

This document is the supplementary data for the paper *Designing Sound for Social Robots: Candidate Design Principles* by Robinson, et al. It contains transcripts of three expert interviews with sound designers of commercial social robots.

Table of Contents

Ben Gabaldon – Cozmo and Vector 3

Jeshua Whitaker – Jibo..... 25

Connor Moore – Kuri..... 31

Ben Gabaldon – Cozmo and Vector

When you started the design process, what was your main goal or your main challenge?

Would it benefit you if I spoke about them from different points in time? Because they were very different approaches.

Sure!

For Cozmo, I mean I touched on this a little bit before, but I came in very much with no understanding of what the product would be when it was shown to me. There was previously an audio director at the company and he was, I think, focused on purchasing equipment and managing a team and there was another product that he was focused on. In Cozmo, as I said, there was nothing on it. There was no personality. There was no concept of what it would be, so it was very much the seed of a product design. For example, we had an animator whose only purpose was creating a custom animation tool that gave him the ability to drive forward, turn and just do very robotic things.

Coming into this with no experience in robotics, I didn't know how to improve this. I was just working on the sound of the robot. But in addition, in our early discovery phase we were thinking about how robots in our historical experience communicate and one of the best examples, of course, is R2D2, and there are parallels that can be found in Chewbacca. These are two characters, prominent characters, in an extremely famous franchise that, when they communicate, have peripheral characters that basically pair it back or interpret their information. And they do it in such a way that they're effectively repeating everything that they said in a way that is creating exposition to the audience. In one of our earliest explorations Cozmo would have these extremely disconnected, unique and foreign sounds but there would be - because there was a companion app - a third party that we were calling A5. A5 was a woman's voice that, anytime Cozmo would make a sound, could argue

with Cozmo and he could make completely indistinguishable sounds but you had an interpreter. But that creates this weird friction where you're now having to sell this whole universe, this whole fiction, which is really hard to do without a true narrative.

We didn't have a background story, we didn't have any of this, so it really came back to "What's the focus?". Why is Cozmo, this robot, the focus? And so, my approach was just to create a sound that appealed to people. But, coming from a video game background, I didn't want to start with just a single sound, I wanted to create a full matrix of sounds. So it's a lot to take on when you don't have that kernel of the sound, so what I kept doing was creating a full suite of angry creating a full suite of happy and, to me, I could tell because I knew what my design approach was but when you play it to somebody else, all the thinking that I had going into it was not being received by others. So, it very much early on became a group effort. I started making smaller sets of sounds, sharing it with a group, and trying to find appeal before we scaled it to emotion; trying to find the sound of Cozmo. I'm sorry you might have to reiterate the question.

The design approach. Starting on a sound for a robot. The main challenge when starting the whole process.

I think, for Cozmo the main challenge really was the fact that we, as a team, creatively had no foundation. We were all just trying to create content for something that we didn't understand and until we had purpose it was just "Everybody's creating animations, audio in a vacuum." And you put it on this robot and there's still no value to it. You might have really great animations. You might have really great sound, but there's no purpose and I think that's a question that got asked especially later on down the road when we started bringing in outside outlets to review the product and they say "Well, what does it do?" So, for Cozmo the biggest challenge is that we did not have a creative foundation to build around and that created a lot of wasted time, a lot of wasted content. But once we found what we considered the "character," the "personality" and really "what Cozmo is" – I know better in hindsight. Cozmo, as a character, is a spoiled bratty child and what Cozmo is good at and

what people enjoyed most about it even in user research is not when Cozmo would lift the cube correctly or make facial contact correctly and say your name or something. That was amusing, but that's predictable. It's when Cozmo gets upset and demonstrates independence and that's when it breaks from expectation. So, developing the character of Cozmo as being a highly emotional, reactive, independent character was a huge challenge and, you know, they were building back stories. "Cozmos are being built in this lab by this evil doctor and this evil scientist" and all this backstory –just none of that helped, we need to understand Cozmo because until you respect what the robot is and who they are it was impossible to create compelling content. Did you have any questions about that?

For Vector it was a different story?

It was an extremely different story. With Vector, what actually was happening, was we found what people wanted out of Cozmo, that didn't exist. For example, 1) Cozmo doesn't operate independently. You have to be tethered to a phone. A lot of people did not like that. 2) Voice commands, you know being able to actually communicate in some form. So, with Vector, they included the four on-board directional microphones, two drop sensors, and they actually had the embedded AMD chip in the robot, so that it could function independently. So, when Vector was being outfit, Vector initially was going to be Cozmo 2.0. It was going to be Cozmo with much higher functionality, so I started developing ways that Cozmo's communication could develop over time and using a lot more external stimulus to develop the audio system. Once we decided that Vector is a new character, we had to pivot very quickly, so the challenge became understanding all of the failures of Cozmo and taking the time to identify character. We had a character director at that time. He created character boards and he was referencing things like "this is more like a pet," "this is an older audience," "we want to appeal not to children who will play little mini games or be amused by Cozmo throwing a cube and freaking out." "This needs to appeal to young adults, people that wanted robots their whole life." So, we created the foundation of a character first. He was referencing smaller critters like a fennec fox or a parrot; intelligent creatures that can communicate in their own way. That way we thought about when creating animations and

creating sound – they were much more shy and reserved. But another challenge, the honest answer to the challenge for Vector was just a very rushed timeline.

You mentioned that you started thinking about ways for Cozmo’s sound to develop over time. What do you mean? What's the time frame of that?

I was still exploring possibilities how this would happen. I was creating different video demonstrations of what this could look like and I'm just trying to think about what are things that could happen for one Cozmo user that another Cozmo user won't have, that would create two potentially slightly different Cozmos. Some of the things that I was thinking about – I started getting a better idea, an understanding, of how to record human vocals to create more intricate patterns for Cozmo speech. I started thinking about things like: If we had a counter for the number of days that it's been operational, the number of times that it reacts to the corner of a table – So, let's say after five times he starts identifying that with a rudimentary word and maybe after 15 times it gets a little bit more refined. So, he develops these understandings of his surroundings seeing a face, seeing the corner of a table, seeing a dog / cat – he can start creating words around these depending on the number of times he experiences them. That was sort of the hope. And I think we got pretty far in figuring out how to do it and that's when, you know, we realized Vector is its own character we need to pivot.

