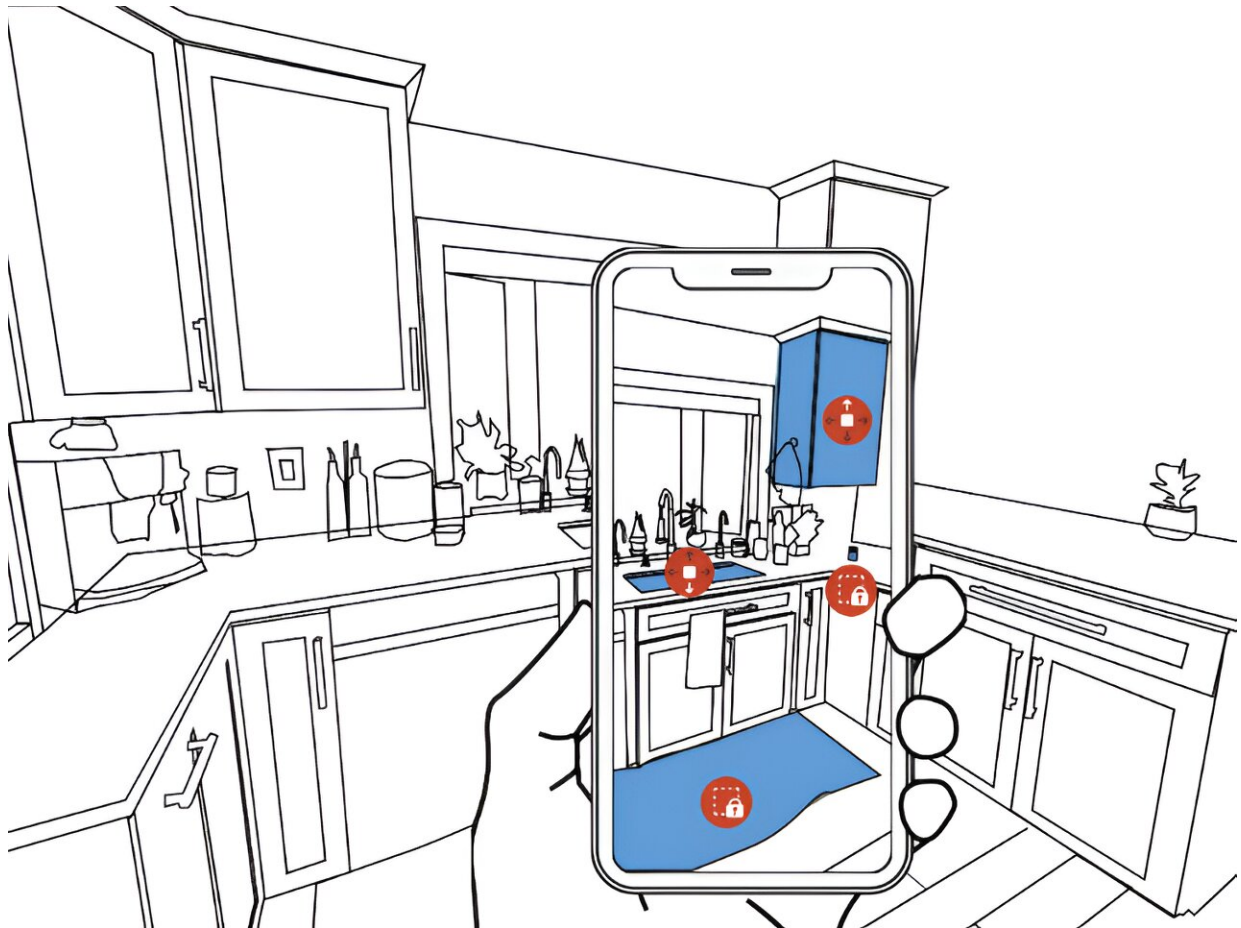# Q&A: Researchers aim to improve accessibility with augmented reality

October 17 2023, by Stefan Milne

RASSAR is an app that scans a home, highlights accessibility and safety issues, and lets users click on them to find out more. Credit: Su et al./ASSETS '23

Big Tech's race into augmented reality (AR) grows more competitive by the day. This month, Meta released the latest iteration of its headset, the Quest 3. Early next year, Apple plans to drop its first headset, the Vision Pro. The announcements for each platform emphasize games and entertainment that merge the virtual and physical worlds: a digital board game imposed on a coffee table, a movie screen projected above airplane seats.

