Welcome everyone! Great to see so many people here on a thursday afternoon, hopefully you're not too worn out.

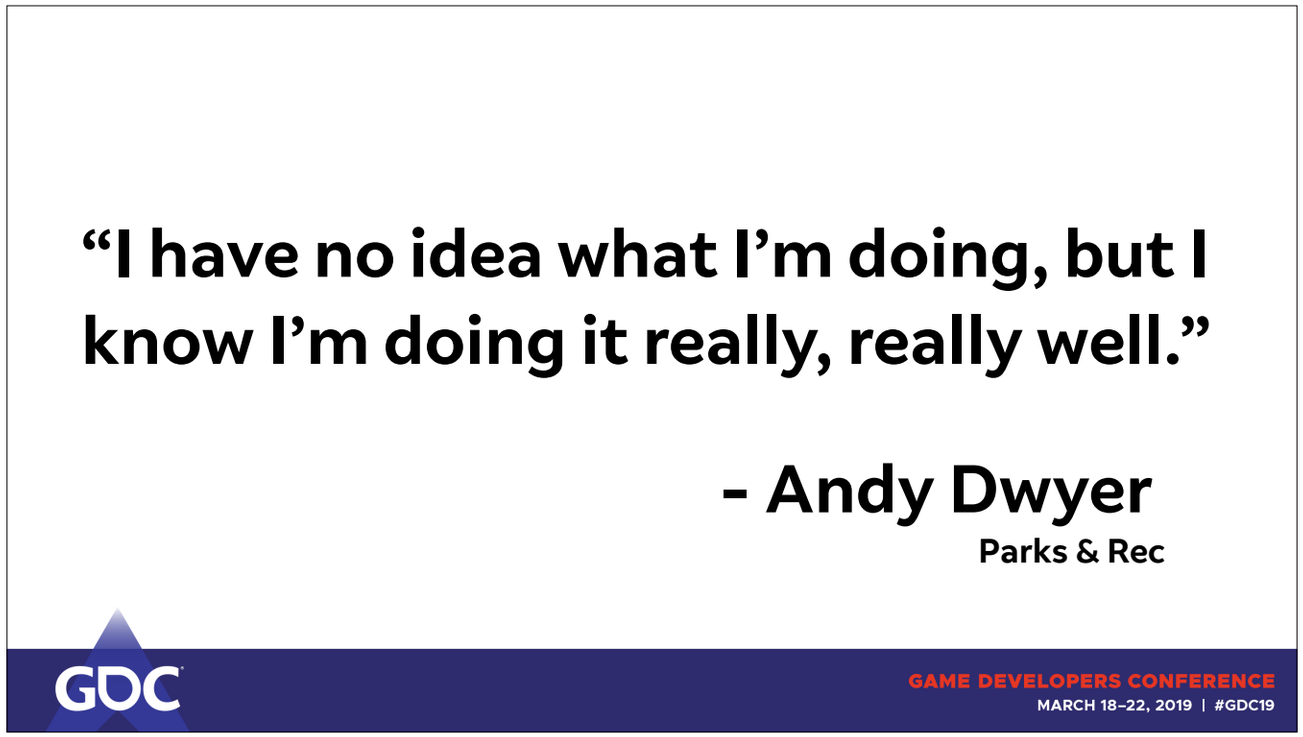
It's my first GDC since 2015 so I'm really excited to talk about what we've been working on at Magic Leap, and how it might be of help to those of you working in these new VR/AR/MR formats

quick show of hands - how many of you have released something for one of these formats?

currently working on one?

just interested?

a little about myself, I'm dave shumway, lead audio designer and composer at magic leap studios, the first party shop inside Magic Leap. I joined in late 2014, and pretty much felt like this on day 1



Earlier that year, I’d played some of my work at the GANG demo derby here at GDC. Through that demo, I met with George Sanger, who unbeknownst to me, was about to become Audio Director for Magic Leap.

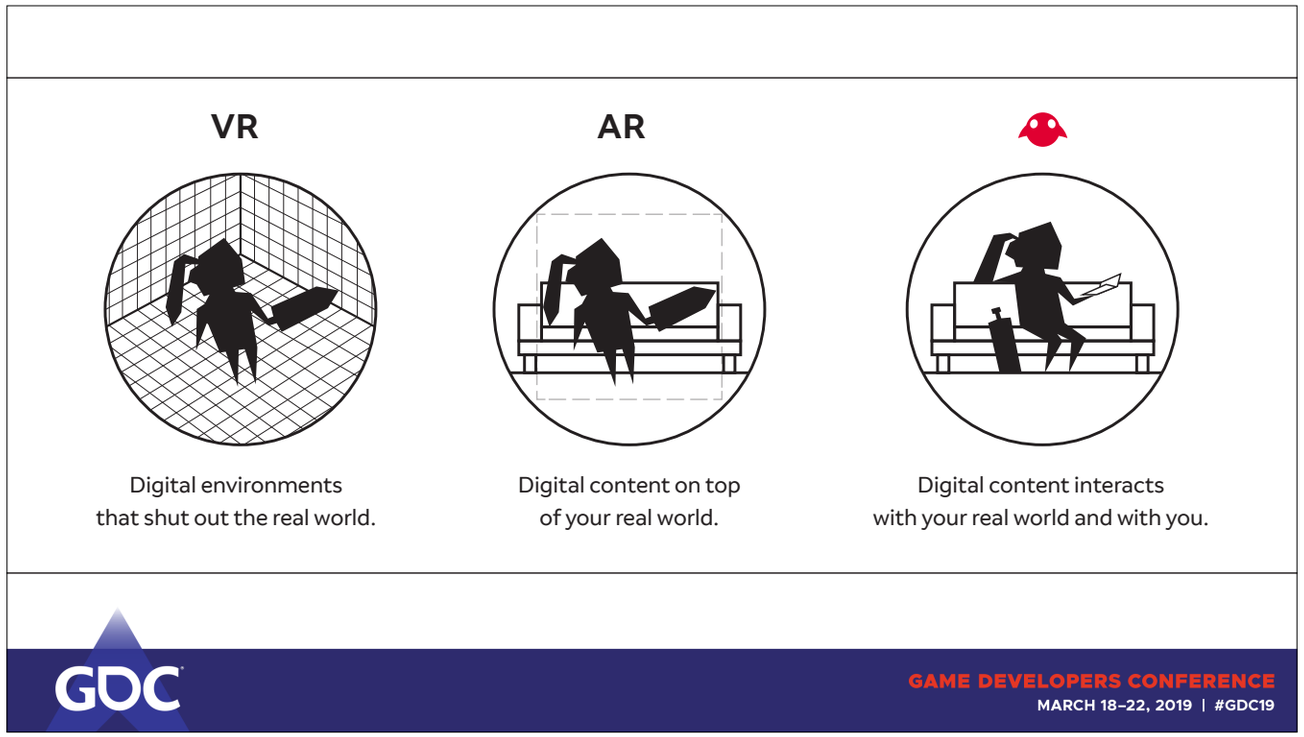
We had breakfast at Denny’s, like ya do, and over the next few weeks/months, he began to drop cryptic hints about this new thing he was getting involved in. By December, He’d brought me on and we started to build out an amazing audio team. So if you’re here George, thank you!

But from then until now, most days I feel like Andy Dwyer. Best of intentions, no idea what I’m doing.

For me, this is why I came to Magic Leap - stand at the edge, peer into the unknown, and hopefully figure it out before I hit the ground.

But I mentioned the team - luckily, we have an awesome team at magic leap of programmers, tech sound designers, and more traditional sfx/music folks that work together to push the boundaries of this new medium.

So to all of them back in sweaty florida, I salute you!

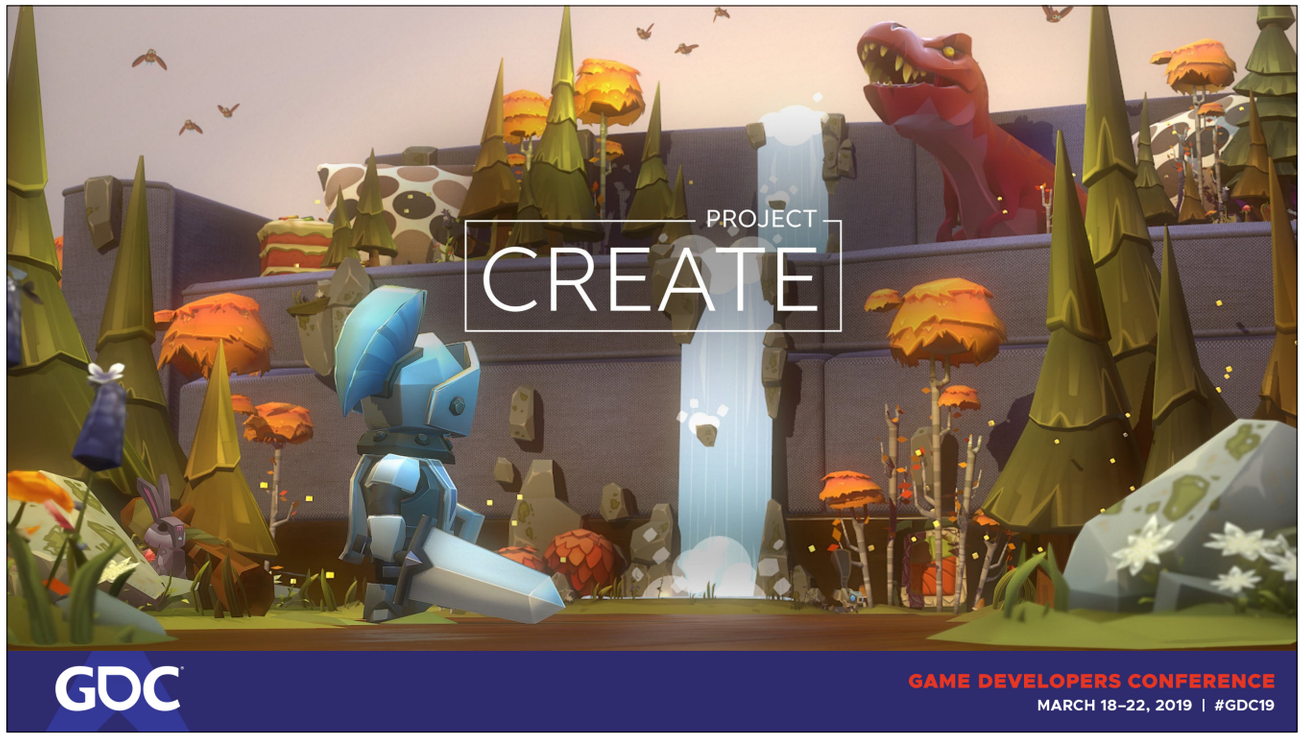


Now to lay some groundwork for what we'll be talking about today, we use this image often at Magic Leap to distinguish the differences between VR/AR/MR

Ss you see on the left, working in VR allows us to completely control the environment the user inhabits. AR, enhances the user's world by overlaying digital content. And MR (or spatial computing, depending on who you ask), allows the content you create to interact in a meaningful way with the geometry and objects in a user's space.

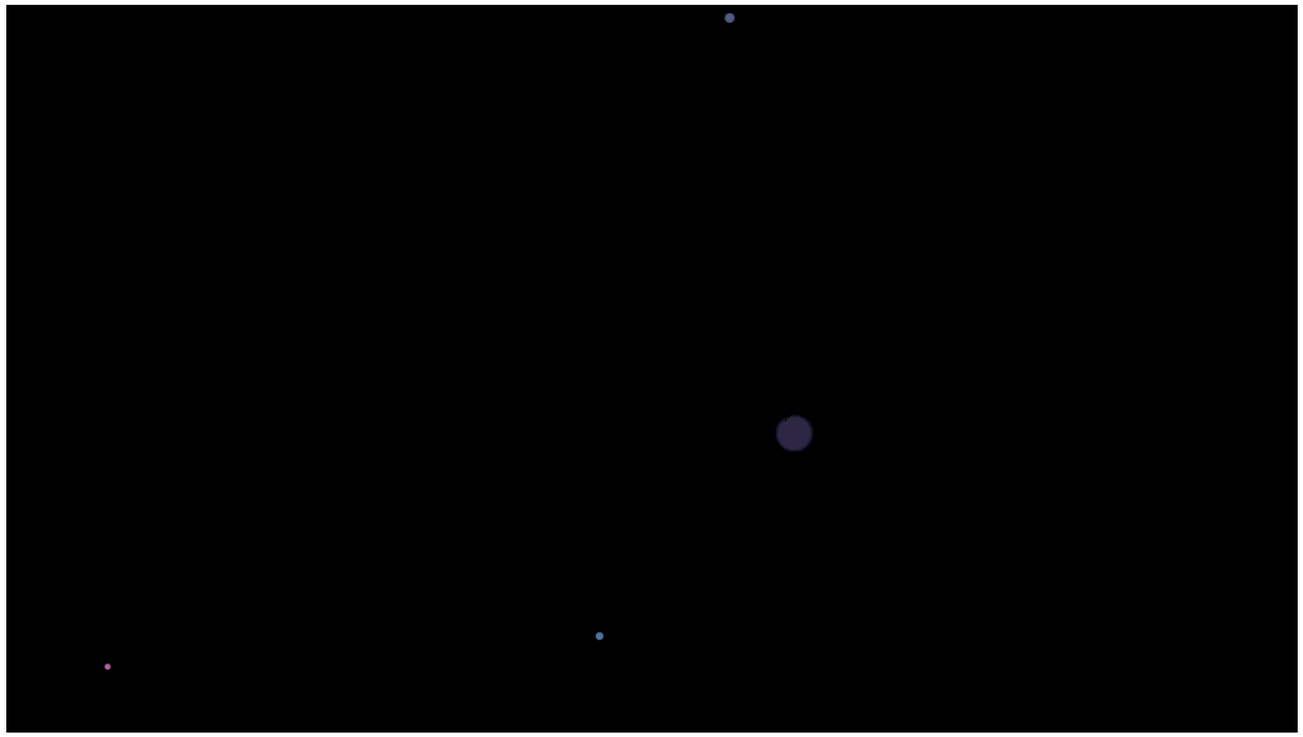
The experiments we've done at Magic Leap span across this spectrum, and again, will hopefully apply to what each of you might be working on.

Today will focus more on the right hand side of this image, as we talk about Project: Create



I functioned as audio lead for Project Create

- a chance for people to experience what interactions can look, feel, and sound like in mixed reality
- paint in 3D space
- build structures and gadgets with physics
- and create environments for AI-driven characters to explore
- All at once. In the real world.
- It was one of the first big things we built as a studio, and it’s been a much different challenge than we expected. But for a little context, here’s the launch trailer



Create has two goals. To introduce ML users to the concept of Spatial Computing, and to allow them to have a lot of fun exploring this new medium.