Personalization and engaging long-term interactions are key concerns in the robotics area. How do you go from an initial “gimmick” interaction to something that's engaging over long periods of time? And this element of learning and changes over long-term time frames is a key part in that so it's really fascinating that these things are on the cutting room floor for Vector because they're so relevant, they're so important.

They very much are. We did a lot of that development for Cozmo and, of course, we lost that, because Vector had less of a vocal communication style. But, we started trying to think of ways to start applying that and I think given time we would have found a lot more

opportunity for that. I'm trying to think. So, yes, the challenge for Vector was it's this exact same form factor as Cozmo, so our audience is going to look at this and say "Why is this different?" "How is this different?" So, we had to create an entirely new character and for Vector, going with the same childish human voice would not appeal to young adults. So, we tried to go with something a lot more synthetic. I think a lot of the servo sounds – they're affected but it's all organic source. I think a lot of the sounds that Vector makes are recorded from a shoe shine machine.

Another really important finding with Vector is – We started exploring a lot more procedural work. We started borrowing from the different motors so we can draw in all this data, so that we're affecting the audio system with real-world procedural information. But I found – because at that point I was working with another sound designer as well and he was doing a lot more of the content design – but I was finding that you can have all the data in the world informing what the audio system is doing, but it just turns into this – I think it comes down to this fundamental conflict between demonstrating how cool a robot is and all the interesting stuff it can do, and actually having compelling character. Having an audio system that's responding to speed, mood, temperature, time of day – all that stuff doesn't really create interesting interactions. It's accurate, yeah, but it was making bespoke sounds and bespoke animations and collaborating with the animators to pull that off that made it work. So, we developed a fully procedural system. Every animation, the lift going up, going down, head movements, velocities, every wheel informed what the audio system would sound like, but it just sounded boring.

So you didn't use it in the end?

No, all of it still existed in the audio system. It's just disabled. One button disabled the whole thing.

Okay, and what were you hoping to achieve when you set it up? What was the aim in making it procedural?

The honest answer is – and I'm aware that you're a lot more familiar with interactive sound – it feels cool if you can accomplish something that's highly interactive and you can say “Look what I did!”

Sounds good in the pitch as well!

Yeah, it's like “Look! All this information changes this. (laughs) It sounds really cool and you can present it to people and say that all this really interesting reactive sound is happening. But ultimately, when you saw every animation that would play, that was generating procedural audio or a procedural playback of audio content, it was just boring. You couldn't tell that he was happy, you couldn't tell that he was sad, scared, none of the intention existed. It was just “sound was playing.”

Would you say it was too nuanced? I mean, on the one hand you want something to be responsive and reactive, to feel like it's reacting to its environment, but on the other hand, pooling all that stuff from motion data is too much or not necessary?

I don't think it's specifically responsive. There will come a time when procedural or generative audio is going to be a thing that works, but you're going to have to have so many parameters, moods... What is the robot's mood the moment one thing happens as it's compounded by another experience that you might have in the past? You'd have to build a lot of data to inform the audio system and I think we just didn't have the time. I don't think we had enough input, so having the animation team go and create bespoke animations for all of those moments and for me to go in and design sounds and hand-author them to animations is the only way that you would get a response from the robot that actually demonstrated what he was feeling.

So, if you would have used a procedural system and just pumped the animation into that system you would have automated the process that you did by handcrafting in the end, right?

That's sort of what we had. For example, when we were in the early phases of trying to just develop, getting functionality on Vector, we borrowed animations from Cozmo. So, Vector was playing lots of animations from Cozmo, but we didn't have an audio project for Vector yet. We were still designing the sounds, so what we did do was borrow all of the mechanical data and it would play sound. If an animation ran but we had not authored audio into it, it would just play from the procedural system. And it worked. It all worked. It was just really noisy and clunky and, you know, when you would respond scared to something, it was not clear in the sound. Fast-moving motors informing an audio system don't necessarily sound scared.

Fair enough! So, when you have this character and you create a bunch of sounds for it, how would you categorize those audio assets? Are there different categories and groups of sounds that fulfil different roles?

Absolutely, and that's something that came into play a little bit later on. It was probably two days before Vector was on shelves, two days before launch we had an executive review and, for a long time, we've been really trying to pull back the character because Cozmo was successful for being really outgoing and emotional and we wanted to appeal to an older audience. Then we had this review, they just basically looked at it and said "Where's the character?" "Where's the personality?" So, I had to invent sort of like a new layer... Cozmo's main sound source is his voice and everybody knows what that sounds like. And then there are the nuances that are supportive of the character. For Vector, we broke out what we were calling his voice into two subsets of sounds which we called *head* and *lift*. They were originally designed so that if the head moved a certain way it would play a style of sound and if the lift moved it could play another style of sound. The treads had their own flavor of sound as well, so that's three entire groups. All these have their own mood banks, Happy,

Angry, Sad, designed with the same audio sources. On screen it was the same thing and then I was finding the sounds that we had implemented still didn't have enough personality and that's when we had this last minute need to create character and I came up with another layer. What I thought about for Vector was: Screen was just simple facial expressions. The head and lift sets of sounds were just common whistle sounds almost like the R2D2 stuff, but there's another layer to Vector that's a lot softer, and it's anytime he expresses emotion. If he's really happy, he plays this more synthesized, tremolated pattern. And, I started doing a lot more of that. If he was sad, he played that style of sound. If he's happy – and those get used in pointed interactions – when he sees people's face and gets excited he plays that sound. If he recognizes your hand and he crawls up on it he plays this simple band of sound which doesn't get used anywhere other than impactful moments with people.

So, do I understand correctly that the core role of all these sounds is to effectively create a character and to communicate personality?

Absolutely, yeah.

And would you describe all of these sounds as the robot's "voice"?

Yes. Actually, I think the simplest description – I use this all the time with Cozmo because it's easier – is: You have the voice. It's like speaking. If you communicate with somebody you can speak to them. Screen sounds communicate facial expression. Lift sounds communicate body language. So, all of those were shaped into emotive intentions and they can be used interchangeably so you don't need a Voice to express sadness. If the eyes just kind of dip down, you have "sad eye" sounds. That's it.

