The Future of Audio in XR

Insights from Exploring Future Reality hosted by

NYC MEDIA LAB

Why sometimes, hearing is believing

It's often said that seeing is believing. Yet if VR, AR, and MR (XR collectively) experiences are to fully immerse viewers, they must cater to all the senses, including hearing.

Some argue that humans' sense of hearing is harder to fool than sight and thus more important in XR. It's more deeply tied to sensing danger – when a twig snaps behind us, we first hear, then look – and when sight and sound come into conflict, the latter often wins.

"If someone taps on an object that looks like metal but sounds like wood, they'll most likely believe it's painted wood," explained Agnieszka Roginska, who studies immersive auditory environments as Music Associate Professor at NYU Steinhardt.

> "We're very attuned to our dimensional sound environment."
>
> **David Al-Ibrahim, Art Director, Warner Music Group**

In XR, audio technology has lagged behind visual technology partly because fooling the auditory sense is so difficult. Our ears are exquisitely sensitive instruments and can detect minute variations in sound waves. When sound is muffled by nearby objects or bounces off walls, we notice. "We're very attuned to our dimensional sound environment," said David Al-Ibrahim, art director at Warner Music Group. Humans use dimensional auditory input to build a mental map of the world around them – where objects are, what they are, and where they are moving.

Insights from NYC Media Lab's **Exploring Future Reality** conference, November 30th, 2017

Research based in part on panels with:

- **Agnieszka Roginska,** *Music Associate Professor, NYU Steinhardt*
- **David Al-Ibrahim,** *Art Director, Warner Music Group*
- **Matt Hartman,** *Partner, Betaworks Ventures*
- **Terence Caulkins,** *Associate in Acoustics, Audio Visual, and Interactive, ARUP*
- **Steven Feiner,** *Professor of Computer Science, Columbia University*
- **Ken Perlin,** *Professor of Computer Science, NYU Future Reality Lab*
- **Youg Liu,** *Professor of Electrical Engineering, NYU Tandon School of Engineering*
- **Adaora Udoji,** *Investor & Producer*

Today, most audio systems used in XR can't convey enough dimensional information to be believable because they were developed for 2D environments like the cinema. XR creators face a number of challenges in adapting existing audio technology to immersive 3D worlds.

Glossary of XR audio terms:

- **2D Audio -** Audio that conveys sound in two dimensions.
- **3D Audio -** Audio that conveys sound in three dimensions.
- **6DOF -** Six degrees of freedom, in reference to an XR viewer's freedom to look around.
- **Acoustics -** The science of sound wave behavior.
- **Ambisonics -** A technique for creating full sphere surround-sound.
- **Binaural audio -** A technique for capturing sound the way that humans hear it, usually with two speakers mounted on a model head.
- **Interactive 3D audio -** 3D audio that can cater to a user's movements throughout a virtual environment.
- **Stereo -** Audio that comes from two or more speakers arranged around around a listener.
- **Timbral capacity -** The range of tones a microphone can capture or produce.