To that end, a handful of design pillars were established

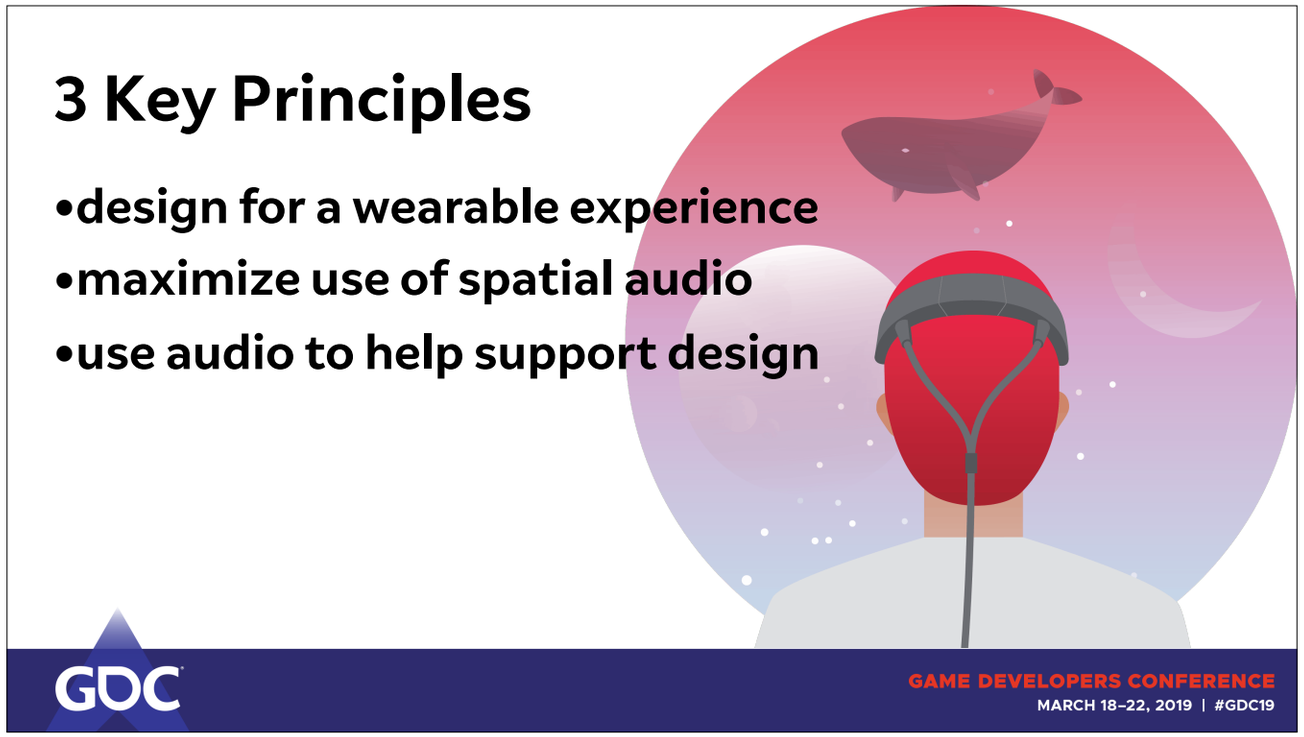
1.- Comfortable - We wanted as many people as possible to be able to operate and interact with our experience.

2.- Intuitive – We wanted people to effortlessly understand how they could interact with these new virtual objects.

3.- Authentic – Virtual objects should fulfill user expectations by reacting to inputs and understanding the physical world.

4.- Safe to Explore - Minimized failure states to encourage experimentation and discovery.

5.- Immersive - Different from VR. we want to make you believe virtual content exists in your space and combines with it, rather than transporting you somewhere else



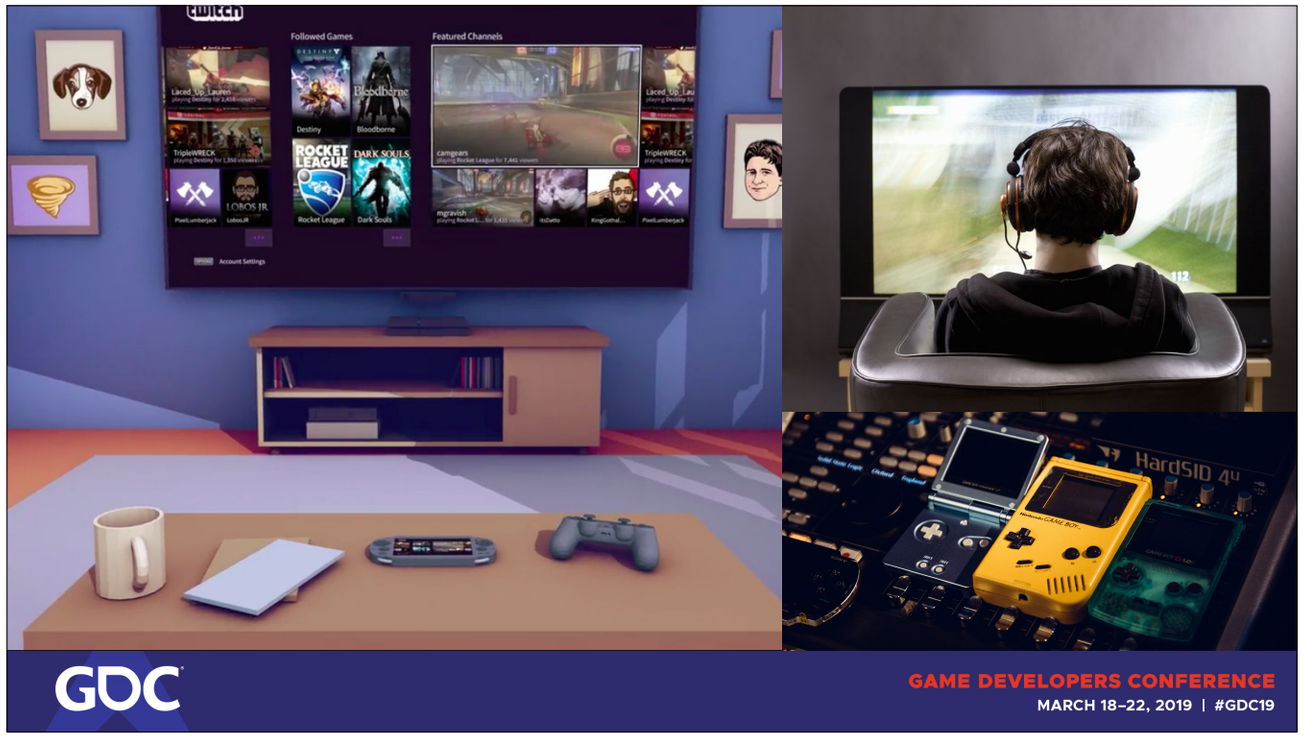
Armed with the learnings our team has been gathering since those early days in 2014, I want to highlight 3 things we did to uphold Create’s design pillars.

A lot of the things we cover today might not be new, in fact I’m sure they’re not. But they are definitely what we’ve felt are most important when transitioning into this new medium, and the context of what this space feels like is important.

My goal for today is to show as much detail as I can in the time we have about the thinking behind our design choices, and how we implemented them

Key Principle #1

- design for a wearable experience



People have been designing and experiencing content on a screen for decades now, and we've got a lot of things figured out. One of things that continues to be tricky, however, is not knowing what a consumer's playback setup will be.

Playing a game on a screen puts physical distance between content and user. Most of my game playing only happens when my kids are in bed, so the volume is either super low, or I'm using headphones.

Hopefully, most users get to experience our content at a reasonable playback volume, maybe even in a great sounding surround environment. We can plan for the best case scenario, and hope that will translate into the best user experience, but without a lot of certainty regarding the end result.

Wearable audio!

As we move away from 2D screens, we find that headsets are increasingly including built-in audio in their hardware designs. In fact, just yesterday oculus announced that the Rift S would follow the same path as the Quest, which is shown in the picture on the right.

This is something I remember George Sanger preaching early and often at Magic Leap - that there be “no silent experiences” when using our hardware. Our old demo station at the time was mostly just optics bolted to a desk that you could look through, but George made sure there were speakers zip tied to either side so that you could look in to see the visuals and automatically be positioned for optimal audio playback.

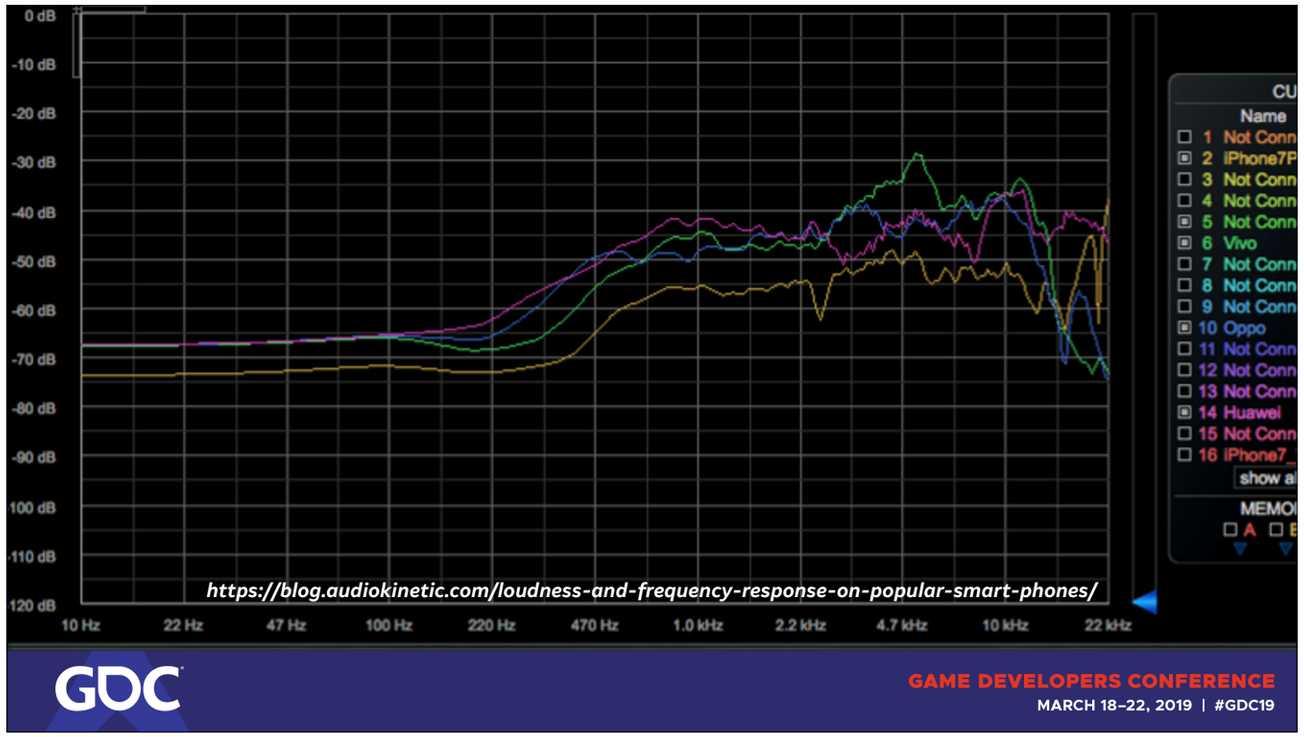
even if no audio had been created yet, there was a great psychological effect of having the two senses inseparably linked

And as we look at the current generation of mixed reality devices (I couldn’t find something illustrative for HL2, but it’s the same idea), audio is physically part of the hardware design. You can’t take them off.

So outside of users selecting their own headphones to bring to the party, we know the exact playback specifications of every platform we target. Which is a HUGE win.

But it also adds an increased level of responsibility to ensure that sound design, music and implementation perfectly complement the experience

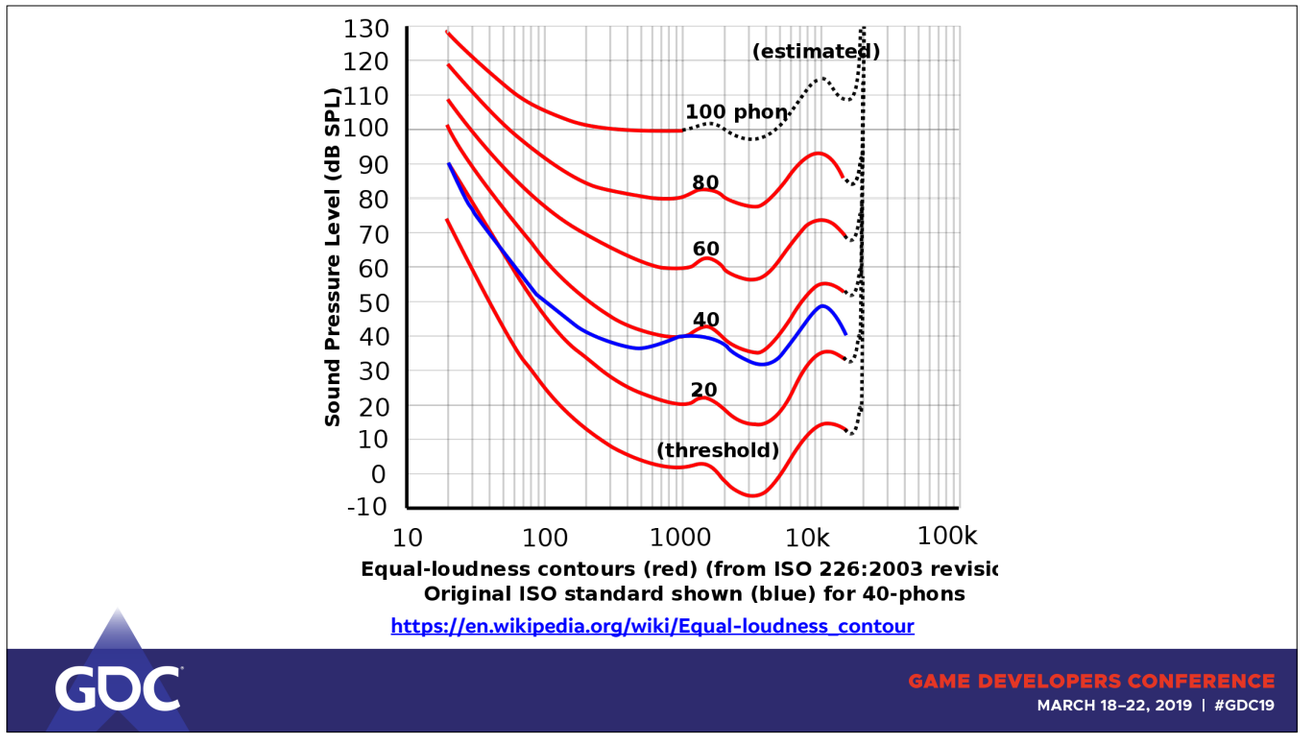
potential for subjecting users to fatiguing sounds or skewing the desired outcome of an experience looms large over our work when those sounds are coming from *HERE* rather than *HERE*



Before we get too carried away, it’s important to establish that from the days of zip tied speakers up to now, we’re firmly in the camp of using a device’s built-in speakers to deliver audio in a meaningful way, and you’ll hear me reference that throughout the talk.

However, that does come with a few caveats, the first being frequency response.

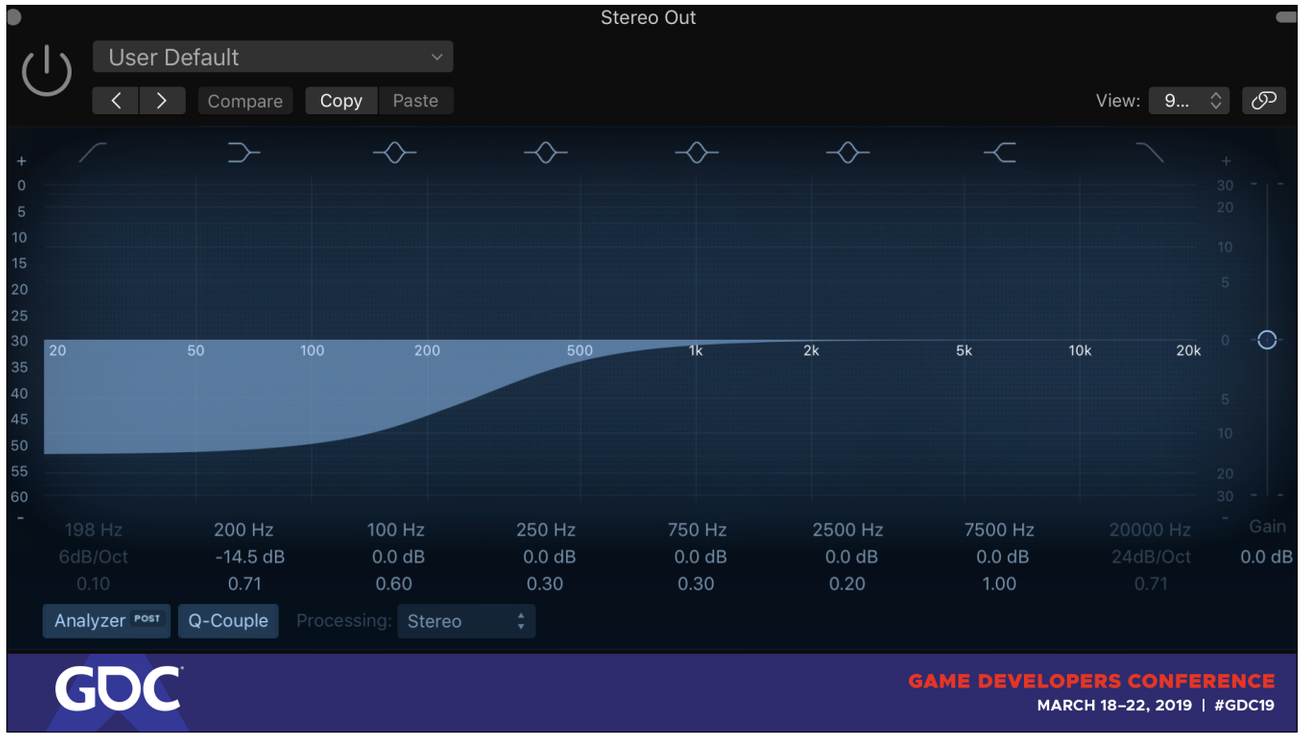
While this image from our friends at Audiokinetic is of flagship phones and not any of the devices just highlighted, you get the idea. And for anyone that has worked on a mobile title, this is not news to you.