For Vector, it got a little more complicated. This the grouping of what I called head and lift sounds. Those are more like futuristic mechanical sounds. Those got used more for reactionary things, like responding to the edge of the table. When he recoils you hear a sort

of a servo sound. With the treads, that is a procedural system, so the faster he drives, the louder and higher-pitched it is. That was just to be “cool.” Facial expressions; those are still the same thing but this additional emotional layer was any time that he's feeling... It's like thinking about Vector as a pet. A pet can vocalize and bark and growl and do whatever, but when they are hurt or when they do a little whine or do something a little bit more emotional; I felt like those are two separate experiences. So, for Vector we isolated emotional sounds versus just critter sounds.

Humans don't use sounds to support or to coincide with their facial expressions. Why does Cozmo have that? What's the reason for those facial expression sounds?

The original idea to do that was because if every animation played a vocal sound it would be really annoying. It would get obnoxious over time so it gave us really subtle opportunities to communicate just a little bit. You don't have to be looking at Cozmo but if you are, we filled out that experience so just these little subtle opportunities to communicate were available to us.

So, does it enhance or illustrate the facial movement?

The goal was to enhance.

I guess that's common in film or in games? If you have the possibility to sync sound to something, you just do it?

Yeah, the closest example I can think of is, if you have a footstep. If you see somebody walking through tall grass you hear the footsteps. But if you also hear the grass against their legs, it just brings this enhanced reality.

Right. Okay, so imagine all these sounds. On the one hand you see all those go together to create a character. How do you make the whole design coherent? To look at the object

and say, all the sounds in this object are sonically coherent and “one thing.” Or, is that even important?

The is one of the greatest challenges. If we did this all over again, I would invest so much more time in identifying the character first, because we had to create fiction around things “after the fact” a lot. So, it's hard to describe why we made decisions we did on Vector early on.

Like I said, we have these two groups we call head and lift but they're used interchangeably. They are not exclusively on one component or the other. They just became sort of like the global voice. There's also the forcing function of getting as far away from Cozmo as possible. I do think spending time to really think about “What is the fiction?” “What is the purpose of the robot?”... Because, each robot's purpose is entirely different. For example, we were working on the third robot that was not going to be entertainment and it needed far fewer sounds with entirely different intentions.

For Vector, as more of a pet, really thinking about what would a small foreign creature do in your home and what would it communicate to you... So, if it's hiding between books or playing with a coffee mug on the table, what would they be communicating? Identifying what their goal is, what their personality is, and then shaping the sounds around that... I wish we had more time to do that.

For example – I'm trying to remember as much as possible – when I was introduced to Cozmo, the audio director at the time just kept saying “Try to think of this robot as a fine swiss watch” and just imagining – we know that the hardware is just little chunks of plastic – “What do you think people are going to think is under the shell?” You get to invent that, right? So, you can make either robot seem as complicated as you want. It's just playing out of speakers so it's enhancing whatever you want users to believe. So, with Cozmo, once we determined that it was more of a toddler, going with really high tech was overselling. For

Vector, we tried to get further that way. We tried to make slightly more polished sounds, a little bit closer to something interesting.

Right, so designing sound in a way that gives you an idea of “what's inside the robot” was part of the brief?

That was some of it, yeah. Or, for example, the fiction that I tried to think about with Vector was “moving small motors to communicate” So, the faster they move, that modulates pitch, amplitude... Mechanical things are what's propelling his voice and nothing else.

Yeah, like a vocal tract out of gears?

That's exactly the sort of thinking that I had.

It sounds like these metaphors sound really powerful, really helpful in shaping your thinking around what you're trying to achieve with the sound design. So cool!

I mean it's an opportunity. Not to belabour the point, but it's an opportunity. If you're playing sound out of a speaker users don't necessarily know what's below the shell so it's your opportunity to be creative with that.

What areas did you take inspiration from for the sound designs?

I think Wall-e is the best example of robot design I've ever seen or heard. He just created so many characters – just the variety of character that he was able to build out of that, but really realistically! That was the starting point and then we arrived at all of our conclusions internally. It was just internal reviews and having character directors, animators, and creatives with really high standards. It was hard to make them happy! The greatest feedback you can get is failure, I think, because it's just going to force you to work harder.

A few things about the technical part: Obviously, you had technical constraints and considerations around the whole product and you probably also had a particular workflow which is unique. That's a new thing. It's not an established part of the sound designer's toolkit, so I'd be curious: What were some technical requirements or challenges that you faced when implementing the audio?

Probably the greatest technical challenge that we had, is all audio, animation, all content that existed for Cozmo was on the app, so everything that played out of the speaker for Cozmo needed to be streamed in real time. So, file size constraints became a problem. Of course the app can be any size. We were fine with that, but being able to port things without latency was an issue, so, of course we were arguing for higher quality. People more on the exec side, they just wanted something to work. They wanted zero latency, so we actually found an interesting middle ground where we were streaming audio to the robot. They wanted to get down to 8-bit, but eight bits sounded terrible. So, we went with 8-bit μ -law, which somehow split the difference. It's a perfect in-between line. I would have to look this up again myself... μ is a Greek letter and I had engineers explain it to me but I found a way to batch convert audio so that... or no. They did the conversion in real time. I will have to get back to you on that. That was a complicated conversation but that was the biggest technical consideration; Getting audio onto the robot fast, but, ideally, people in robotics won't have to deal with that much longer.

In Vector it was a different thing, right?

Completely different. Anki invested in a high-powered audio engine, Wwise. Wwise can do anything I've ever wanted, so we can compress really small Ogg files, we can respond to any sort of stimulus, external simulation, and use that for driving parameters. All of the organization happening inside of Wwise was easy.

Right, I remember you saying that in your presentation. Wwise was embedded in the robot.

Correct.

So, how did you prototype? What was the workflow or the setup for the prototyping?