The chief XR audio challenges

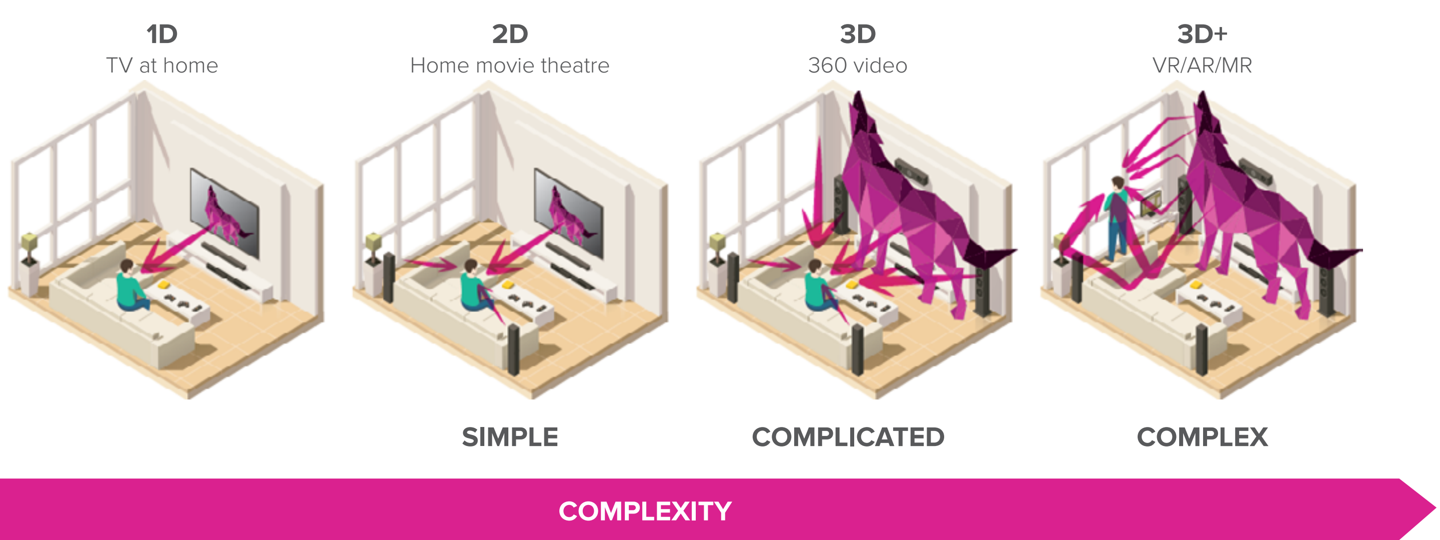
To understand why most 2D audio is so poorly suited to XR, it's helpful to think in terms of a home theatre. Imagine you are seated on a couch, facing forward.

This is a 2D sound environment. Speakers are arranged in four directions around you: front, back, left, and right. That accounts for two dimensions. During the movie, each speaker plays different sounds, giving you a rudimentary sense of immersion.

2D audio immersion only works because both you and the screen are fixed. When you are in a 3D environment, such as a 360-degree video, it's like being in a movie theatre with screens in every direction, including on the floor and ceiling. Your head, no longer fixed forward, has what's known as six degrees of freedom, or 6DOF. A 2D audio system can't generate sounds above or below you, so the noises you hear won't match what you see. It's the equivalent of wearing a pair of VR goggles which, no matter which way you turn your head, always show the same fixed frame.

Now imagine a full VR experience where you can get up and walk around the theatre. That 2D audio system must now generate sounds in relation to which way your head turns and where your body moves, neither of which it was designed to do. The more freedom the viewer has, the more complex the audio challenge becomes.

Spactial Audio Environments



To collect dimensional audio information, XR creators use 3D microphones. Much like a 360-degree camera, a 3D microphone captures audio in all directions from a fixed point. That is sufficient to make 360-degree videos believable, but to cater to viewers that have freedom of movement within VR, creators must affix each sound to its virtual source. That means that the sound of a dog barking cannot simply emanate from the viewer's left; it must come from the dog itself.

These sounds can be incredibly difficult to collect. Rather than record single sounds, technicians must gather them at variable distances from recorded objects and then digitize an environment where those sounds can be modulated in relation to the viewer.

> "Every location has its own unique sound signature."
> **Terence Caulkins,**
> **Associate in Acoustics, Audio Visual, and Interactive, ARUP**

The challenges do not end there. Sound waves interact with their environment as they travel. A lone person clapping, for example, doesn't sound the same in a room with a wood floor as it does in a carpeted room because every surface reflects sound differently. "This means every location has its own unique sound signature," explained Terence Caulkins, associate in acoustics, audio visual, and interactive at ARUP, a firm that performs acoustic modeling.

For sound to be fully believable in a VR world, every virtual object must, in some way, impact the audio. That includes the users themselves. When sound waves reach a listener, they bounce off the individual's body and are distorted. Human ears are attuned to these slight variations, and the time lag between the when the sound reaches each ear allows us to determine the direction of the noise – a process known as sound localization. This means that a lone person clapping sounds different to each listener based on their location in relation to the clapper, and the sound changes depending on what else is in the room, where the listener's head is turned, and what the listener is wearing.

To partially compensate for these variations, some XR audio specialists use binaural audio equipment. These recording devices fit a model head with microphones inside each ear to capture sound as a human would actually hear it. Most 360-degree videos use a technique known as ambisonics to play binaural audio as full-sphere surround sound.