Combine that small speaker form factor with close proximity to our ears, and this becomes potentially more complicated because of the way our ears work

The equal loudness contour curve shows us that we perceive certain frequencies louder than others, even when played back at the same volume

In practice, it’s similar to the effect of subtractive EQ, where the lack of one frequency band immediately makes others feel more pronounced



Ok so how does this sound in practice? Here's an object in Create and some examples of how we approached the sound design. It's an explosive block that you can build with, but that triggers a countdown and explosion on collision.

Again, this is not entirely new, we have to deal with this problem every day on mobile devices. But it's something to keep in mind as we build the case for designing wearable audio.

play examples

- v1
- v1 plus EQ
- v2
- v2 plus EQ

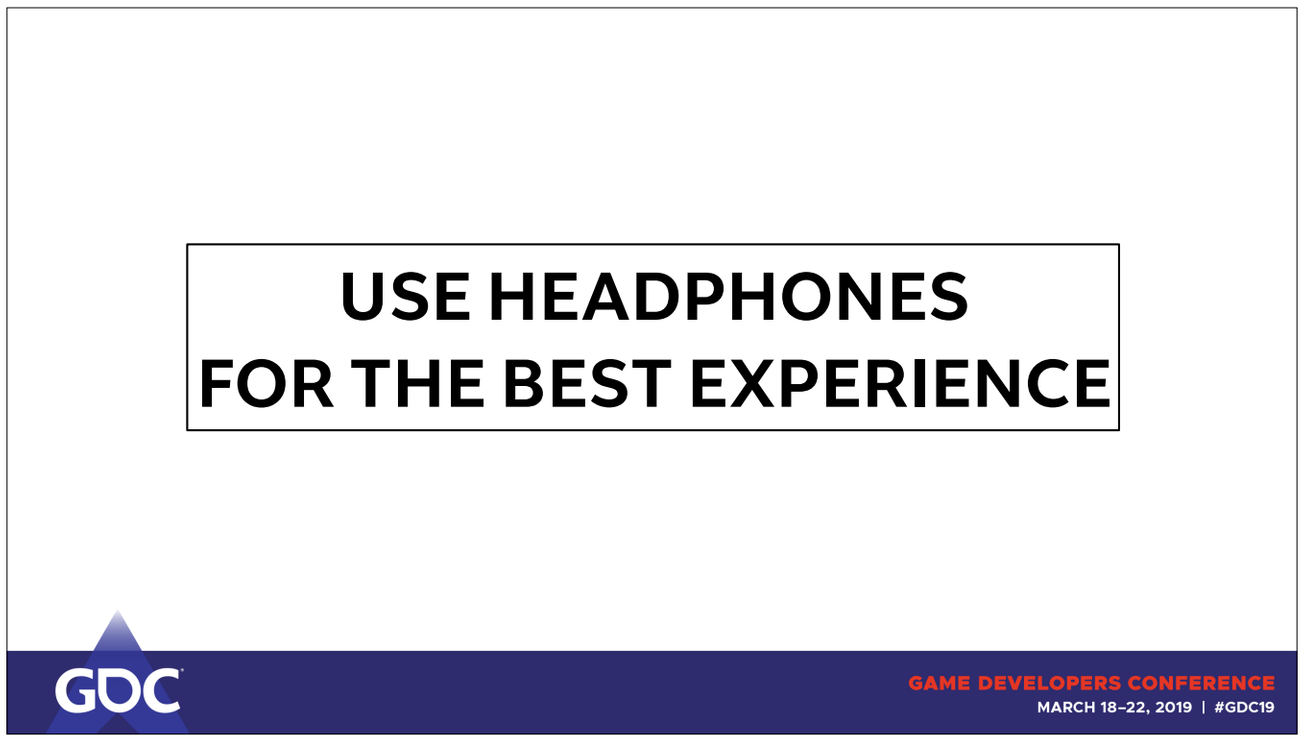
So at this point you might be thinking, cool story - it’s just a bigger mobile device. I want full range speakers! and 12” subs!

During much of our early days developing Create (and other prototypes), we’d often demo in rooms with surround speaker setups so that larger groups could give feedback at once, and because the on-device audio hardware was still in flux. This approach definitely had its strengths, and will still be used widely in LBE

But when we were eventually able to move exclusively to on-device playback, we were surprised at how much more personal and accurate a user’s audio experience became. More on this in a bit…

Even if you’re watching netflix on your phone, an explosion in a movie will need to feel larger than life. It’s how our brains are wired. If it’s happening in your room, however, that’s likely an entirely different scale, and a different expectation.

Stefan Schutze gave an excellent talk on Tuesday about using a real car to learn how audio might work differently at scale for VR experiences. Check it out on the vault for some more valuable insight on this principle.



So if we need the playback to be local to the user as opposed to speakers in a room _and_ we want low frequencies, then we just put one of those cards at the beginning of the app instructing users to wear headphones like mobile apps do.

Maybe, maybe not…

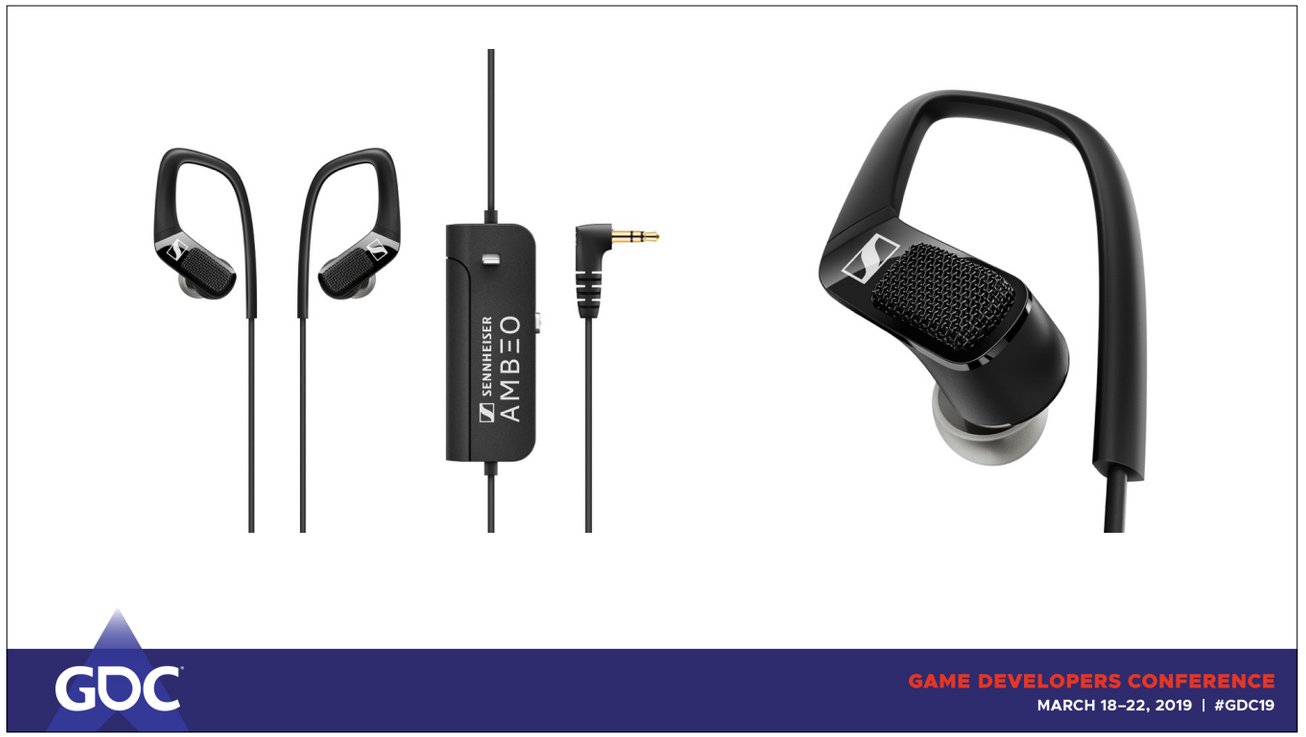
For those that may not have heard of it, Tonandi is our collaboration with Icelandic artists Sigur Ros. It’s a uniquely personal experience that combines the sounds from the band with otherworldly visuals, both of which react to your gestures.

When we first showed Tonandi to the press, the team asked me to choose a headphone to pair with the device for the demo. I gathered a handful of high-end earbuds, ran through the experience with each, and made my selection.

Having completed my task, I went through one last time using the device speakers. Obviously, scientifically, some of the low end that I’d just heard through earbuds was not nearly as present, but I can tell you that my only reaction was THIS IS HOW WE HAVE TO DEMO. Even after having run through the experience half a dozen times, that was the moment when the loop closed and I believed these Icelandic spirits were in my space, reacting to me.

I quickly fired off an email to the team sharing the earbuds I’d selected, but recommended they run the press through without, using the device speakers instead.

one of the most powerful things about mixed reality is how it allows a user’s visuals to remain mostly unoccluded, and despite the less than full range nature of the built-in speakers on all of these wearable devices, there is magic in leaving a user’s ears unoccluded as well. It’s an incredibly compelling piece of data that we did not expect.



That said, there will definitely be scenarios where you'll want to anticipate, and even encourage your users to wear headphones

Sennheiser has recently released a set of earbuds that employs outward facing mics to passthrough the audio of your space and blend it together with the digital content of the experience.

They call it "transparent hearing" and it's an interesting option as you design for wearable audio

Another thing to consider is that you'll be competing with your users ambient noise floor, which again, can be a pro and con. Yes, there's volume to have to work around, but if you can design and mix your sounds to account for that, there's a force multiplier for immersion that can happen.

During development of Create - we'd intentionally mix in different rooms both controlled and open, have a handful of people in during playtests, and stimulate conversations between people in the room and people on device

If the user either couldn't hear us talking to them or started speaking in elevated tones that was my cue that the mix might either be too loud or too busy

as a "worn" device, lots of sound at close proximity to a user's ears can become fatiguing very quickly, calling back to the core design tenet of "Create" to be comfortable and safe for its users

Keep it simple

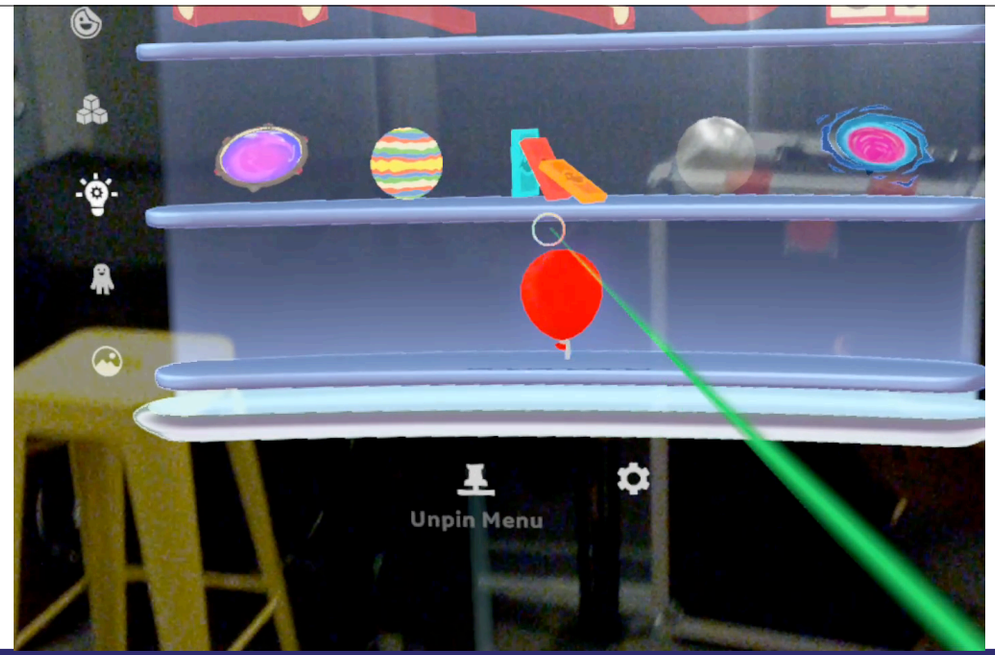
Taking these concepts into account, we found that you may get more mileage out of simpler sounds that can communicate clearly without masking or being masked

using sounds that are less processed might read more successfully especially when you’re asking the user to believe that they are truly existing in their space

Some of the biggest wins in Create came from the simplest sounds, like the drop of a bouncy ball or a domino topple

The visual aesthetic of Create came to be known as “low-poly plus,” and we tried to follow and support that on the audio side in terms of simplicity

I’m going to play some footage of some dominoes, but I’ll mention here that as we’re talking about more sound design focused elements, the clips I’ll play have the music turned off to hopefully be more illustrative in a big echoey room.


Unpin Menu

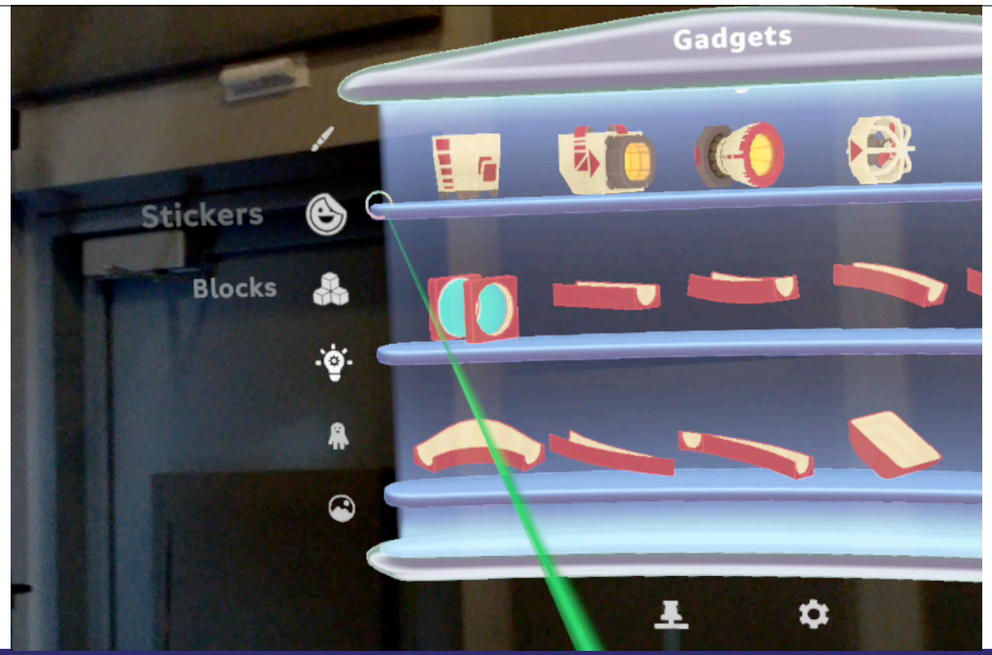
Now that’s not to say that we shouldn’t add detail and sonic interest to the sounds we create, just do it in layers

At a distance, a sound can be relatively nondescript, but have additional layers reveal themselves upon closer examination

Not only does this help maintain a cleaner mix, but meets a user’s expectations and also encourages them to explore the virtual content in their space

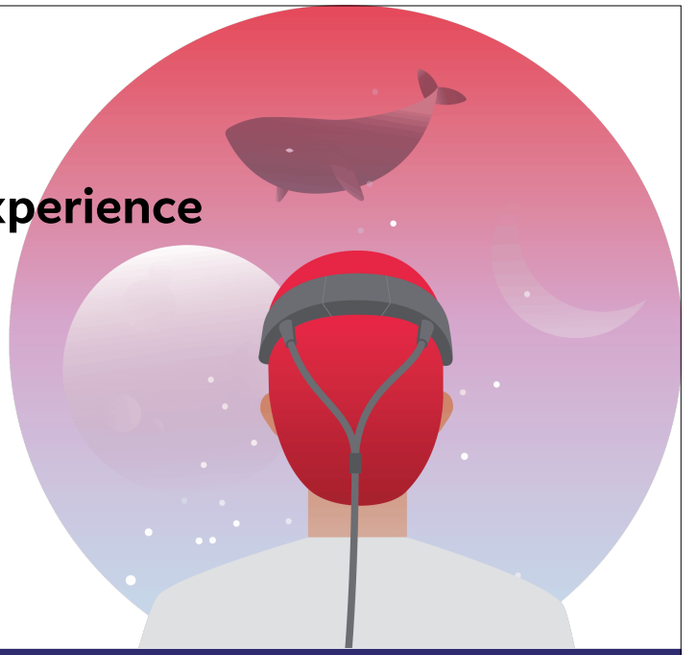
A sound doesn’t need to sound cool or unique or flashy, but it needs to sound right! We found that you won’t necessarily know this intuitively at first while you’re designing in the studio, but you will know immediately if it works when the sound is played back on device.