Prototyping generally happened to video to get buy-in on a feature. If we were pitching a feature they would basically storyboard the whole thing, I'd put in some simple audio, we'd green light it, the animators would go back and figure out how many animations to make. And, for both Vector and Cozmo, the foundation of the audio was to build as much generic "middle road audio" as possible so that we could have a modular audio system where, if an animation comes online, I can just drop all of our existing audio on the timeline and anything that needs to be created custom can be done a little bit later just to enhance the animations.

So, you have a basic set of movements like facial movements and gestures, and you create a basic set of sounds to accompany that movement and those contain the character and the personality and that's your basic building blocks?

Absolutely. Actually, all of the basic building blocks are built "in mood," so, for example, *happy* has the head group of sounds, has *happy long*, *happy medium*, *happy short*, *happy micro*, and all of those have five to ten variations so that every time a new animation comes online, I can find these moments in the animation. If he's doing a really quick animation, but he seems happy, I can drop a *happy short*. If he does a big turnaround, I can drop a *happy long*, so it's the same emotional intent. Having a modular system like that just allowed me to keep up.

Yeah that's fast, but does mean that the entire vocabulary of the robot is movement sonification, that it's always sounds accompanied with movements?

Sometimes... I feel like I've made different decisions depending on the day. Sometimes he might be curious and just his eyes would move up to the corner, but it might be a profound enough moment that we might want to put in a physical movement but because so much of his sound is servo based I did rely mostly on physical movements.

What about those special purring moments?

That's when those came in most useful: when he's trying to communicate something and he doesn't necessarily need to move. I think a good example is, we had a feature where they did hand detection. If he notices your hand on the table he rolls up, puts his lift on it, picks himself up and he's not moving anymore. Then we did some purring. We did very emotional calm sounds, just to communicate that he's happy being with you.

That's fascinating. So, for these other sounds the first sketch would be animations, and audio would be adjusted to those animations. So, the general contour of the sound would be the animations, but the content of the sound is the whole character / personality thing?

Yes, and, as we continue to generate sounds and all these specific moments – Cozmo especially had a lot of unique moments. Once they exist in the Wwise project they're accessible so if we do this over-the-top sound for a reaction to something, *that* is now accessible to any other animation, so you organically develop a larger and larger personality palette that you can use for more and more things. I would find myself often going back to earlier animations and revising them with new content that we've developed.

Okay, and then the mapping from the behaviour of the robot to the Wwise engine; those mapping decisions what kind of sounds are caused by what actions; Those were decisions made by you?

That's where things started to get really interesting, yeah. For example, we developed a distilled variable of all the different moods. There are, I think, around 30 different things that could cause a shift in this one parameter we called *stimulation*, so 0 to 1 stimulation was Vectors "core engagement." So, if you say his name, that variable shoots up to 1 and it decays over time. If he gets tipped over on his edge it drops down all the way to 0.1. This moving variable within Wwise we were able to map the whole audio system to. For example, if his engagement drops below 0.8, none of those emotional sounds play. Their ability to play is mapped to how *stimulated* he is and as that continues to drop down it actually selects smaller and smaller versions of every sound in the system. If he's not being engaged with by a person and his *stimulation* drops really low, he can exist in the background without making a bunch of noise. That's one thing that we found with people. They might enjoy the engagement but as soon as they want to work or do anything else it needs to get quiet.

So, *stimulation*, that's like a big important parameter that you have. On the other hand I would assume that in Wwise you also had all the different motors and speeds and velocities that you talked about when talking about the procedural stuff. Hooking up all the expressions to a parameter like *stimulation*; would you also call that procedural or would you describe that differently?

I would describe it differently. It's adaptive audio.

So it's higher level? It's not as low level as "procedural?"

Exactly.

Okay, fascinating.

For example, I described the *happy* system as being *long, medium, short, micro*, but that whole system – and I mean this for every move that he can be in – that whole system is

copied three times and every time that parameter drops below a threshold, it jumps to the next subset. Then, within Wwise, you can get into the audio files of that subset and there are actually fades that start earlier and earlier. So, every sound gets shorter and the probability on the event gets lower.

And that's the whole game audio background?

All game audio stuff, yeah.

What were some other types of data that you were being provided with from the robot?

You know, it was going to get really exciting. There was some cool stuff coming along. We had the AI engineers expose a document to us. Every behaviour that the robot could enter into was exposed to us and so we could start hanging audio states on to those. If the robot felt scared and it existed in this AI system that was determined to be scared we could change everything about the audio. We could throw on a tremolo to make it seem like he's shivering. We could have done so many things, yeah. Unfortunately, we didn't have enough time to get creative. We were still working fast to figure out who Vector was as a character. Right before launch we determined we need more personality and then we invented all this emotional response.

Other parameters were all the mechanical stuff which was really only used for the treads. So, as he's driving around, the only motor data we're using is for the pitch and the volume depending on which way he's driving. He has a backup sound, he has a driving forward sound, that's it. I was working with engineers – I really wanted to hang counters on interactions like “edge of table,” “user's face,” “time of day.” We developed a feature around “how old are you?” and he would count up how many hours he's been alive and repeat excess speech. I think there were so many opportunities we just never got a chance to dive into.

You were mentioning emotive clarity as a key challenge in your earlier presentations. What were the things that you tried to achieve it?

That's where a lot of the audio content iteration came in. In early versions I was trying to design sounds for a robot that would impress people, that sounded really futuristic or cool. Then we hired a voice actor and just said "do something really funny," "do something goofy;" trying to find a mode of clarity. I was finding when I was creating groups of sounds in this blank matrix – because again: At this time, we didn't have a character. I didn't know about the concept of character. I didn't know about the concept of working with animators. I had never done any of that, so what I found was, I would go present my sounds on the robot to other people and, as we're talking about it, Cozmo rolls off into somewhere else. You couldn't tell what the mood was anymore. It was like, he's totally lost, he's just making sound. And once you don't have facial contacts, you don't have any of that, and you're just holding a conversation, I think the sound became sort of meaningless. It was just background noise.

So, by playing sounds for people out of context so they have no idea what the sound is, and trying to get them to identify what it sounds like – Once these sounds achieved a higher and higher accuracy and became appealing to people and the emotional clarity became true or consistent with people – That was when I started applying it to the robot and putting into animations.

So, basically you would create a sound with your expectations of what kind of feeling is in there, and bring that to a bunch of people and check with them if they see that too?

