Work Instructions applications for HoloLens and MagicLeap: <https://vimeo.com/714471703> & <https://www.ptc.com/en/products/vuforia/work-instructions>

Additional Examples

- Audiovisual Video Games
 - The Last of Us Pt II
 - Gears of War
 - Animal Crossing villager earcons
- Audio-only games
 - Swamp
 - The Vale
 - Hero's Call
 - Blind Drive
 - Blind Knight
 - Pitch Black
 - Road to Rage
 - Shades of Doom
 - Games from audiogames.net
- Accessibility products and media
 - Seeing AI
 - Soundscape
 - [James Webb Space Telescope - images set to sound](#)
 - [Audiom](#)

XR Access & XR Association

- [XR Access website](#)
- [XR Association website](#)
- [XR Accessibility Project on Github](#)

Research

- [An evaluation of the communicative ability of auditory icons and earcons](#)
- [Age-related hearing loss](#)
- [Audio Hallway: a Virtual Acoustic Environment for Browsing](#)

Participants

Format:

Name, pronouns, organization, position

- [Dylan Fox](#), he/him, XR Access Head of Community & Outreach
- [Aaron Gluck](#), he/him, Clemson University PhD Candidate - VR Accessibility

- [Emily Kasnter](#), she/her, Meta, Program Manager
- Bob Dodd, he/him, Digita11yAccessible, Director- Delivery
- Sukriti Chadha, she/her, Product Manager, Spotify
- [Larry Goldberg](#), he/him, Accessibility consultant
- [Indira Knight](#), she/her PhD student Oxford Brookes researching XR for built environment co-creation
- Meredith Veit
- [Brandon Biggs](#), He/him, CEO, [XR Navigation](#); Researcher, Smith-Kettlewell Eye Research Institute; PhD Student, Georgia Institute of Technology, [contact me](#) if you want to research audio games and XR
- Stephanie Montgomery, she/her, XR Association, VP Research and Best Practices
- Jazmin Cano, she/her, Owlchemy Labs, Accessibility Product Manager
- [Peter Galbraith](#), he/him, Owlchemy Labs, Accessibility Engineer
- Peirce Clark, he/him, XR Association, Program Development Manager
- [Keita Ohshiro](#), he/him, New Jersey Institute of Technology, PhD student
- [Brian A. Smith](#), he/him, Assistant Professor of Computer Science, Columbia University; Director, [Computer-Enabled Abilities Laboratory \(CEAL\)](#)
- Andy Slater
- [Vishnu Nair](#), he/him, PhD Candidate, Columbia University / Computer-Enabled Abilities Lab (CEAL)

Discussion Topics

- What does the future of audio cues in VR look like?
- How do audio design for AR and VR differ?
- When is text to speech vs. earcons vs. other types of sound most effective?
<https://smartech.gatech.edu/handle/1853/49940>
https://www.tandfonline.com/doi/abs/10.1207/s15327051hci0401_1
- What are the biggest hurdles to getting better audio cues implemented? What is the timeline?
- For the idea of placing a beacon on a moving object is that a physical device that goes on the moving object or can that be assigned through another means to identify the moving object?
- What kinds of HRTF rendering options should users be able to choose from?
- When adding audio in VR, are there best practices we should work with? Can we reference something that describes how people may understand high pitched vs low pitched pings, the rate at which a sound plays, and the kind of sounds we should use? Are there guidelines for what not to do with audio cues?
- How to play volumetric sounds - representing roads, buildings, rivers, busses...

- Should we consider some kind of sonar for VR?
- What are your thoughts about acoustics+image recognition+AR to discover your world without vision?
- [Web XR Spec language precludes non-visual uses](#)
- is there a SDK or something that would allow integration of soundscape into an AR app?
- What work has been done on handling frequency ranges for earcons. I'm thinking about how men tend to lose their lower register as they get older. I see it in the statistic for preferences in female narrators in audio books
<https://smartech.gatech.edu/handle/1853/50593>
- can you name any xr spatial audio design aps thats accessible for blind folks?
- Q? Any tips for managing cognitive load when attempting to maximize communication of the affordances of a real environment to people with visual impairments for mobile scenarios? Example scenario; Managing too many verbal descriptions may overload the user, separating given real world audio cues with digital augmented cues to help a user.
- What are the most important sounds for the first mainstream VR platform to become accessible?
- We all know what most emojis mean, but we don't share understanding of haptic signals; how about development of "hapmojis"?
Stanley Cantrell: cantrell@gatech.edu is doing his PhD thesis on audio emojis