So getting your sounds in early will allow you to dictate what sounds are enough to carry the weight at a distance, and what layers you should bake into that cake.


Gadgets
Stickers
Blocks

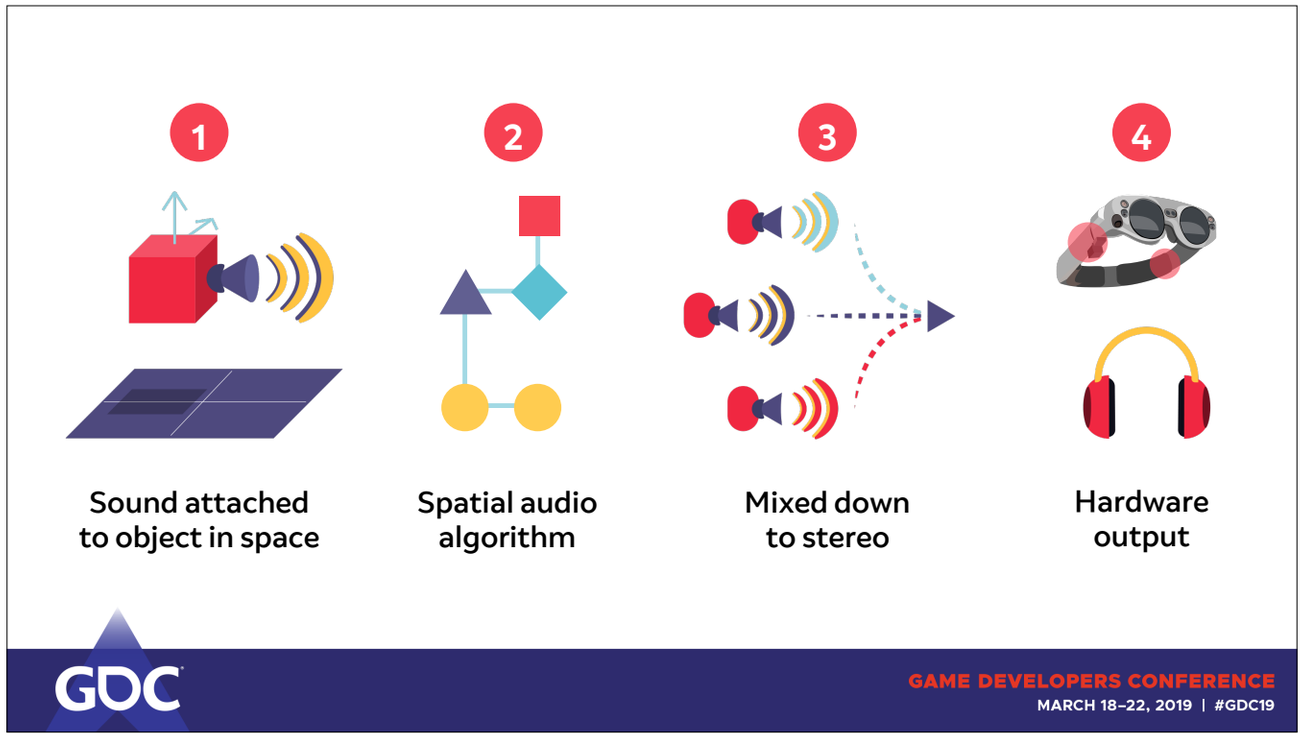
Key Principle #1

- design for a wearable experience
  - ✓ *know frequency response*
  - ✓ *mix for your environment*
  - ✓ *keep it simple*
  - ✓ *layer with love*



Key Principle #2

- maximize use of spatial audio



Now, on to the big buzzword of “spatial audio”

oooooooooooooo

This is an image we use at Magic Leap to describe spatial audio, but it is a simplification, and doesn’t necessarily show all the principles we just discussed that would take place BEFORE we get to step one on this chart

Assuming that part is done, we’ll go ahead and attach the sound to play back at an object’s position in 3D space, then employ a spatialization algorithm of some kind in order to make sure that the positioning feels right

Otherwise, we only get panning to the left and right of the listener position, and volume adjustments (attenuation) based on proximity. No above/below, no front/back, and much less compelling left/right and distance information.

That is then mixed down to stereo, and passed to the device speakers or headphone output.

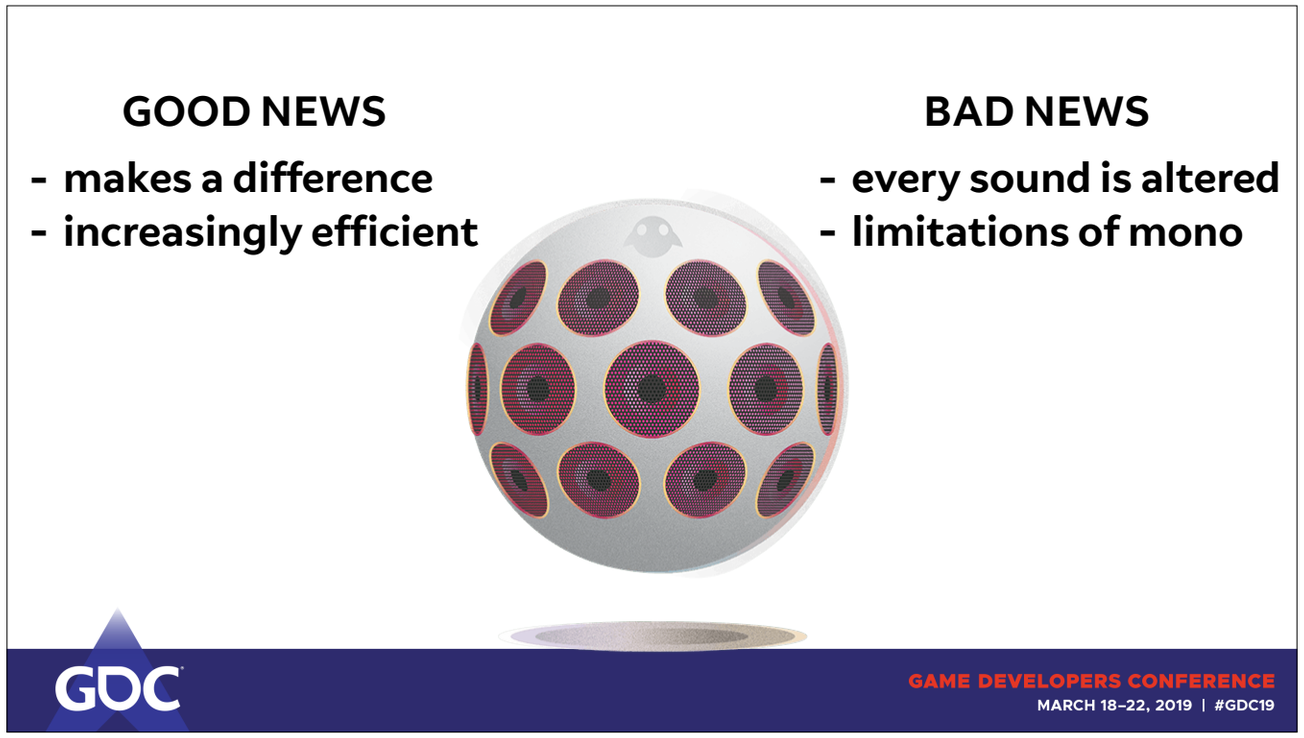
TA DA!!!

Not so fast…

Now it may seem like a given to attach your dino footstep sounds to its feet, and the roar to its mouth, but it’s important.

Few things break immersion faster than interacting with an object and hearing the sound come from an entirely different place

Even things you’d normally accept as static/2D like UI feel strange when you see them floating in front of you. If that object exists in space, even if it’s a menu or some kind of button, the sounds should be spatialized



Now there’s good news and bad news about these spatialization algorithms

After spending years listening to algorithms of all types across a wide range of content, it makes a big difference in allowing the user to localize the object and feel connected to it

We were also nervous that running so many sounds through the algorithm concurrently would tank our performance, esp considering the hardware, but ended up just fine!

This enhancement does have some cost, however, and you may find that after designing a sound in your studio, the playback on device may not live up to your expectations

These artifacts may take the form of a duller transient response, and a potentially strange sounding EQ across the frequency spectrum

In theory, this happens as a result of using a generic set of head sizes/shapes/densities when creating the algorithm, as opposed to a user’s unique ear position and shape

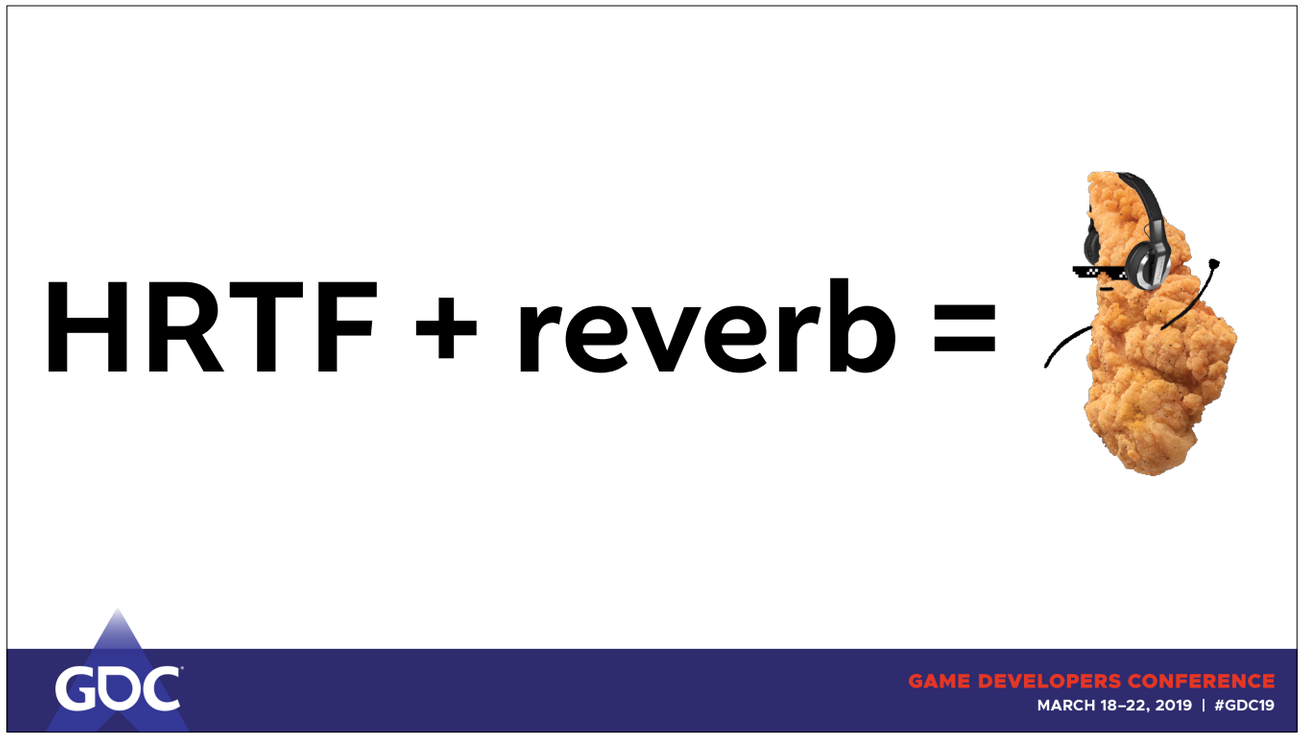
Once we are able to personalize that filter to each user, some of these artifacts might go away

Until then, you might need to revisit your designed sounds in order to make sure they convey the proper effect, given that they will not come out the other end exactly as you designed.

Some of the more difficult sounds in Create included things like wind, waterfalls, and impacts, and required lots of source variations to find something that worked.

I would love to demo some of these distinctions here, but I tried at home with our team, and the point is kinda lost over speakers. So you’ll have to take my word for it for now, and try out Create afterward :D

sounds that we know very well and expect to be played back in stereo will most likely be mono point sources, so we’ll need to get that width through other means



Algorithm gets you spatialization of direct sound, but we need reflections in order to externalize (get it out of our head)

In theory, the holy grail of room tone might be the ability of your device to understand not only the bounds and surfaces, but also materials in the user's playspace

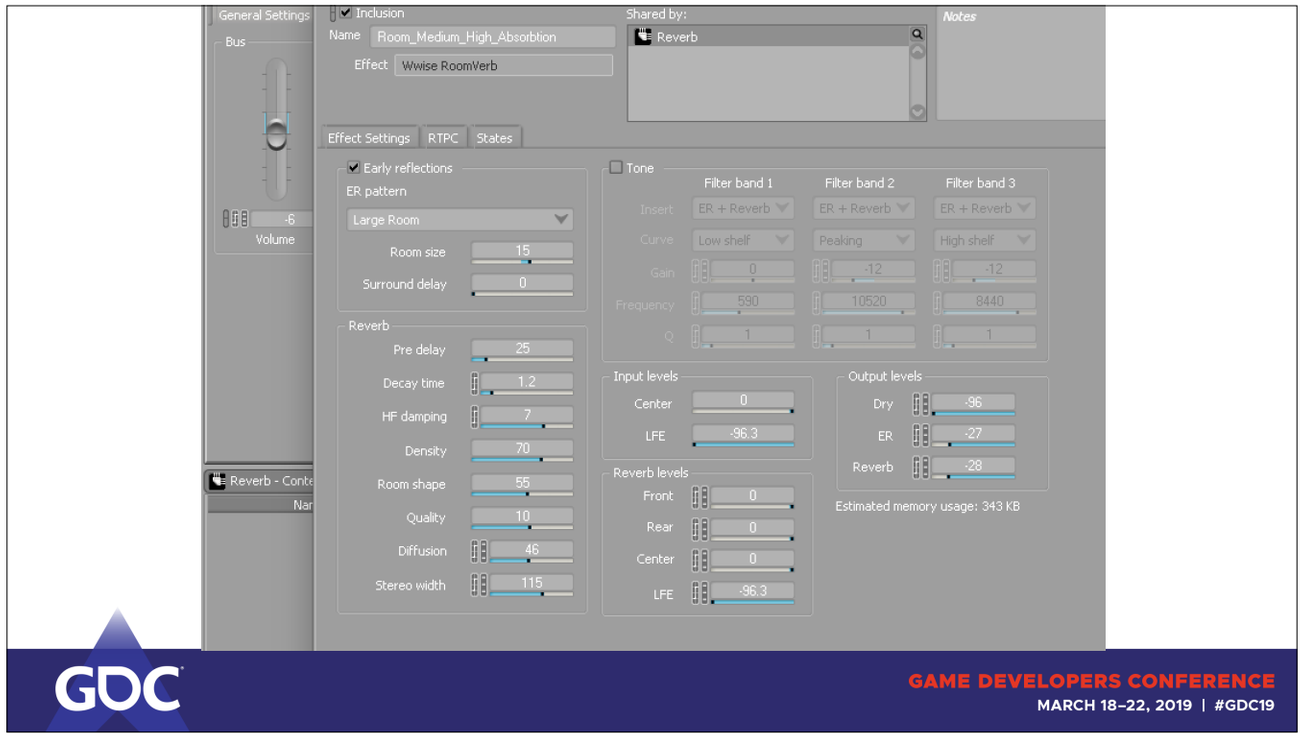
Armed with that information, your audio system could generate early and late reflections with precision, uniquely tuned to each user's environment

Not only would you get walls and ceiling, but the tops of tables and the reflective surface of a computer monitor

This does propose an interesting philosophical question, however - how accurately do we want to model our generally crappy sounding spaces?