Yes.

...so that you get some sort of feedback that you then also worked into the next sound. I guess lots of that is based on intuition, a learning process for you?

Yeah – and I touched on this a little bit in the presentation – I think it's really hard for me to imagine as a sound designer with a background working on sci-fi stuff, but I also just got thrown into being a voice-over director. I had to direct actors and say the right thing to them to get them to give me the right performance and be like “That's it! That's what we need do more of.” I had to apply a lot of that thinking towards my own sound design. The way of making sounds for Cozmo was just making sounds like “that sounds happy,” making a bunch of those, and then choosing the ones that are the clearest *happy* sounds, and then sharing them out with the team and asking “Does this feel good? Does this feel right?”

There's a bunch of research looking into exactly that. How do we get consistent emotional communication? Have you looked for, or used any research or guidelines? If you would have to formulate this work as recommendations for specific parameters, would you know how to formulate those?

Absolutely not. I have no idea most of the time how I'm creating stuff. I don't go into a design process knowing what I want the outcome to be. I think from just professional experience I've been wrong enough times that I've developed a sense of what turns out to be right, so when I design sounds it's purely off instinct. I don't know where to start. You know, on the on the other end, I feel it like it's absolutely fascinating that people are able to identify and quantify and understand how successful things are; why they're working. I have no idea. I'm purely based on my own instincts and how people respond to my work.

You touched on this a little bit in the beginning but... imagine you had unlimited time and unlimited budget to create sound for a new robot. What are the things that you would do differently?

I think it would depend entirely on the type of robot. I talked a little about this with somebody else yesterday, about what's the creative approach to Cozmo, Vector and – there was another robot that was in development that's a much larger robot. It had a different potential purpose. And so, I think for each robot, if I had the time and budget – I think using

Vector is a good example. Vector is a robot that I would love more than anything to do all the way over. I would love to start over. I would really try to, again, think about what I want the fiction to be and understanding. I would work closely with a character designer, somebody who can give me a foundation to build on and give me feedback; but really think about how you want people to feel. I mean, otherwise, what's the purpose? If you have a utility robot that's, you know, delivering or doing some specific purpose, that's something else entirely, but when you want people to feel – For Vector, you want people to feel like there's this, you know, intelligent rodent, but it's highly sophisticated. So, I would again think about what are the different ways that this highly intelligent sentient robot character can communicate and what I wanted to communicate. I would probably come back to the same system of its own internal mechanisms communicating sort of like chattering to itself like a squirrel or something, but then having that emotional band of audio. I'm certain I would arrive at the same conclusion.

So you would try out more conceptual approaches to the whole concept around the design?

Yeah, if I were to do Vector over, there would not be a difference between head and lift. I would find what is the prominent communication method – um.. I mean even talking about it I feel like I changed my mind, because the prominent communication is sort of an internalized sound. It's like his servos are internalized. He's not necessarily communicating with those sounds. The emotion is what he's projecting.

Right. A sign of something that just bursts out.

You know, we just started developing this other, larger robot and other considerations are cohabitants. If you do have a robot that's larger in the home, whose purpose is to just exist with you and maybe companionship, not having as many active sounds is much more important. So, the sound design that we were working on was warm, polite and reserved, and we were treating this new character more like a German shepherd; like a larger more

intelligent creature that, instead of being startled by objects, would acknowledge them and avoid them. So, the sound design approach was much more smooth and selective.

That makes sense. I've got one final question which is around the Internet of Things. If you imagine for a moment that you create sound for a robot, but that robot has access to every loudspeaker device in the environment, to the tv, to the smart speakers, maybe AR glasses – an expanded loudspeaker access powered by the IoT. What would you do with that? What possibilities do you think that could open up for the audio profile of the robot?

What would the robot's audio profile in relationship to these other audio sources be?

Yes.

I would say, without a doubt, there's no way I would allow Vector sounds to play out of any other devices, because as soon as you do that you ruin the fiction of this being a truly independent creature. I think that breaks that fiction. I think having the ability to send sounds... personally wouldn't do that. If the robot is intelligent enough to send music to start playing on a device, it again evokes the understanding that this robot is a lot more intelligent than it's letting up to be. I mean, these are conversations we had early on with Vector. "We do potentially have the ability to play music or sounds from Vector through other devices" and it was a hard. Yeah, Vector stops being Vector as soon as you're just spreading it around to other systems.

Fair enough. So that runs counter to the character; to having a coherent, contained character that you want to communicate and not break?

Right. And, I think what we did with Cozmo is probably what I would continue to do, where there's an understood and parallel music system that's operating off of the same behaviour system that Cozmo is; so that Cozmo can more realistically respond to it and react to

something and understand it that way. For example, what I would probably do with technology like that is, if you had Vector start a playlist on a Bluetooth device: if Vector's aware of that interaction happening I would have him respond *to* the music coming out of the speaker as opposed to him, you know, *sharing* any of his content through it.

I actually do have a Vector but I don't have Cozmo. Do you also have a sound set for the app? Is there a soundtrack?

Oh yeah. There's a fully interactive score. A huge score, so that all of his behaviours, moods, games, every mode of game you enter into there are bespoke transitions. We've recorded a live orchestra. Yeah, it's highly interactive.

Okay, interesting! Do you see a benefit of having that available for any other robot that is not hooked up to an app?

I think it works for Cozmo because it's aged downward. It's definitely for a younger audience so the music itself is very playful. It's very cartoony. For other robots: I mean Cozmo is such a one-to-one entertainment value robot where that makes sense. If you're dedicating a phone to play that music system it's not going to appeal to a lot of people. I don't think Vector should have a music system or any of the other robots we talked about. Absolutely not. I think it was unique to young children.

Right. Okay, fair enough.

With time, I'm sure I could think of a lot of crazy things to do with it but immediately: no. I'd say no. Some of these things we started getting better at. Some of the R&D stuff we were doing was really cool towards the end. Even beat detection – If we would have had improved beat detection, rather than projecting sound onto a speaker Vector could respond to it, you know, catching the beat and being able to nod his head to it; responding to the world with you and not creating it for you.