Notes

Saqib Shaikh presentation

- Seeing AI at Microsoft lets you use your phone to identify things around you
- Have been looking at created audio augmented reality experiences
 - Many ideas apply equally to VR and gaming
- A lot of original inspiration drawn from gaming, e.g. Shades of Doom showing audio cues via a stereo soundscape
 - The direction of the wind, the sound of your footsteps, or the direction an object is coming from all matter
 - You can tell from the echo when you're approaching a wall
- Seeing AI - started working with LIDAR to tell how far away things are, build a 3d model of the world and present it through spatial audio
 - Can use spatialized speech to e.g. have the word "chair" come from chair
 - Three modes:
 - What's right in front of me?
 - Scan, describe what's around me
 - Hear from left/right/forward/back via headphones
 - Place beacons, use sound to e.g. center yourself on a door
 - What if you placed beacons on moving things you want to follow or avoid? Or place them on a path you're following?
 - Working on head tracking, so you don't need to look where the camera is

- Similar to Microsoft Soundscape, which is focused on making soundscapes when you're walking outdoors, e.g. describing intersections and buildings as you walk
 - Uses GPS
 - Also placing audio cues in the real world
- VR is easier in some ways - you know exactly where things are

Tim Stutts presentation

- Current director of product design at Cognixion, a neurotech AR startup working on HMDs and BCIs
- Looking at examples of audio & XR - have brought in audio and sound design on many projects
- At Magic Leap, worked on team of designers designing OS
 - Spatial audio was very important, as was audio as cue/notification for things outside of user's field of view
 - Vice-chair of Leapable group for accessibility - did e.g. bluetooth key
- More recently at Vuforia
 - Products geared towards work instructions in a factory, where there can be blinding light, loud sounds, etc.; hands-free interface very important
 - Big parallel - workers in factory face similar challenges around mobility, sensory issues
- Cognixion - direct focus on accessibility; primary users are those w/ cerebral palsy and other severe motor disabilities

Robert Ridihalgh presentation

- Technical audio specialist with Xbox in Microsoft; long career in games history
- We have a slew of new technologies as audio designers
- W/ AR, we have to be accurate about our virtual representations
- Pillars of audio tools:
 - Spatial sound - this is what lets us accurately give location information about sound in a virtual world
 - Many ways to render spatially - home theater can reflect off of ceiling to give very immersive feel
 - Headphone rendering is commonly done through HRTF (head relative transfer functions)
 - HRTF very dependent on the shape of your ears, size of your head, other aspects of body shape; doesn't work the same for everyone
 - Need to give options to end users for different kinds of rendering
 - Acoustics - how sound propagates through a world
 - Is sound coming through a doorway behind me? Behind a giant rock in front of me?
 - Many ways to represent acoustics - consider e.g. early reflections, reverberation (how big is a space? How reflective/absorptive?)
 - Tech today can represent acoustics accurately - see Project Acoustics at Microsoft Research

- Calculates wave transformation of how a space would sound
 - Makes it more immersive for everyone and more usable for blind people
 - Synthesized voice (text to speech) - e.g. for screen readers
 - So much potential to describe the world
- To navigate through a virtual world, it's not enough to just hear ambient sounds; need a beacon system or similar
- Also need to know if my path is blocked
 - Can do this with acoustics
 - See Microsoft Soundscape
- Haptics is a very important tool as well, esp for hands
 - Work in a similar space to audio - frequency, intensity, etc.

General Conversation

- From my research with accessible VR for people who are blind / visually impaired, audio does provide a good information source about the environment, but I have found that it does not do a good job with information that requires a fast or instantaneous response. So thoughts on an earcon library of sounds that developers can use to create a standardized dictionary of sounds that can mean the same thing from game to game when an instantaneous response is required?
- Beacons
 - BB: Swamp uses beacons to track objects, NPCs
 - Auditory icon representative of character, vs. earcon like a beep or musical cue
 - **Icons e.g. footsteps are more successful than arbitrary beeps**
 - TS: Menu systems are often anchored to field-of-view or pinned to real world; users needed both
 - Could “handcast” by pointing palm at a menu; it would have to grow at a distance and make a sound to indicate this
 - Made a “pitch up” sound to indicate growth, “pitch down” for shrinking
 - Menu stuck in landscape + user moving away = need for menu to grow; decided to have it play the sound regardless
 - RR: in Soundscape, the beacon has multiple layers; **changes texture based on your orientation**
 - If it's behind you, it's a thump; as you rotate towards it, you get a clearer ping
 - Frequency and rhythm are two things you can really tie into
 - In gaming world, Gears of War has beacon system designed to fit in game world and specific game sounds
 - TS: how to make those seamless with the game world?
 - RR: high-frequency sounds localize better, low frequency sounds spread out
 - **Perceive a sound in pinpoint location -> use high frequency sound**