In the short term, this can be achieved to great effect by simply using an ambisonic reverb to taste.

It might not be accurate, but it will sound good and provide the desired effect of gluing your content together in 3D space.

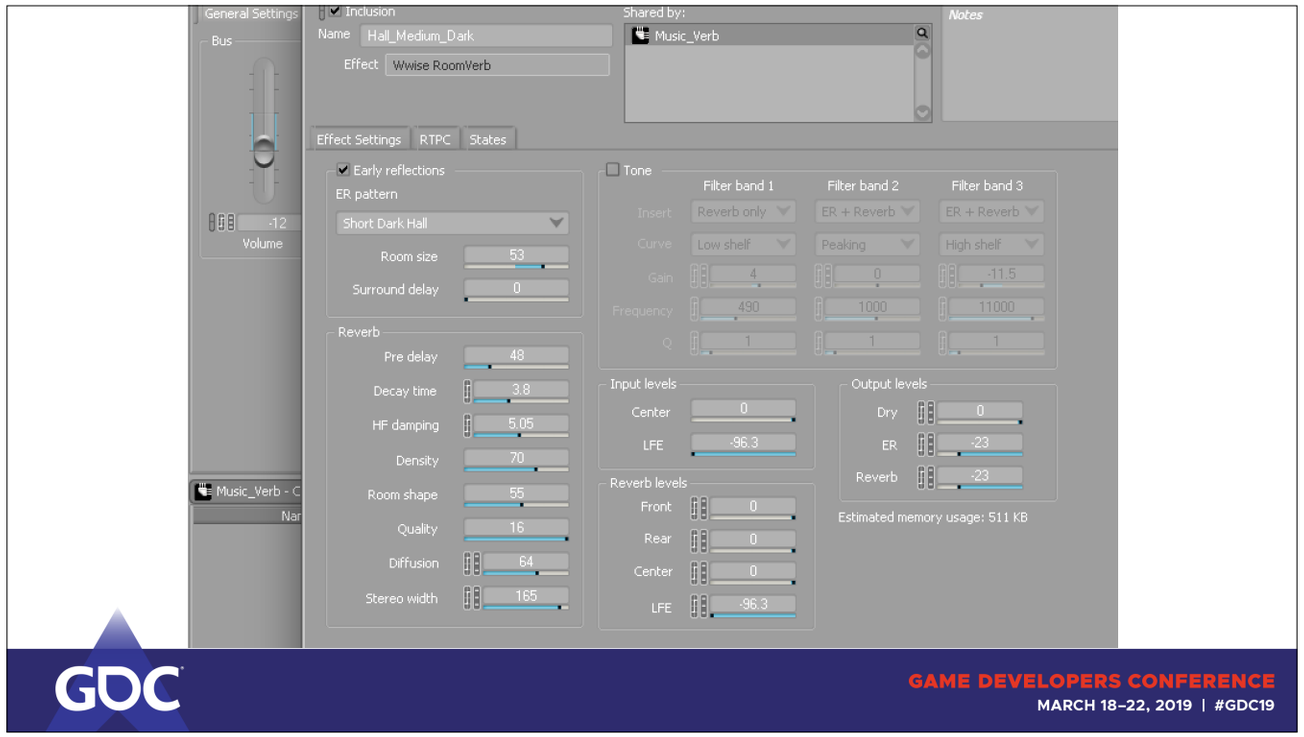


CAVEAT: This also may not translate very well on big speakers in a big room, but again you’re welcome to try it out on device after the talk :)

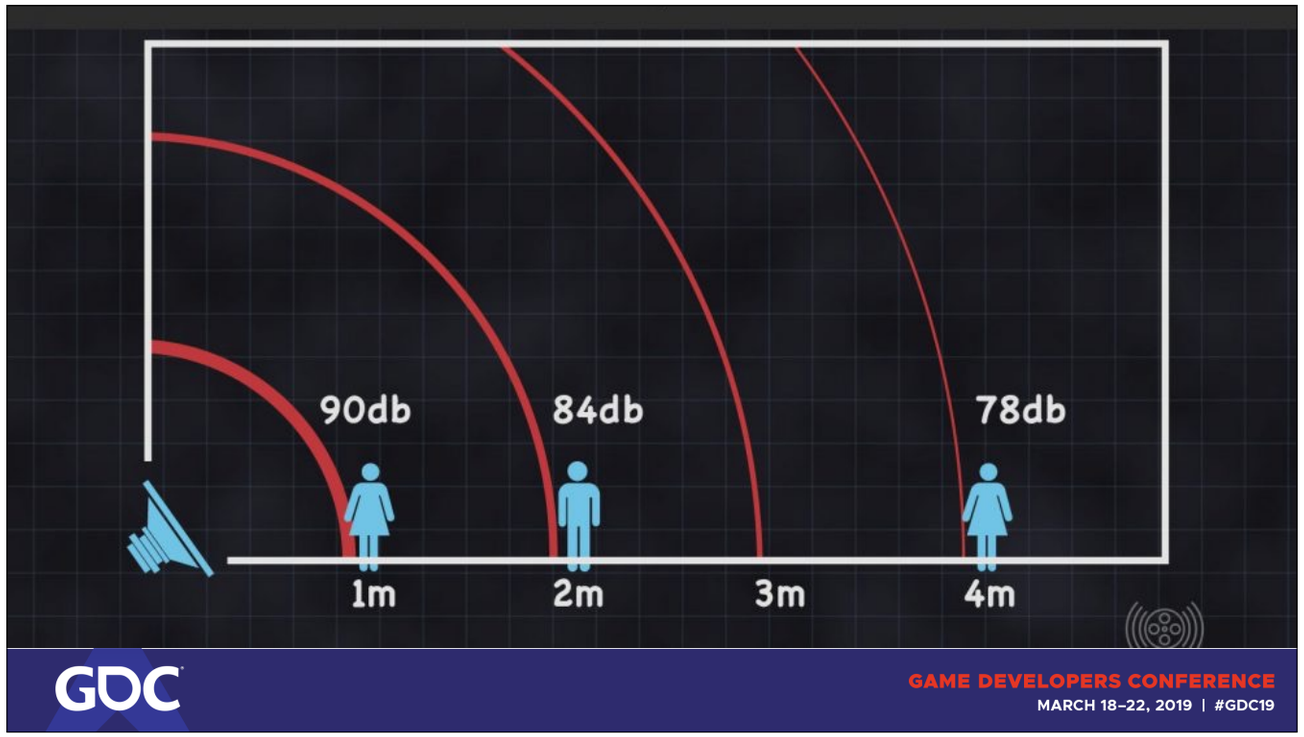
In Create, we used a third order ambisonic reverb that all objects were sent to (separate bus was used for music playback) and adjusted to taste.

It wasn’t based on any specific dimensions, but sort of generically tuned to a living room style environment (see Wwise preset above!)

It’s almost felt, more than heard, but very important to make sure that your content didn’t just feel like a single mono point source of whatever you happened to be standing in front of.



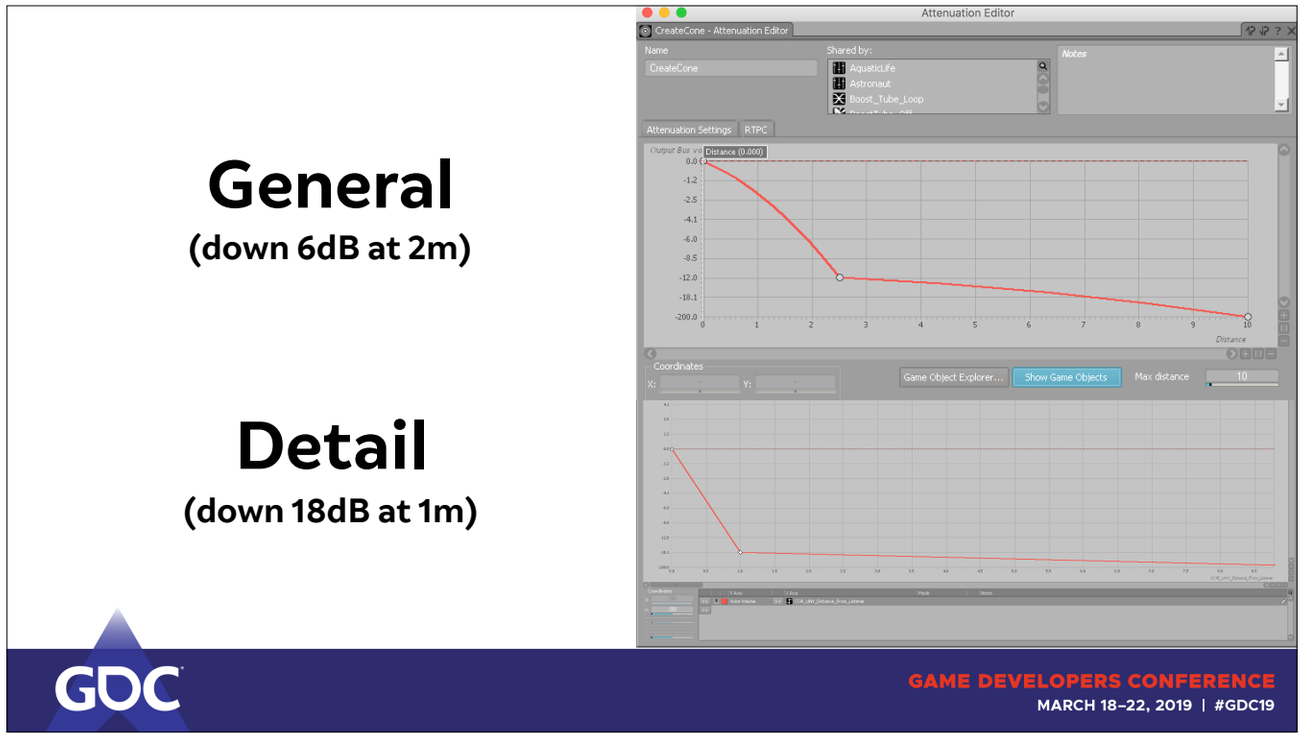
music verb was more “musical” in nature, but even quieter in the mix - we’ll circle back around to music in a bit



ok, so we've got our sound designed to play nicely with the HRTF algorithm, attached properly to an object in space and run through the spatialize and ambisonic verb.

what about attenuation?

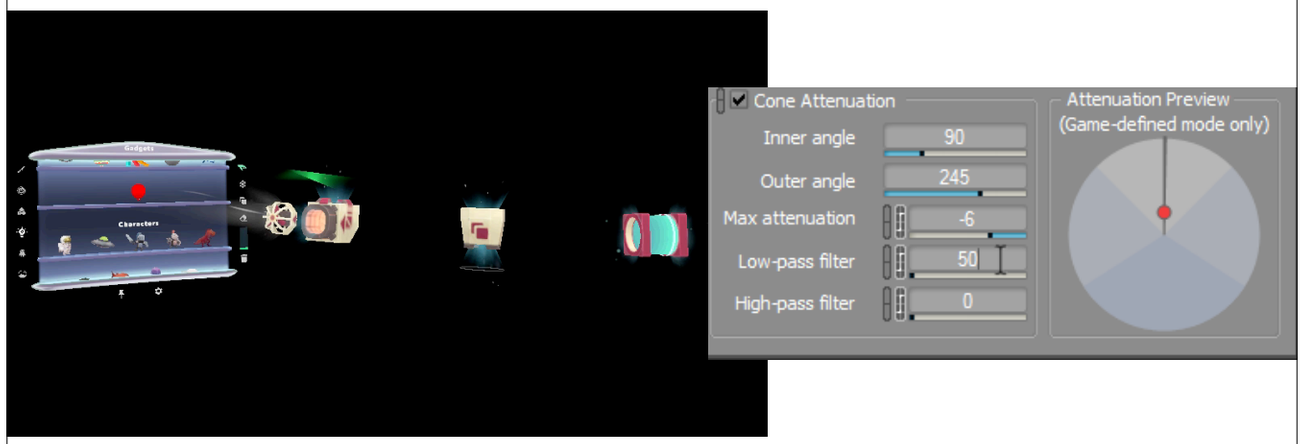
inverse square law teaches how we perceive that in nature



In practice, we started there, and then dialed in using wwise remote while running on device, which was super helpful - down 6dB by the time you’re 2m away

this illustrates a more aggressive attenuation curve for the detail layers mentioned earlier - down 18dB by the time you’re 1m away

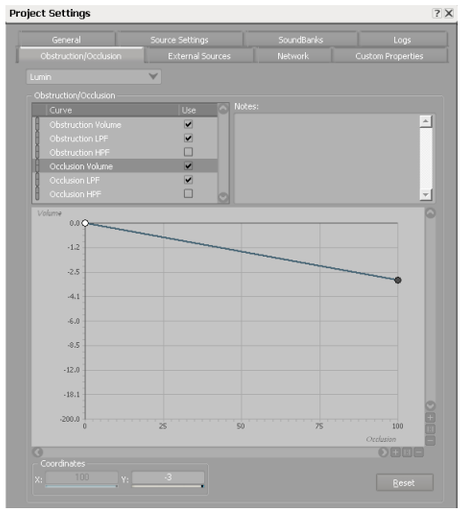
Cone attenuation

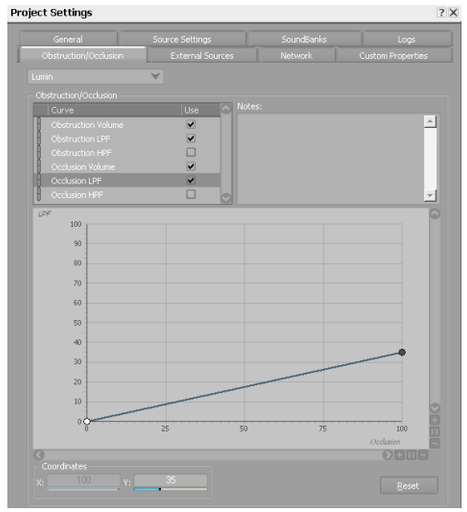


In addition, As sounds are not omnidirectional in nature, our brains will also expect that as we orient ourselves off axis from objects, the sound of that object should change.