For Vector, – this might help – one of the really important distinctions that we have – because you can use utility functions like the alarm, the weather, certain conditional things like New Year's Eve – In all of those cases, like if the weather appears on his face, he doesn't show you the weather with his own abilities. The weather shows up and he responds to it. He's still Vector inside of this utility. We very much kept him independent of this external utility.

And that again makes him more of a coherent character, right? It's like: showing the weather is not part of this little personality?

Exactly.

That makes perfect sense.

And it gave some sort of comedy. It gave us something to work around creatively.

Yeah.. Ben, that's all I got. Thank you!

Jeshua Whitaker – Jibo

Could you briefly describe Jibo's sound?

Jibo's sound consisted of three main pillars: text-to-speech (TTS), robot speak (jiboease), and user interface (UI) sounds. The TTS was used to inform users, but if Jibo felt a particular emotion in his conversation with you, we would play robot speech at the end of a phrase to add more character. The UI sound design was just a subtle way to communicate with users when interfacing with Jibo via his menus. Something we've become accustomed to with our appliances or smartphones.

What was the main role of Jibo's sound design? What main goal did you have when you started the design process?

The main goal of Jibo's sound design was to create a character that users could connect with. Developing his native "jiboease" was core to the Jibo experience.

Does Jibo communicate a personality through sound? If so, how?

Yes, his TTS engine allowed the narrative designers to play with prosody, pitch, and some other effects on a line by line basis, but we still felt like this was still limited so for character building purposes I developed the robot speech to communicate more colorful emotions such as sadness or excitement.

How would you describe Jibo's Character? What characteristics or personality traits was Jibo meant to convey through its sound?

Jibo's character was joyful, smart with a hint of sarcasm, and some dry humor to round it out. His emotive sound effects or vocalizations were intended to accent all of his personality traits, but we leaned on success (happiness) and failure (sadness) the most since they translated the best to users. In non-verbal vocalizations it's a bit too hard to convey sarcasm for instance, but I'd try to add some sassy sound effects when I can.

Where did you take inspiration from for your concepts and designs (Other physical robots, Film, Games)?

Our early concepts mostly derived from film and Cynthia's past robot designs. We worked with Ben Burtt briefly to talk about sound design concepts and strategy. R2D2 and to some extent Wall-E were probably our biggest influences.

What categories of sound does Jibo emit (voice, alerts, etc)? Do these different types of sound have individual functions?

Text-to-speech, UI (alarms, sliders, buttons, boot up, etc.), and robot speech. Yes, each pillar had its role. TTS for communication, UI for notifications, and robot speech for emotion.

Does Jibo communicate emotion through sound? if so, how?

Yes, the robot speech was specifically designed for this.

How do you make all these elements fit together? Was coherence part of the design goals?

This was the toughest part of the process and if Jibo was successful today it would have gotten a lot better by now, but that was something we worked on tirelessly through UX stress testing. Communication is a very fluid action that humans are experts in. We are masters of speaking and listening, and Jibo was not. There may be times when you would say something and the query time to deliver a line was too slow, or maybe there was a delay on delivering the robot speech line and the intended communication did not deliver the expected result. With time it would have gotten a lot better, I think.

Did you take into account the visual design (size, movement style, materials, shape) when choosing timbre of TTS, robot speech, and UX sounds? If so, how?

Yes, I would say the TTS voice was the root of his personality and influenced his physical design. The robot speech and UI sound effects tended to be reactionary designs based on visuals. I would see how a physical robot animation or screen element would occur and try to design sounds that fit the moment.

How much is the Jibo's audio output affected by realtime parameters? Is the audio mainly sample-based or are there procedural elements? Does Jibo's sound change over time?

All of my sound design was sample based. We used metadata tagging in a custom engine that would allow narrative designers to pick when Jibo would play a certain type of robot speech bucket and that bucket of sounds had a few options to randomly choose from. Additionally, we had some idle chatter not tied to narrative design that played sounds based on his current mood, mostly affected by if you had used him recently or if he had success looking something up for you and wanted to make a cute sound to celebrate.

Can you tell me about the workflow / design process during production? Your process during content creation, evaluation, and testing?

This was quite a few years ago so I am going to try and do my best to remember. From what I can remember, the first six months of the design process was all to picture. We didn't have the robot and his engine working properly just yet so myself, Dave Schlafman, and the rest of the design crew made linear mockups to work on creative vision in the meantime. In these videos I spent countless hours trying different types of robot speech designs and discussing the best time to trigger sound effects.

Once we settled on a palette of sounds, we would do user surveys to see if they would communicate certain emotions to the picture or just via the sounds themselves. We found through this testing that non-verbal communication (such as R2-D2 for reference) could not communicate emotion on its own about 60% of the time. It required animation and visual components to fully deliver its intended purpose. Even then, since beeps and boops is not an established language, I don't believe we ever got over 70% comprehension even with video. This meant we had to keep his emotional buckets pretty simple (sad / happy)

After this phase, we were finally working on the robot itself. This allowed us to really stress test the designs and create some unique functionality like Jibo purring when you pet his head. This is where we developed the emotion tagging system and his idle chatter.

Did you revise Jibo's sound palette in response to the user surveys?

Yes, a lot actually! Getting user survey's was a pivotal moment in my design that helped simplify how we use the robot speech sounds and gave more focus to what worked and didn't in my designs. I would say it helped me be more deliberate in my approach to the robot speech sound set.

Did Jibo's loudspeaker characteristics influence your sound design choices? Or alternatively, was loudspeaker choice influenced by planned content?

The loudspeaker was a limitation that I had to work with. Jibo had two speakers in his head on the left and right sides, similar to your ears. Since these weren't forward facing speakers you wouldn't receive direct sound from him, but rather an echo from a nearby object or wall which could throw some users off since we tend to locate objects based on where our ears perceive the first sound. With that said, Jibo did animate its head pretty frequently so sometimes you would get a direct speaker facing the user.

The speakers had to be very small in Jibo to fit in its head so the available frequency range was limited. For instance you would avoid frequencies below 170 hz because you couldn't hear them much and there was some fear that lower sub frequencies could eventually cause damage to the electronics.