- We have a very good memory for patterns and rhythms
 - BB: posted paper about audio icons vs. earcons
 - TS: **Easier to distinguish rhythm than tone**
 - RR: Challenge in games
- How would you recommend avoiding a noisy scenario? Different tones, frequencies, multiple beacons, etc? Should there be baseline system-supported sounds?
 - BB: There should be defaults but **players should be able to adjust volume on each kind of sound individually**
 - RR: Have seen some directors that say “this is my mix,” others that want total user control
 - TS: With music, notes and frequencies tend to be more spread out; can be used to make a more tonal sound design
 - **If 10 sounds all play at once, you need some kind of ducking algorithm to avoid blasting user’s ears**
 - Ideally all sounds still come across and are heard; they need different timbres
 - E.g. **different sounds for dialogue, button click, etc.**
 - **Recorded app and tried to make the most chaotic mix of sounds**
 - **What happens in situations with clutter?**
 - BB: Simultaneous audio recognition maxes out between 6 and 20
 - They spatialized sound, which increases recognizability
 - TS: with haptics, you can have stereo - possible to roll controllers and have the sensation of a ball moving across your hands
 - RR: in the future looking forward to “spatial haptics” - enough actuators to give sense of positionality
 - RR: For simultaneity, we can take learnings from games; everything has an audio engine, priority systems
 - Consider engine sounds in racing games; to have a sound cut through it, you need to cut a space in the frequencies for it
 - Takes some work and CPU cycles but it’s worth it
 - Results in feeling of being able to hear everything at once
- What are the most important sounds for the first mainstream VR platform to become accessible?
 - SS: when thinking about what audio cues/earcons we need to make VR accessible, made me think about in-game/in-app audio experiences
 - Value in consistent sounds across apps/platforms, esp in context of screen reader
 - BB: When one platform does it, others will imitate it - that’s what happened in audio games
 - Collision information is vital - need to know when you’re run into things
 - Screen reader scan telling you what’s around you is vital
 - SS: Challenge in discoverability; you need to learn a whole audio language to be proficient

- Is game creator controlling full soundscape or is it coming from platform/screen reader?
 - Games have rich audio; want assistive things to slot right into that
 - RR: Want sounds to fit into my game's universe, esp on sound frequency and tonal qualities
 - Could have some kind of consistent patterns modified per game
 - BB: Not enough data to know what that kind of constant between platforms would be
 - SS: In soundscape, you control the full audio experience; also true of virtual world experience
 - May need to mesh assistive and native sounds
 - AS (Andy Slater): have tried to make our own backdoor way in Unity to create a soundscape-like experience, but is there a way to implement it directly?
 - Would be great to have SeeingAI or Soundscape directly
 - SS: Can discuss
 - BD (Bob Dodd): Seeing VR wraps around Unity Apps (or can be built in). Maybe Soundscape would work the same?
- Passive vs active sound
 - JA (Jesse Anderson): Lots of audio games that have a "ping" feature to learn what's nearby via earcons
 - Last of Us pt I and II let you hold down right bumper then press square for enemies or circle for objects
 - Would love to see some kind of consistency. Ok for games to have their own sounds, but something cross-platform would be great
 - Some games have narration, text to speech, etc. that work on PC but not Steam Deck or consoles
 - On an engine level (e.g. Unity, Unreal) it would be helpful
 - BB: Swamp has a sonar, but it's passive when you activate it. A Hero's call has a passive ping for objects, doors, and walls.
 - My hands are busy navigating and interacting with the environment and I don't really want to use my hands to ping around me. Maybe to beep where someone I'm tracking is.
 - You can put a little beacon to follow, or put a looping sound on the walls like [cyclepath](#) for a racing game:
 -
 -
 - AG (Aaron Gluck): Have been working on XR with no visuals at all
 - Examined sonar as a technique
 - Worked well in a slow-paced environment where you can take your time
 - Needs to cycle; can't send out one 360 burst or you get too much audio noise, but a rotating ping (even fast) is more effective
 - Anything that needed instantaneous response, it's too slow; same for words
 - Had to switch for haptic feedback for instant reaction

- RR: Publisher for Killer Instinct worked with Sightless Combat, a blind fighting gamer
 - Design of sounds was distinct for every character, every move
 - Need to target sets of tools for engines, middlewares to make this easier
 - Models for doing these things
 - A good screen reader for every engine
 - Audio middleware like Firelight FMOD Studio
- BB: often blind folks like to use their own screen reader if possible
 - RR: for Minecraft, had to cover 20+ platforms to run on
- BD: Maybe APIs is where we need to be for screen-reader UI? I'm thinking of the Second Life UI
 - @Bob I really liked the Second Life UI for setting labels and description as a starting point
- AS: middleware isn't accessible to blind creators right now - stopping blind users from participating
 - Still have to work with sighted people to get that done
 - RR: Will try to connect with FMOD; tooling in general has poor accessibility, bare minimum compliance
- BB: Only platform that's kind of compliant is ____, because it's open source
 - Any way for Unity to feel comfortable letting blind hackers come in
- TS: Making a user interface relying on head pose for Cognixion
 - Have sound cues as well as haptics
 - What users say is both announced and displayed
 - Larry suggested "hapmoji"/"audiomoji" - recognizable earcons along with speech