So the cone attenuation feature in Wwise was really useful in dialing this in

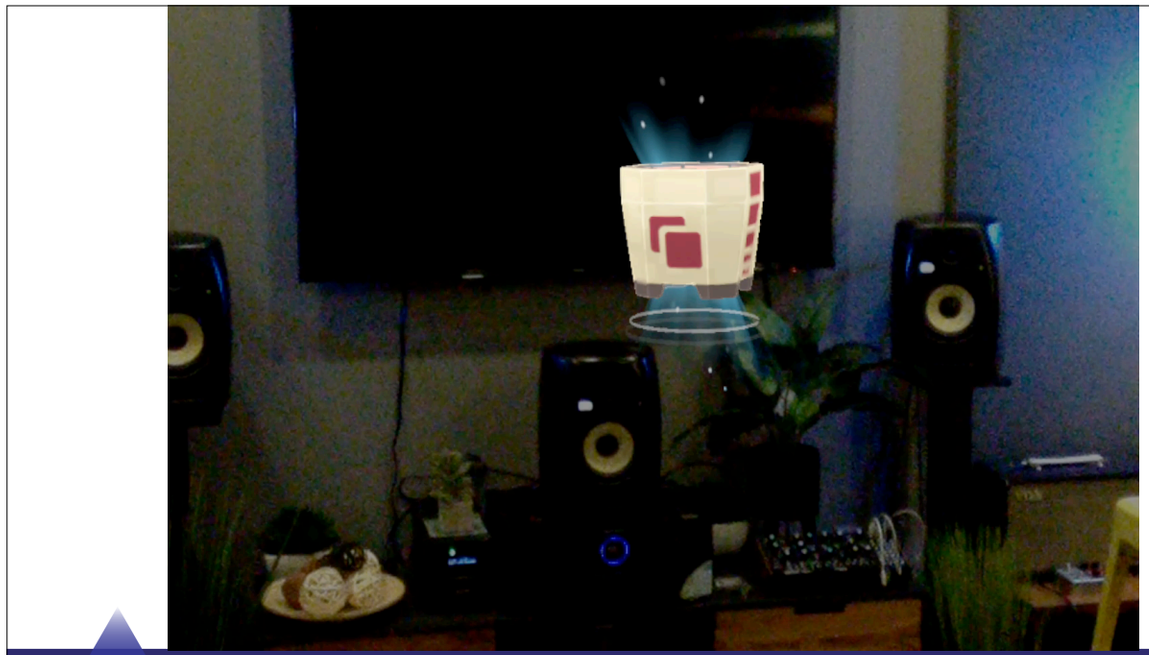
Occlusion

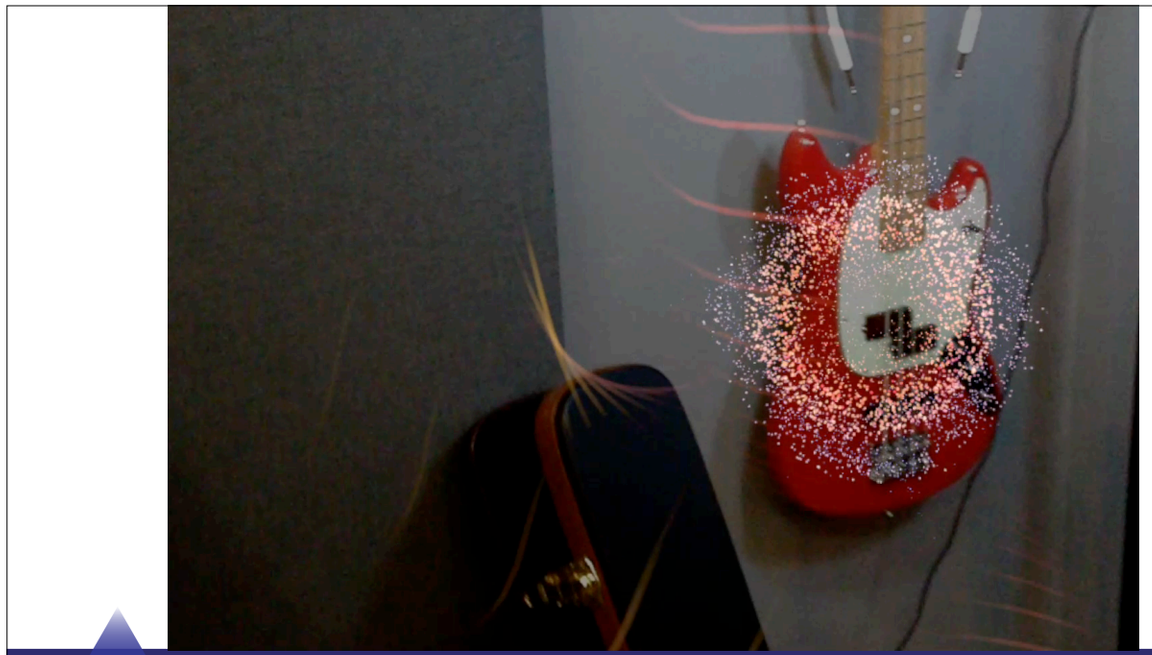




Real and virtual objects should also apply some element of occlusion to the extent that your experience allows (Sennheiser Ambeo app does a great job of this). We’re using a simplified version of this in Create with Wwise driving a low pass filter as a raycast from the listener position comes in contact with an object in Unity.

One caveat here is that these things are very hard to get right because we turn our heads really quickly, often without even noticing it. If we’re not able to make that filter feel natural, it may be more distracting than if you just let them play out omnidirectionally without occlusion.



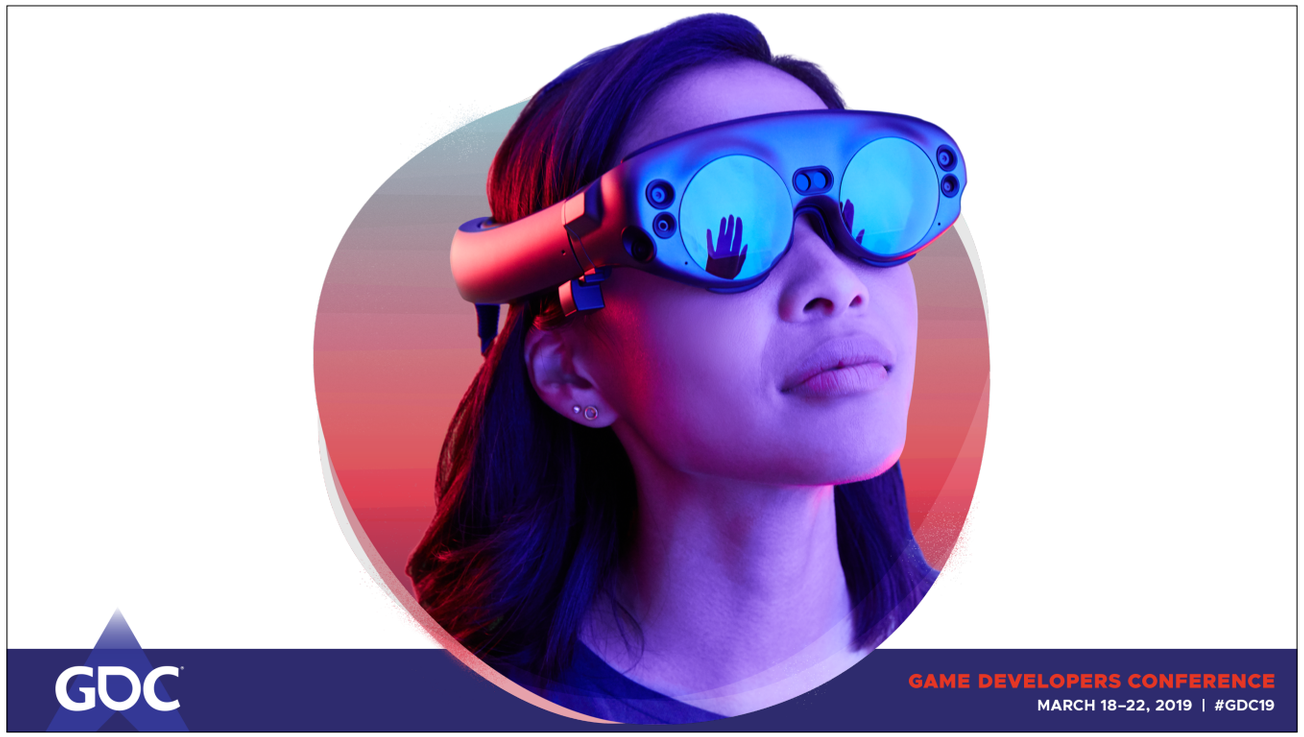


Key Principle #2

- **maximize use of spatial audio**
  - ✓ *place sounds at XYZ position*
  - ✓ *use a spatializer*
  - ✓ *add ambisonic verb*
  - ✓ *natural attenuation (cone and occlusion)*

Key Principle #3

- use audio to help support design

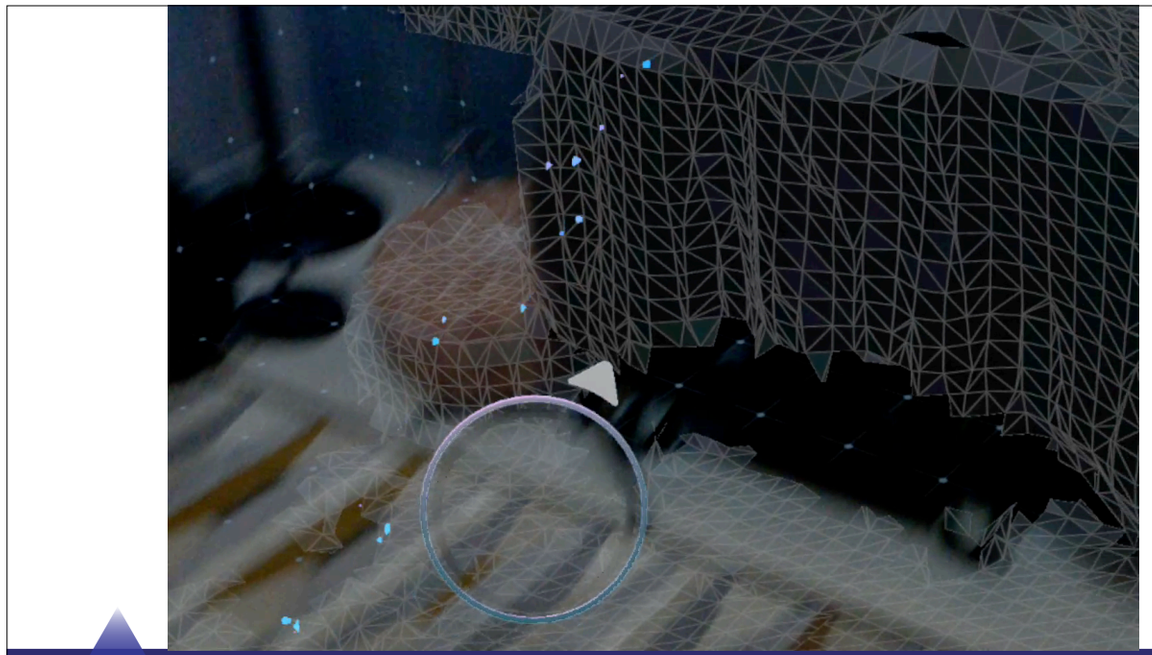


Working in these new spaces, we’re having to teach users a lot of things we’d normally take for granted.

We’re really good about knowing that hitting ESC on our keyboard will probably get us to an options menu, that right mouse click will aim and left click will shoot. Spacebar jumps, shift will sprint.

In mixed reality, even with some platforms using simple controllers with a few buttons, it’s not always clear what the user is supposed to be doing and why.

Audio becomes an integral part of the onboarding and learning process as users are trained on the core mechanics of your experience.



There are some psychological barriers to using this new technology.

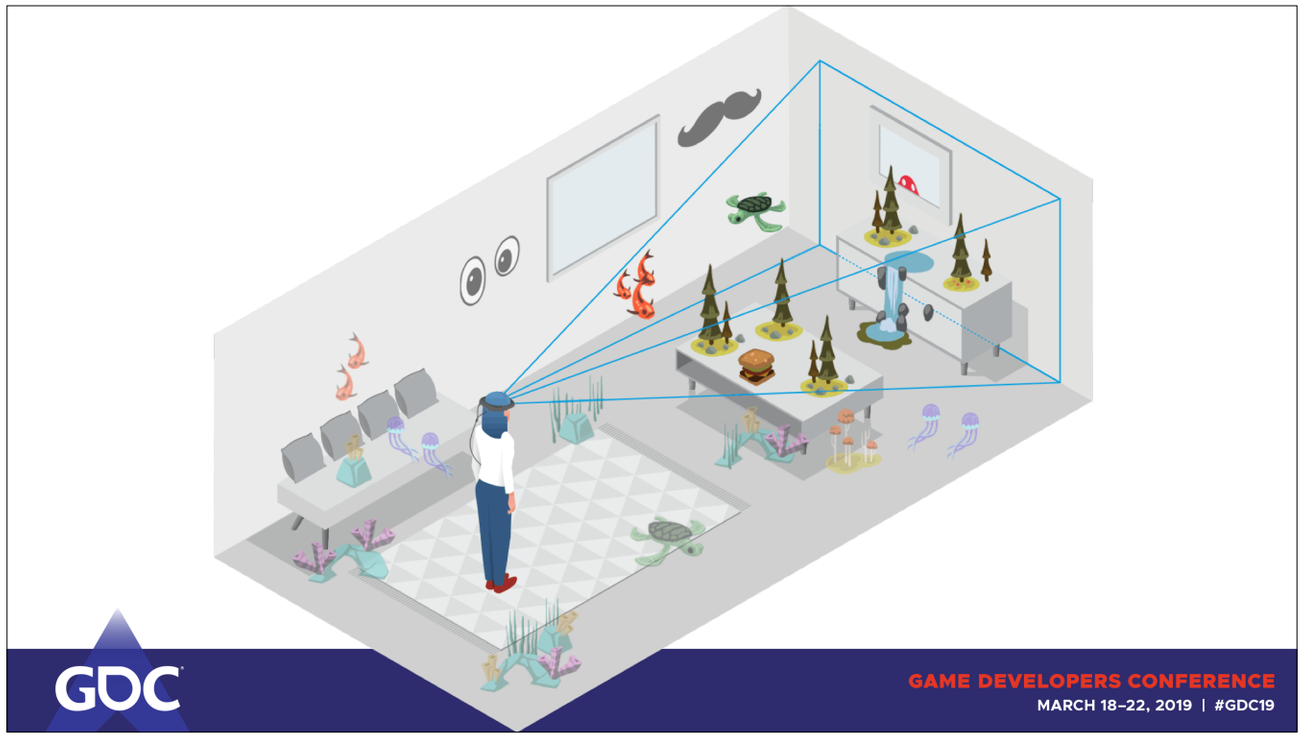
It feels strange to wear a headset, even if you’re able to see through it, and users aren’t sure what they’re supposed to do, where they can walk, what they can interact with.

They might not know that audio will be coming out of the device at all, and from where.

It’d be very easy for us to shock, frighten or otherwise make a user uncomfortable with audio playback.

From the first moments of Create, there’s a simple ambient musical bed playing, and light tones and whooshes as the user explores their space during the onboarding meshing sequence.

This helps to ease the user in and get them to move around, understanding that the sounds they hear are coming from actual interactions in their space.



Along with the potential awkwardness of wearing the newfangled headsets, the field of view that visual content can be displayed on will be constrained.

If you were playing on a screen, you would not expect to hear much that isn’t currently being displayed.

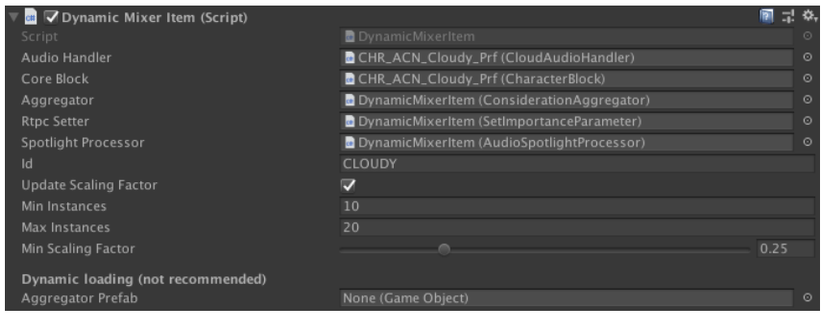
However, as soon as you are wearing a device, you expect your content to persist even when outside your field of view.

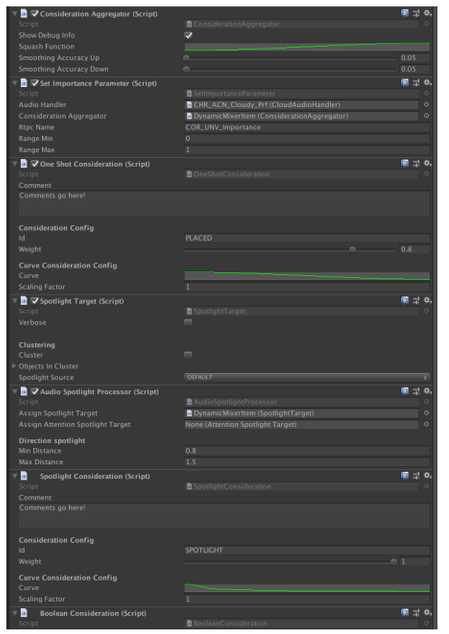
In Create, objects can exist anywhere in your room, and behave in a variety of emergent ways you don’t expect.

Using audio effectively allows us to clue in users to those moments of surprise and delight and help sell the magic of this content truly being part of your space.


Worlds

Context-driven mixing





So we spend all this time crafting everything to be just so, and then essentially hand over the camera to the user, to focus on whatever they want, as close as they want, for as long as they want. This can happen to some extent in a lot of games, but it is never so painfully obvious as in VR/AR/MR

There is very little you can predict about how a user will explore the content you create.

In Create, this means that until a user reaches an arbitrary performance limit, we could be playing a LOT at once, which could easily become overwhelming.

In order to accommodate this, we try and understand the context of what a user might be experiencing at any one time.