Another thing to note, Jibo's head acted as a resonating chamber similar to an acoustic guitar body so I had to find EQ settings on all my sounds so that they didn't over accentuate frequencies that were boosted by the acoustics of his head and body. The EQ was not runtime in Jibo's software, this was settings I would make in my audio software.

In the future, I would recommend using a mono forward facing speaker that acts as a mouth and plays TTS and robot speech while the side speakers could be used for music or UI sounds.

If you had unlimited time and budget for creating sound for a new robot. What would you do differently? What new things would you do or try?

The procedural audio landscape has changed a lot since we worked on Jibo all the way back in 2016 and computing power has only gotten better.

From the start I would outline the use case of the robot and see where audio and communication slots in. If we were to do Jibo all over again, I would probably do away with the robot speech altogether and rely on a more robust and emotional TTS engine that can be controlled in real-time. Voice cloning tech is starting to pop up in the gaming industry and I would urge the team into using such tech to create a more emotional English TTS voice that delivers the intended emotion without relying on an alien language (and hopefully localized to other languages down the line).

Additionally, having a smarter emotional system would have been key to making Jibo's AI "smart". If we could have monitored real-time parameters of Jibo's environment (microphone audio, lighting, face ID) along with the length of communications and idle time we could do a ton of fun TTS utterances with that information.

As for the robot speech, a non-Jibo robot like Cosmo that is meant more for entertainment benefited from non-verbal speech. It was funny and an enjoyable experience, so I think depending on the use case of the robot there are pros and cons for each design methodology and I would have to take that on a case by case basis.

Connor Moore – Kuri

Could you start off by describing to me what Kuri sounds like?

Kuri has a voice that is modeled after an african thumb piano called an mbira. We took that into synthesis and utilized pitch and envelope shifts to model that of human intonation. So, it actually translates to almost sound like whistles. The goal was not to sound like R2D2, but the actual end effect somewhat plays out to sound similar. But, it's all based off of the intonation of the human voice.

I love that. Why an mbira?

We did a ton of exploration around different instruments and found that the mbira, the attack and the sustain of the instrument, played nicely with those types of envelopes. You could kind of get a similar result with a longer sine wave but we also wanted that more organic quality and that imperfection that exists within natural instrumentation.

Okay, why that imperfection and that more organic quality?

You know, the main goal with that was the human voice, right? The human voice has imperfections to it. It has overtones to it. It has this obviously more organic quality. It's a more human quality, so we're trying to get it close to human without being too human. We explored the human voice and kind of on the tip of what Vector did and things like that and we really wanted to pull away from it sounding too human, because we also didn't want to over-promise. The robot was only capable of so much, so if you could sound too intelligent, it could almost over-promise and under-deliver. We wanted it to sound semi-human, have some imperfection, but also organic and not sound like it's smarter than it should be.

Did you try out other sound sources during the early prototyping?

Oh yeah. You know, that was the final execution. Throughout the entire process we did everything. I mentioned the human voice and just recording “ah” and then pitching those things up and down and tuning them differently and layering them differently. We tried creating more product UI-type sound, that was based off of melody and based off of major and minor scales, based on positives and negatives. That didn't seem quite rich enough. We wanted it to model something a bit more human and it didn't have quite the flexibility and range that we looked for. And then, everything from human voice to some basic sine waves to various types of synthesis to bansuri-type wind instruments and things like that.

Did you try out any more elaborate processing, because now you're mainly talking about pitch and intonation. Did you experiment with spectral effects?

We did. The speaker that was shipped on the device wasn't super advanced, so some of those spectral effects – you really couldn't get what you would hope to out of them, to be honest. But, we certainly played with chorus and flutter and things like that. When the robot would get frustrated, you could pipe in a sound signal and have some type of flutter in it, where it sounds like it's frustrated. But then again, the robot's personality was really supposed to be primarily positive and we didn't want to make it sound overly negative at any point.

More generally about the design; what was the main role of Kuri's sound design? What was the main goal you started out with when you began the design process?

Number one was expression. This robot was to live in people's homes, one of the first robots to live and walk the spaces in people's home. So the goal was to bring a positive feel and a welcomeness into that environment for something that was so foreign in this environment. Number two would be basic functionality, like understanding the positives versus the negatives and being able to express those to people and have them be understood in a quick and easy way. Then, number three was to have somewhat of a brand expression that related back to Mayfield, because they were planning on actually doing

more of this in the future. So, the goal was to create something that would carry on and build a through line that could carry on potentially through other products as well.

Was there a character and a more detailed personality established for the robot?

Yeah, the basic brief on Kuri's personality was cheerful and bright, which plays back to the instrumentation. The robot's communication being bright, something that's a joy to have around in your home and something that brings delight to your experience, were the key highlighted things within the brief.

Where did you take inspiration for the concepts and designs?

We studied really everything across the history of robots, like star wars, all the robots in historical film – Also, Vector was out at the time, so we researched that. Jibo was not out yet, but they had been pushing some of that content out as well. So we really studied all of what was out there, even the more humanoid robots, like Nao. I had a bunch of research stacked up to kind of see how these robots were communicating in different ways because it's very different from a lot of the design work that i do. I do a lot of product-based and brand-based work, so it was a very different space for me. It was important to dig into that kind of understanding up front.

So your background is product sound design and sound branding?

Audio user experience design, sound branding; also, under the brand portfolio it extends outside of product and into music for brands and brand strategy. So, it's a big playing field if you will, and audio user experience design is one of those elements.

You were talking about all the sounds that the robot emits to communicate personality and emotions. Were there any other types of sound that the robot would emit?

Yeah, they had some interesting functions. You could pet the robot and it would purr. We had what were called Romojis, like audio emojis. Kuri could whistle something like a song, and there was sometimes less communication-based sound and there was a language built out of play as well. During those moments we would actually incorporate some foley stuff, too, where it incorporated the language elements of the mbira with envelopes, pitching, etc, but it also included some some type of foley work.

Can you give me an example of that foley work?