Some researchers, though, are more curious about other uses for AR. The University of Washington's Makeability Lab is applying these budding technologies to assist people with disabilities. This month, researchers from the lab will introduce multiple projects that deploy AR—through headsets and phone apps—to make the world more accessible.

Researchers from the lab will first present RASSAR, an app that can scan homes to highlight accessibility and safety issues, on Oct. 23 at the ASSETS '23 conference in New York.

Shortly after, on Oct. 30, other teams in the lab will present early research at the UIST '23 conference in San Francisco. One app lets the headsets better understand natural language and the other aims to make tennis and other ball sports accessible for low-vision users.

UW News spoke with the three studies' lead authors, Xia Su and Jae (Jaewook) Lee, both UW doctoral students in the Paul G. Allen School of Computer Science & Engineering, about their work and the future of AR for accessibility.

What is AR and how is it typically used right now?

Jae Lee: I think one commonly accepted answer is that you use a wearable headset or a phone to superimpose virtual objects in a physical

environment. A lot of people probably know AR from "Pokémon Go," where you're superimposing these Pokémon into the physical world. Now Apple and Meta are introducing "mixed reality" or passthrough AR, which further blends the physical and virtual worlds through cameras.

Xia Su: Something I have also been observing lately is people are trying to expand the definition beyond goggles and phone screens. There could be AR audio, which is manipulating your hearing, or devices trying to manipulate your smell or touch.

A lot of people associate AR with virtual reality, and it gets wrapped up in discussion of the metaverse and gaming. How is it being applied for accessibility?

JL: AR as a concept has been around for several decades. But in Jon Froehlich's lab, we're combining AR with accessibility research. A headset or a phone can be capable of knowing how many people are in front of us, for example. For people who are blind or low vision, that information could be critical to how they perceive the world.

XS: There are really two different routes for AR accessibility research. The more prevalent one is trying to make AR devices more accessible to people. The other, less common approach is asking: How can we use AR or VR as tools to improve the accessibility of the real world? That's what we're focused on.

JL: As AR glasses become less bulky and cheaper, and as AI and computer vision advance, this research will become increasingly important. But widespread AR, even for accessibility, brings up a lot of questions. How do you deal with bystander privacy? We, as a society, understand that vision technology can be beneficial to blind and low-

vision people. But we also might not want to include facial recognition technology in apps for privacy reasons, even if that helps someone recognize their friends.

Let's talk about the papers you have coming out. First, can you explain your app RASSAR?

XS: It's an app that people can use to scan their indoor spaces and help them detect possible accessibility safety issues in homes. It's possible because some iPhones now have lidar (light detection and ranging) scanners that tell the depth of a space, so we can reconstruct the space in 3D. We combined this with computer vision models to highlight ways to improve safety and accessibility. To use it, someone—perhaps a parent who's childproofing a home, or a caregiver—scans a room with their smartphone and RASSAR spots accessibility problems. For example, if a desk is too high, a red button will pop up on the desk. If the user clicks the button, there will be more information about why that desk's height is an accessibility issue and possible fixes.

JL: Ten years ago, you would have needed to go through 60 pages of PDFs to fully check a house for accessibility. We boiled that information down into an app.

And this is something that anyone will be able to download to their phones and use?

XS: That's the eventual goal. We already have a demo. This version relies on lidar, which is only on certain iPhone models right now. But if you have such a device, it's very straightforward.