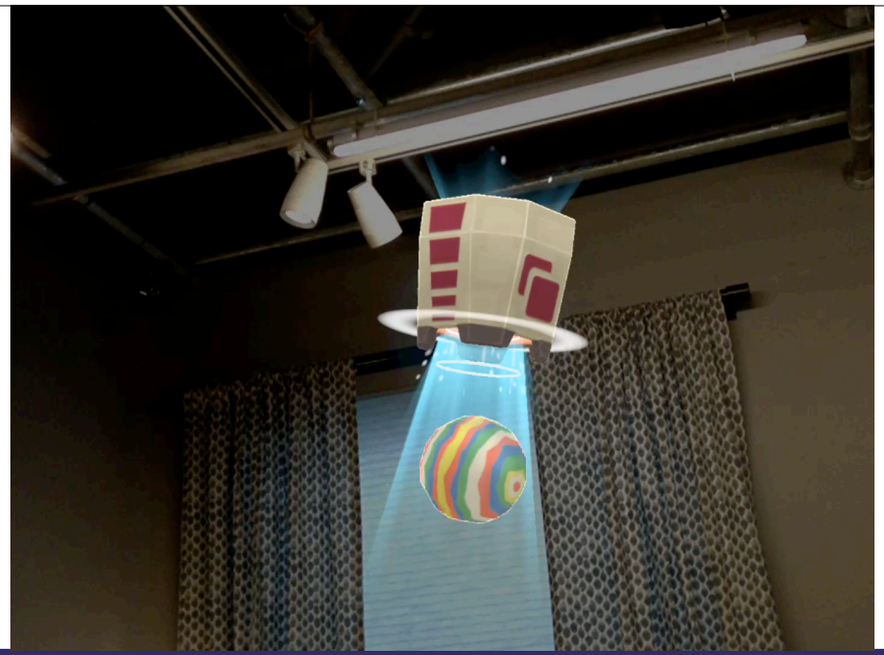
We can know things like headpose to determine what a user is focused on or interested in, and dial sounds outside of a specific angle down to taste.

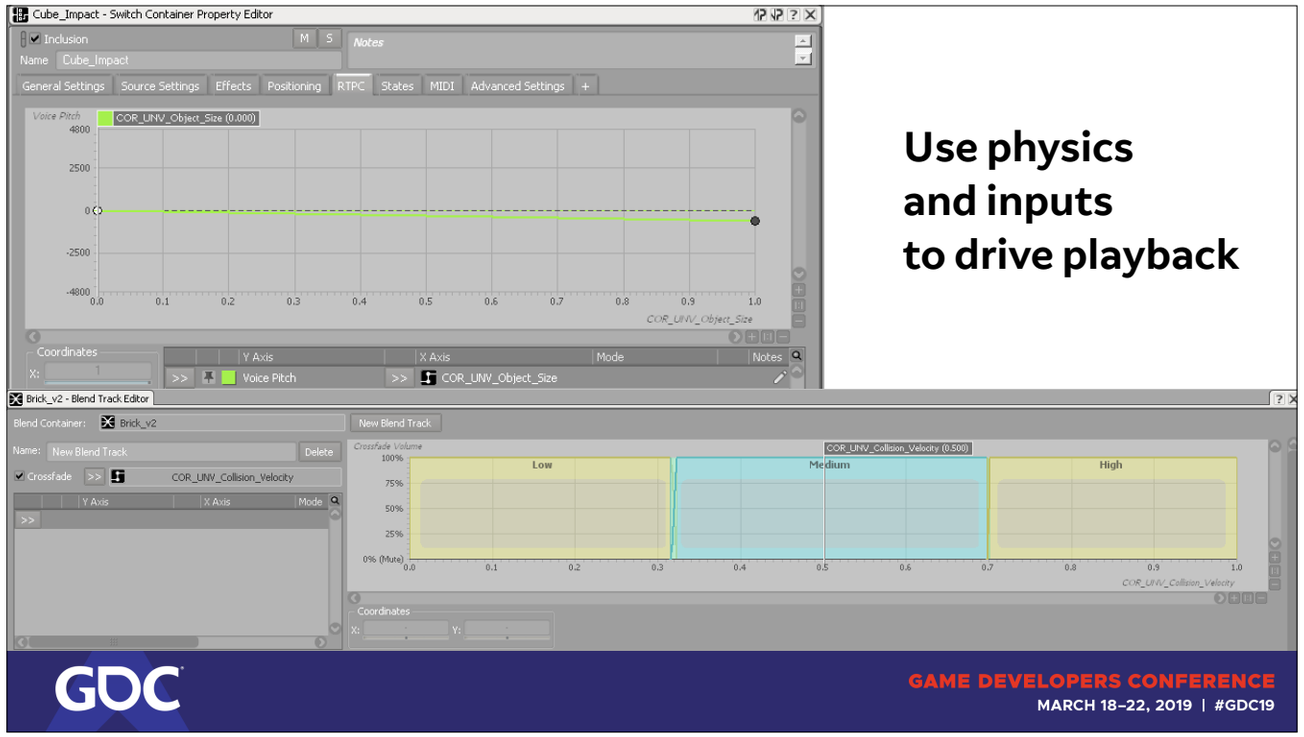
We can also ensure that an object currently being interacted with is given a little bump of volume or focus, or highlight something interesting happening even if it’s out of frame.

We can also keep track of how many instances of a particular object are in the scene and globally clamp them down a bit (a squadron of knights or UFOs, for example).

These elements are not new, but crucial to maintaining a quality mix when you have little to no control over the traditional frame.

These are some of the tools in unity we built to help dial in this dynamic mix, and here’s an example of what it sounds like in practice. You’ll notice that as the camera/user pans to each object, what’s in frame comes into focus audio-wise, while other elements are kept audible, but pulled back.



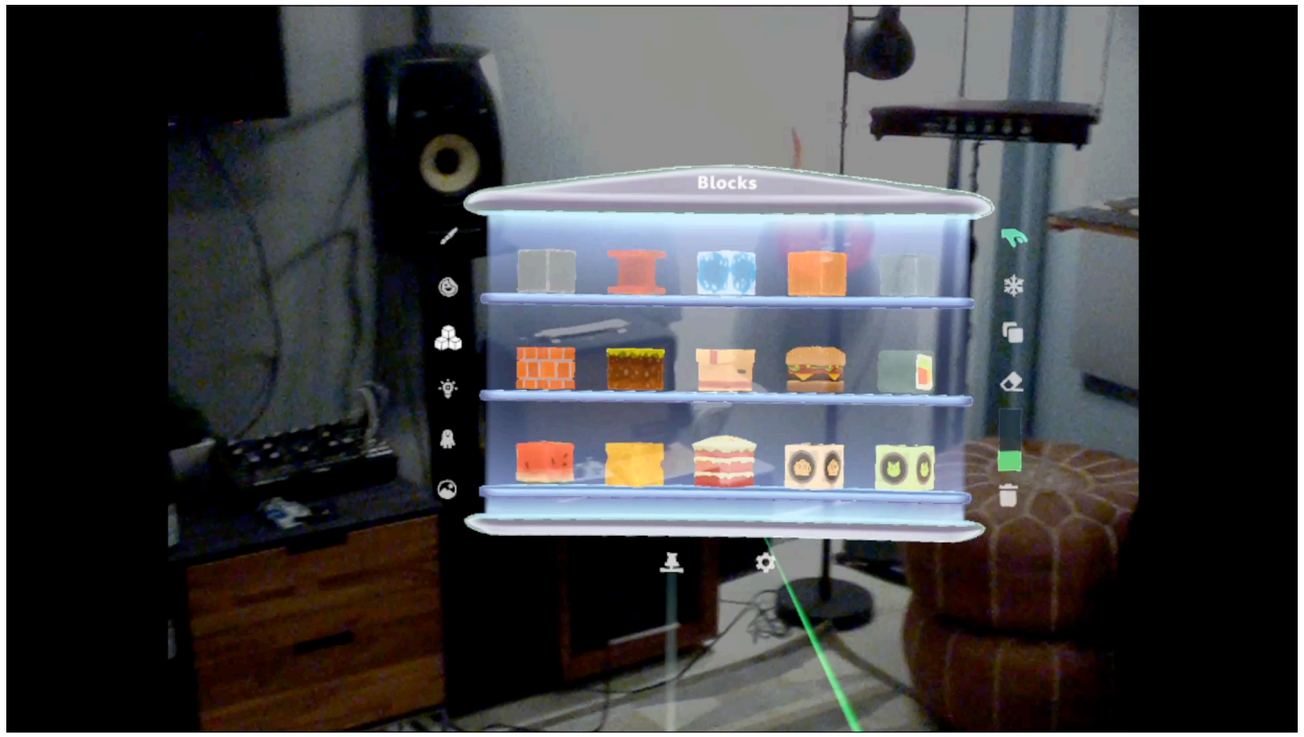


We also want to make sure that we’re accounting for a users intentional actions beyond just where they place the camera. This means making your content as reactive as possible to the physical space it occupies and any inputs you may be giving to the user.

simple things like tying volume to collision velocity, or pitch to size of objects allows you to hit those user expectations in a meaningful way.

Allowing a simple thruster loop to continue to pitch itself up in concert with an object’s elevation (which could vary wildly based on a user’s space) creates a unique scenario that may only happen in that environment.

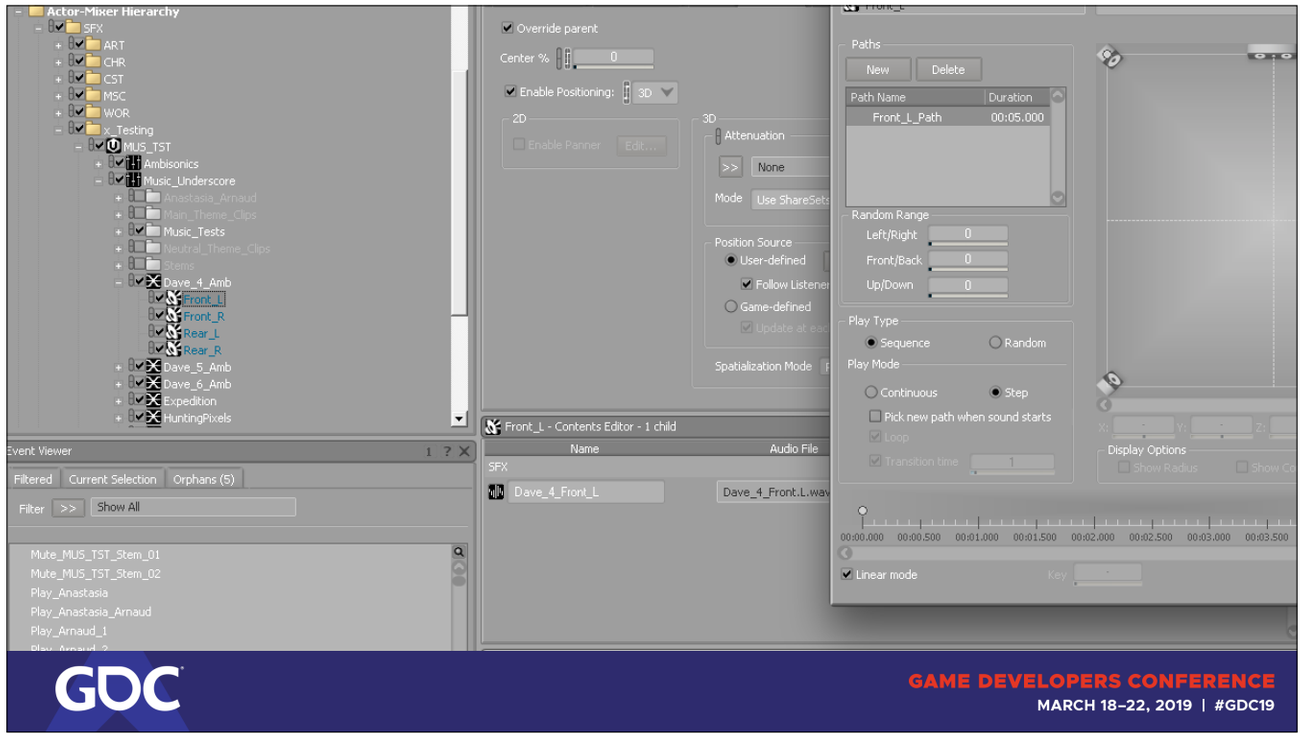
Using the control to hit a music cube in Create allows us to tap into the speed of the Control and vary playback volume accordingly. Very simple, but you’ll notice if it’s not there.


Blocks

music proved an interesting design challenge and a polarizing subject as we tried to align with the experience

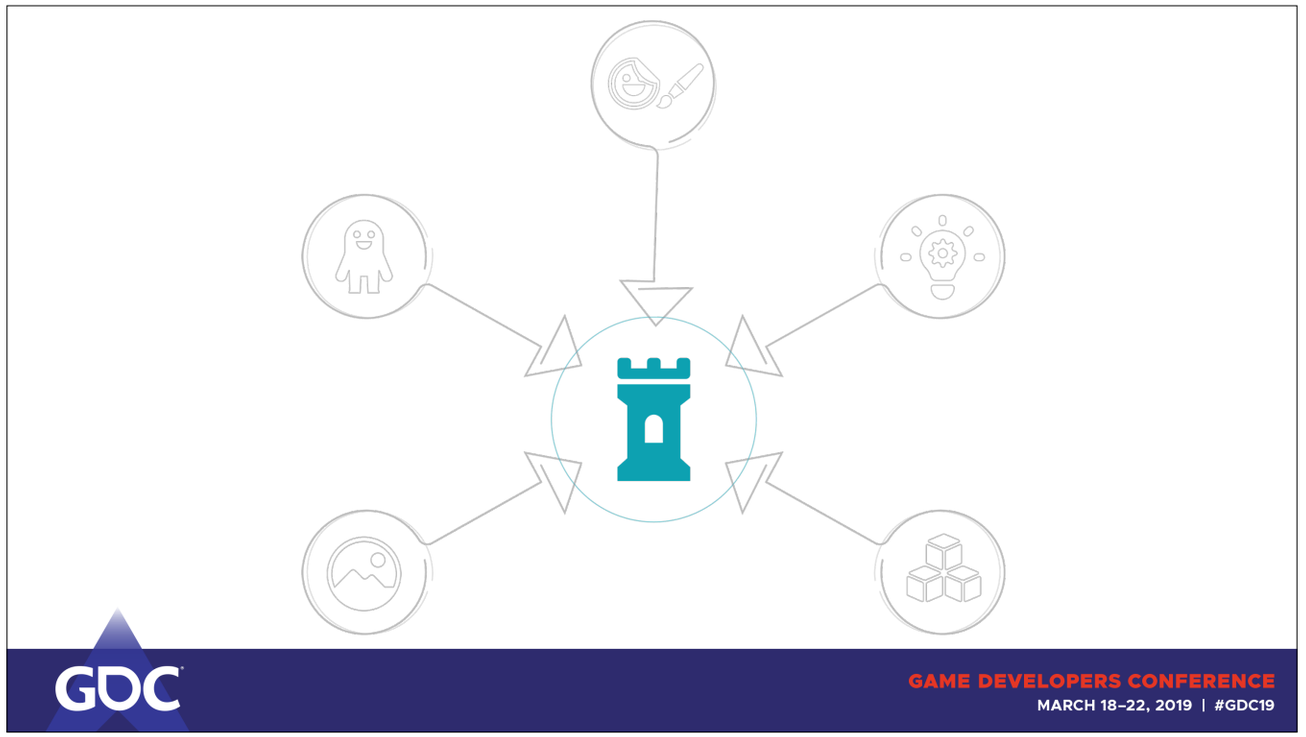
our ears are very attuned to traditional instruments, to the point that many VR/AR experiences will stick with stereo music playback to avoid degrading the palette the composer and musicians have so carefully created. Even though the music might not be present in the same space as the rest of your sound design, it’s totally plausible to assume your user will accept that soundtrack played back in a traditional sense.

We’ve done a variety of experiments in-house to suggest that the design of the experience will probably play a big role in the route you take, but it should be a point of early conversation with your music team



we ended up creating front/rear pairs, and positioning them sort of in a quad layout, with each object being sent through a spatializer, then through an ambisonic reverb to glue together so you had this sort of spherical music bubble following the user.