Photo: 3DIO

At ARUP, Caulkins and his team capture the unique sound signatures of buildings with a technique known as oralization which uses 3D microphones to create a virtual acoustic blueprint. This model can be used to predict building acoustics and to help XR creators model audio in virtual environments.

When creators get the sounds just right, people notice. "A reverberation time of 1.8 seconds doesn't mean much to most people," explained Caulkins. "But put them in a space with 1.2 and then with 1.8 and they have totally different experiences."

> “People want a world in which they can look around, but also listen around.”
> **Steven Feiner, Professor of Computer Science, Columbia University**

But 3D microphones have their limitations. Many models have a narrow timbral capacity, and when recording live shows, audio engineers often supplement 3D equipment with much higher quality stereo microphones. When playing audio back in a 3D environment, it can also help to have a variety of speakers. A bee buzzing around a listener’s head is easier to mimic with headphones whereas the physical vibrations of a rock concert are easier with a subwoofer. To achieve the desired effect, XR audio engineers often experiment with a hybrid of many technologies.

AR and MR audio challenges are complicated when such gear is used in non-controlled environments. Believable audio would have to interact with objects in both the virtual and observed world, and would have to account for sounds emanating from the outside world. Steven Feiner, Professor of Computer Science at Columbia University, believes the demand for this already exists.

Headphone maker Bose **recently released** a prototype of AR glasses that use sound instead of sight. The device connects to a smartphone and uses a nine-axis sensor to allow users to look at objects and hear recordings. Bose’s website lists Yelp and TripAdvisor among the product’s potential partners. Steven Feiner, professor of computer science at Columbia University, believes the demand for this type of virtual audio experience will only grow.

“People want a world in which they can look around, but also listen around,” Feiner said. “They want to see and hear virtual things that are relevant to whatever they’re looking at, without having to reach into their pocket.”

No matter how far XR audio advances, Al-Ibrahim predicts that end-user hardware will remain a bottleneck. “You can design a really beautiful interactive acoustic environment, but when someone listens on fifteen dollar earbuds, all that complexity is lost,” he said.

What the future holds

When XR creators deliver exceptional audio, they greatly enhance the viewer's sense of immersion. Audio is deeply tied to our emotions and, just as musical scores and soundtracks bring movies to life, believable audio makes virtual realities feel more real.

In the online 360-degree video experience **Calling Thunder**, Al-Ibrahim and his team partnered with the Cornell Lab of Ornithology to create an experience where visuals take a backseat to audio. The film uses 3D audio recordings of present-day New York City traffic which fades into bird calls to recreate the sensation of traveling back in time on the island of Manhattan.

Immersive audio can be used to guide XR viewers through 3D environments where they have freedom of choice. Subtle sounds can help viewers navigate by nudging them toward points of interest and warn of danger. When users can use their auditory senses in XR the way they would in real life they are more likely to believe the experience. Soon, “no one will be thinking about immersive audio as technology anymore than we think about the web as technology,” said Ken Perlin, professor of computer science at NYU.

For some, truly immersive 3D audio offers the potential for social interaction. Musicians who are geographically isolated could meet and jam in a high-fidelity virtual world. Today, “the latency requirements are actually extremely straining,” said Yong Liu, professor of electrical engineering at NYU Tandon School of Engineering. “But we'll quickly see advances in that.” And real-life audio experiences such as concerts could be replicated virtually.

Matthew Hartman, partner at New York City-based incubator Betaworks, believes the idea of recreating virtual concerts has real monetary implications both for venues and performers. Artists could distribute experiences in much the same way the internet now spreads their songs. These virtual events could feature audio personalization by giving every individual front-row seats – and perhaps even allow them to interact with the performers. “VR, AR, any of these realities, are an alternative medium by which we can view media,” said Adaora Udoji, an XR investor and producer.

If an opera experience were fully replicated in VR, however, some panelists questioned how many people would still visit the physical venue. Might more refined XR audio become a source of disconnect as fewer individuals choose to leave their homes?

Hartman isn't worried. “We are pack animals,” he said. “We want to see and feel each other. More likely, we'll see VR techniques pop up in fixed venues to transport everyone in a space together.”