JL: This is an example of these advancements in hardware and software that let us create apps quickly. Apple announced RoomPlan, which

creates a 3D floor plan of a room, when they added the lidar sensor. We're using that in RASSAR to understand the general layout. Being able to build on that lets us come up with a prototype very quickly.

So RASSAR is nearly deployable now. The other areas of research you're presenting are earlier in their development. Can you tell me about GazePointAR?

JL: It's an app deployed on an AR headset to enable people to speak more naturally with voice assistants like Siri or Alexa. There are all these pronouns we use when we speak that are difficult for computers to understand without visual context. I can ask "Where'd you buy it from?" But what is "it"? A voice assistant has no idea what I'm talking about. With GazePointAR, the goggles are looking at the environment around the user and the app is tracking the user's gaze and hand movements. The model then tries to make sense of all these inputs—the word, the hand movements, the user's gaze. Then, using a large language model, GPT, it attempts to answer the question.

How does it sense what the motions are?

JL: We're using a headset called HoloLens 2 developed by Microsoft. It has a gaze tracker that's watching your eyes and trying to guess what you're looking at. It has hand tracking capability as well. In a paper that we submitted building on this, we noticed that we have a lot of problems with this. For example, people don't just use one pronoun at a time—we use multiple. We'll say, "What's more expensive, *this* or *this*?" To answer that, we need information over time. But, again, you can run into privacy issues if you want to track someone's gaze or someone's visual field of view over time: What information are you storing and where is it being stored? As technology improves, we certainly need to watch out for these privacy concerns, especially in computer vision.

This is difficult even for humans, right? I can ask, "Can you explain that?" while pointing at several equations on a whiteboard and you won't know which I'm referring to. What applications do you see for this?

JL: Being able to use natural language would be major. But if you expand this to accessibility, there's the potential for a blind or low-vision person to use this to describe what's around them. The question "Is anything dangerous in front of me?" is also ambiguous for a voice assistant. But with GazePointAR, ideally, the system could say, "There are possibly dangerous objects, such as knives and scissors." Or low-vision people might make out a shape, point at it, then ask the system what "it" is more specifically.

And finally you're working on a system called ARTennis. What is it and what prompted this research?

JL: This is going even more into the future than GazePointAR. ARTennis is a prototype that uses an AR headset to make tennis balls more salient for low vision players. The ball in play is marked by a red dot and has a crosshair of green arrows around it. Professor Jon Froehlich has a family member that wants to play sports with his children but doesn't have the residual vision necessary to do so. We thought if it works for tennis, it's going to work for a lot of other sports, since tennis has a small ball that shrinks as it gets further away. If we can track a tennis ball in real time, we can do the same with a bigger, slower basketball.

One of the co-authors on the paper is low vision himself, and he plays a lot of squash, and he wanted to try this application and give us feedback. We did a lot of brainstorming sessions with him, and he tested the

system. The red dot and green crosshairs is the design that he came up with, to improve the sense of depth perception.

What's keeping this from being something people can use right away?

JL: Well, like GazePointAR, it's relying on a HoloLens 2 headset that's $3,500. So that's a different accessibility issue. It's also running at roughly 25 frames per second and for humans to perceive in real time it needs to be about 30 frames per second. Sometimes we can't capture the speed of the tennis ball. We're going to expand the paper and include basketball to see if there are different designs people prefer for different sports. The technology will certainly get faster. So our question is: What will the best design be for the people using it?

More information: Paper: [A Demonstration of RASSAR: Room Accessibility and Safety Scanning in Augmented Reality](#)

Paper: [Towards Designing a Context-Aware Multimodal Voice Assistant for Pronoun Disambiguation: A Demonstration of GazePointAR](#)

Paper: [Towards Real-time Computer Vision and Augmented Reality to Support Low Vision Sports: A Demonstration of ARTennis](#)

Provided by University of Washington