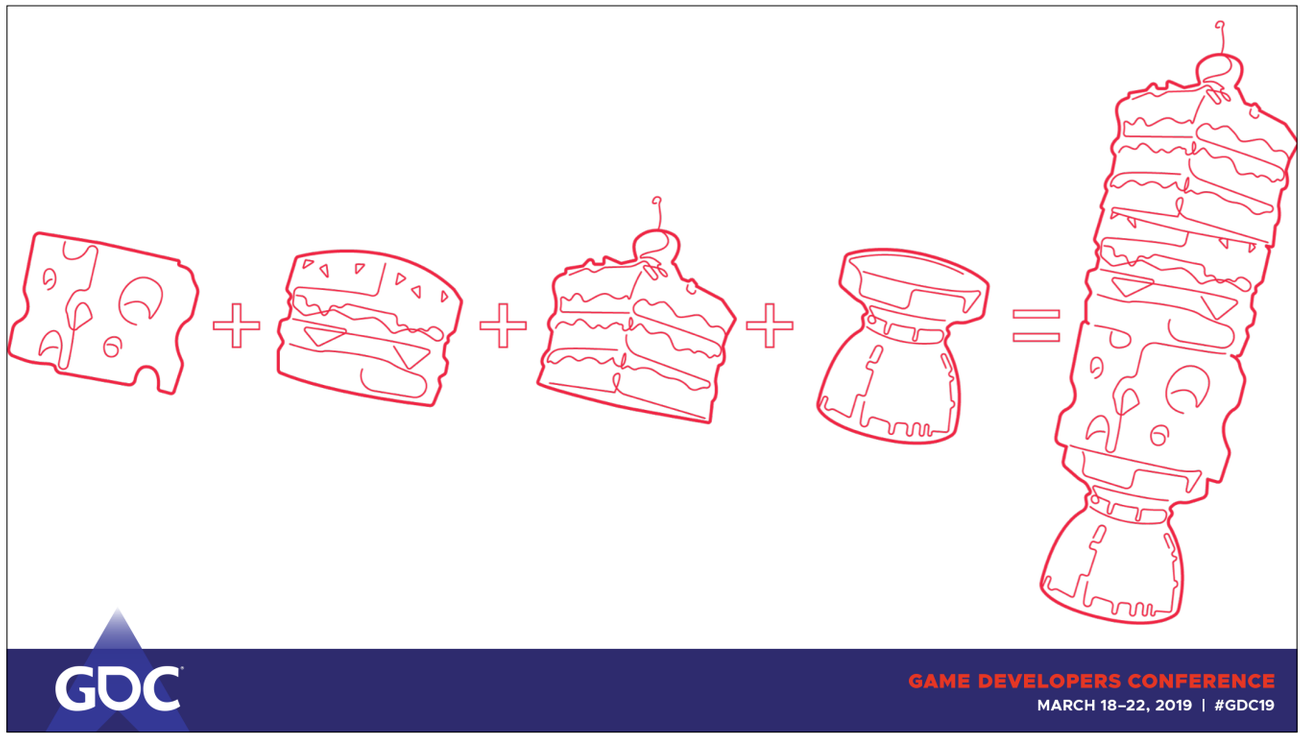
It’s not something I can demo here, but it had a nice effect of extending the music out into a user’s playspace, instead of being stuck to the sides of their head.



ok, but what about dynamic music?

5 categories = 5 layers?


OLD FOOTAGE!



that worked ok, but just because I want to play with gadgets, doesn’t necessarily mean I want a busy soundtrack.

we also noticed that as scene fills up, music gets busier, but so did sfx from those object we added

nice emotional release on “clear all,” prob relief

needed to make user feel comfortable exploring, and that wasn’t doing it

lots of interesting characters, let’s use those!

main instrumentation wouldn’t change

small piece of unique instrumentation per character

Tarrey Town!

could even attach to character

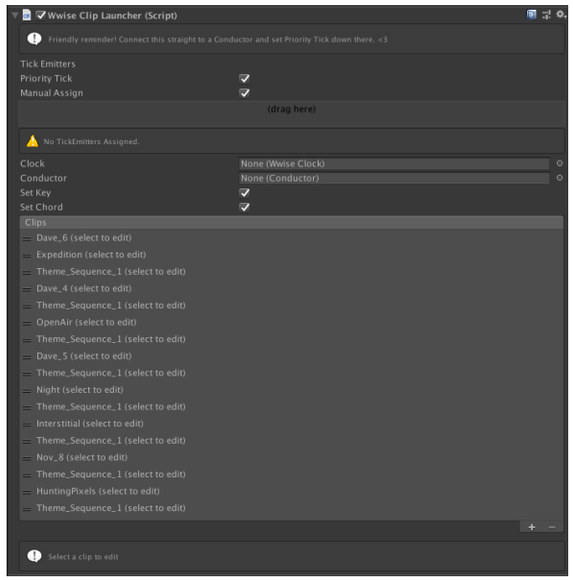
could have lotsa characters tho…

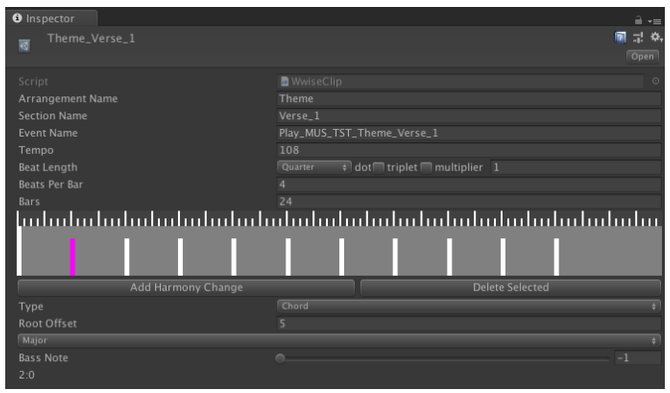

OLD FOOTAGE!

Still pretty intense, and skewed away from the comfortable, inviting exploratory vibe the design called for.

In the end, we found that users responded to a light, evocative, almost safety-net-like background track to allow them to feel comfortable exploring their space.

Musical Metadata

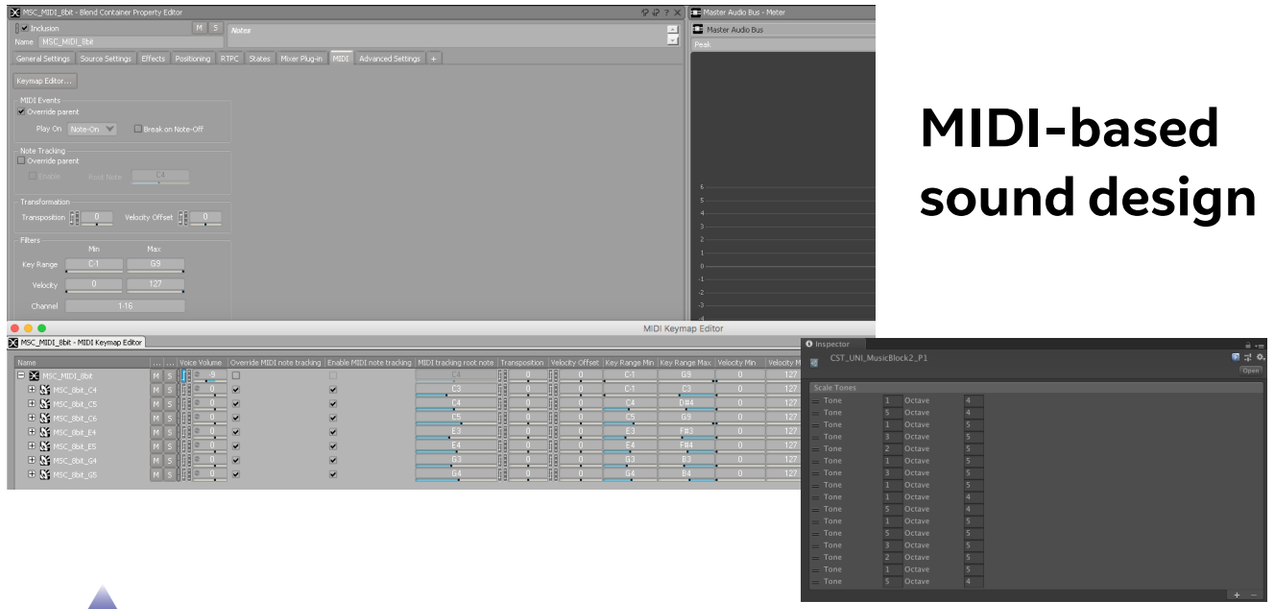




But we still wanted to music to feel part of the experience, and not just something tacked on as wallpaper. So we chose to have static music, but track metadata that we can then use to drive different elements within Create

So what you see on the left is a clip launcher with a bunch of different tracks that we create custom rules for in terms of how many times they loop, random chance that they’re played when reached in the sequence, etc.

Then on the right, you see an example of the metadata captured for each of those clips. We’re tracking tempo, time signature, total number of bars in the segment, and any key and chord changes that happen.

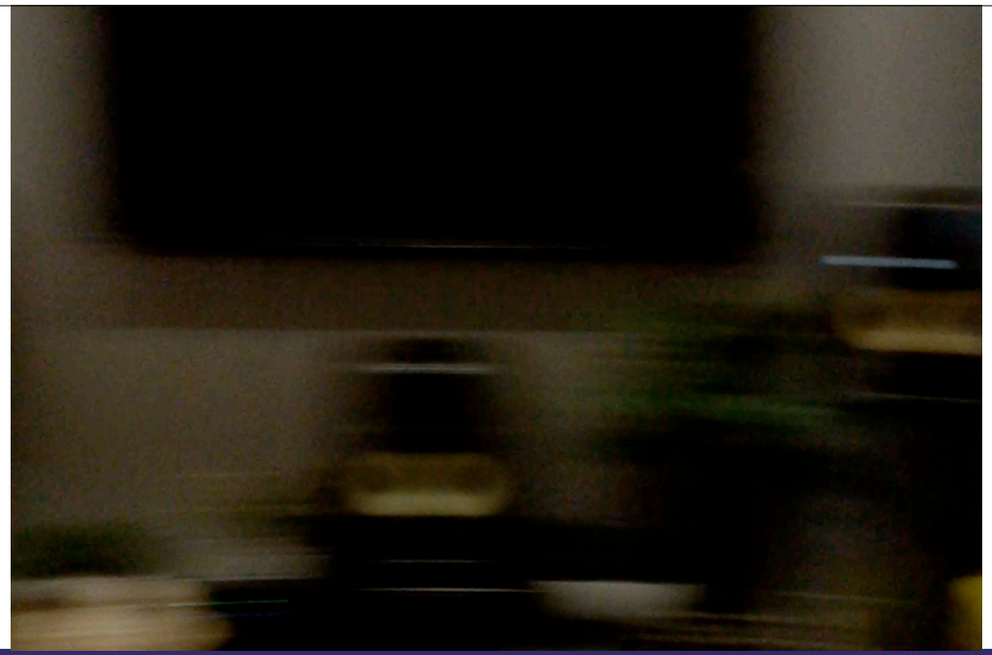


This data then gets passed to Wwise, which uses MIDI to change the pitch of UI elements, music blocks and other character-based stingers so that they’re always playing something complementary to the background track.

So while the music itself wasn’t dynamic, as the user interacted with their world, little pops of musical color that they controlled would complement the music.

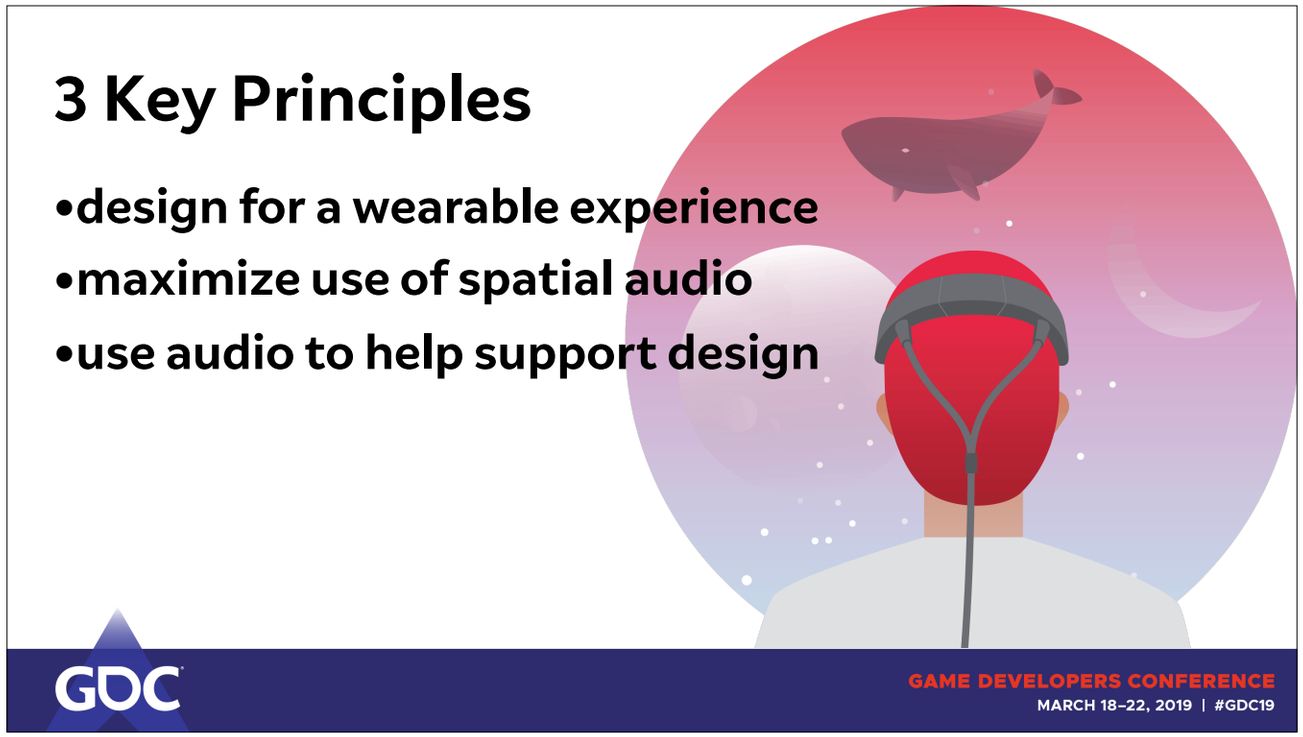
This would allow people to sort of play the menu like an instrument, or physically bang on music blocks with the controller.

Here’s an intentionally bull-in-a-china-shop illustration of what that sounds like



Key Principle #3

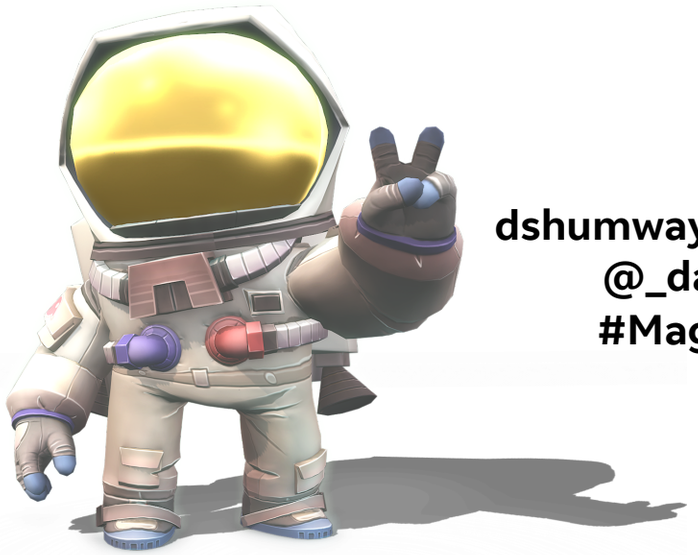
- use audio to help support design
  - ✓ *emphasize UI and on-boarding*
  - ✓ *expand beyond FOV*
  - ✓ *build systems to drive playback and mix*
  - ✓ *music can easily skew experience*



Reflecting back, we probably chose one of the trickiest ideas in making a physics sandbox on a brand new device with an entirely new hardware and software stack, and setting it up as the first experience a user will have on the platform.

We had no idea what would work, but wanted desperately to make audio an integral part of this new medium and have it resonate with users.

We developed Create not only as an introduction to the Magic Leap platform, but as a way of kickstarting the minds of developers everywhere. We’re proud of what we made, and we’re looking even more forward to seeing what you all do next in this brave new world of mixed reality.



dshumway@magicleap.com
@_daveshumway
#MagicLeapxGDC