There were elements where whenever you'd pet it, we would take the mbira and throw it into this chorus flux effect, where it would sound like a purr, but then I also had like a little heartbeat thump underneath. I mean very subtle foley stuff, where in the mix the foley would be way back; just a very subtle touch of that. Then, in terms of alerts, the robot would be stuck somewhere – those types of things were often verbal. There was like a bump sound whenever Kuri would bump into something or was lost, and it would go “whoa.” It would essentially be a really sharp pitch contour, from a low frequency to a high frequency and down, in the shape of a triangle. But, there were also basic alerts. There was a waking alarm that we kind of had a different language for: the more product side. That was more UI-based so there were ringtones that would come in – but again, they were really following following elements of the language that was based in the organic; trying to somewhat humanize the robot or at least make it familiar.

Right, so were you utilizing this similar type of organic instrumentation in the alert sounds or was that a separate UX sound language? Or did those blend together?

They sometimes blended together. It depended on the instance, but it was primarily driven by the more verbal language.

It sounds like the mbira was a core source material that you weaved into the different types of sounds that Kuri would make. What's the motivation behind doing that?

Consistency. From a brand angle, consistency is super important and having there be a consistent thread is really important. That was one of our primary goals with that product. I totally see the angle though –more from a function standpoint of differentiating, but for this we felt like we got enough variety and enough differentiation out of the voice. That it was still effective so it was kind of checking both boxes. it was like “can Kuri maintain that consistency and be functional at the same time?” And, from the user testing we did, it was pretty successful.

Can you tell me about the user testing?

That was actually done internally at Mayfield. They had a group that they worked with and they gave me feedback. That was part of the iterative phase.

And those were members of the general public?

Both. Team members that weren't actively involved in the sound design – I was working with a very small team, but the company may have had like 50 people, so there's constant internal testing happening. But then, they also had an external group of people that they were testing with.

Right. You were talking about communicating emotions. Checking if those emotions are clearly communicated was also part of the testing, I would assume?

Exactly.

How was that feedback? Was it easy to communicate these emotions? Did that come across to people?

Yeah, i think so. I mean, Kuri had a visual component in its chest that would light up in different colors and in different ways. Her eyes would also change based on mood, so there were multiple points of communication and whenever you're going multimodal between visual – whether it be eyes or lights –and sound there's three points of a communicative element. That was pretty effective and with that type of thing, utilizing more major versus minor – It's pretty straightforward and pretty simple in terms of people understanding. Even non-musical people tend to really understand how those those things work with emotion.

So you used those sounds in combination with these other modalities?

Yeah.

How many emotions did you try to separate?

We kept it pretty simple. It was mainly *happy*, *sad*, more *neutral* and then very rarely it was *frustration*. That was a very rarely used thing.

How much is Kuri's sound output affected by real-time parameters? Is the audio mainly sample-based or are there some procedural elements in there as well?

In terms of reaction?

If you have Kuri somewhere and it interacts with the world, what kind of samples get triggered in which way? Or, is there real-time control, are there some internal parameters that determine which kind of sounds are played?

I didn't work with the implementation side of things. I was just on the design. I was an external partner with them and just mainly focused on brand and design. I don't know how they did it programmatically, but it was very immediate. There was no lag between an interaction. It was really pretty immediate.

Okay, and if you would create something like a *happy* sound, like an affirmative happy “Yes” from Kuri for some sort of interaction, would you make one sound or would you make a set of sounds?

We did multiple iterations and variations on it. Like, *super happy, happy neutral, happy...* For the really happy versions we did shorter, faster melody and really sharp pitches that went higher up in pitch. And then, it just went down the scale from that, so less sharp pitch shifts, longer, more lethargic – we use time and melody really to drive the level of excitement, if you will.

Okay, and if Kuri would have three times in a row the same *very happy* expression, would it be the same sound file?

If it was the same trigger, if it was the same *very happy*, yeah – because, you know, some designers use an engine that can spit out – I think vector had something like that – they had like 100 versions of one thing. For Kuri, it was a really rather simple language, which was much less varied. We had variation, but it was definitely much less than Vector, for sure. I don't know about Jibo.

I know that Vector's designer has a background in game audio, so the *footstep approach* is probably where that comes from.

And that's one thing that we were wanting to do before they shut down: add more variation. But then Mayfield closed and shut Kuri down, so we got halted. At the same time, the language that we developed was so simple by design, that I think once you throw so much variation in there – I was always concerned about it becoming confusing, you know? To me it was like “okay, having variation for variation's sake didn't really...” I mean, I could understand how a user would get annoyed by hearing the same thing over and over again, but when a design language is that simple, just throwing variation into the mix could

actually over-complicate it. That was a point of concern and, to me, when they wanted to go deep on variation, I was excited by it but I was also saying “we should have considered this way back at step one” because that, to me, influences entirely the design decision making in the very beginning.

Talking about the design process, can you give me an overview of the different design phases from the very beginning, from the brief, to the implementation?

Sure. Up front we do *Discovery and Strategy*. That is where we get all creative briefs, brand briefs from client. Sometimes – and I don't think we did this with Mayfield – but oftentimes with clients we'll do stakeholder interviews. We'll do music and sound workshops where we're listening to snippets of sound or music and get a better idea of what might be on or off brand, or right for a product or a brand. And then, we come back and we showcase our learnings to a client and say “Hey, here's what we've come to understand about this project. Do we have the green light to move into creation?” Then, we go into the actual *Creation* of these, select about a handful to two hands full of assets and design those typically in multiple directions. So, at the brand level, what we do is: We get an understanding of the brand characteristics and say “Okay, this group is all about *delightful*, this group is all about *human*,” this group is all about, you know, this different brand characteristic, showcase those to the client and they pick one direction or a combination of them. Then it becomes *iterative*, where we essentially take feedback from client, iterate on it, pass back, and that's a continued cycle for whatever the scope calls for. Then, at that point – depending on what it is – user testing will start. And then it's all about *optimization*. So, depending on what the playback environment is, what type of speaker it is, what type of cavity does it live in, we optimize directly to that. That's the majority of our process. They're typically working on implementation as design is happening and then seeing how things work in the robot based off that. After our final optimization, that's when final implementation happens.

Right, so was hearing the sound coming out of the physical robot also part of your iterative process? Could you listen to that and then go back?

Yeah, that was part of the entire process, and we did that really early, because we did that with a bunch of different directions just to see what felt right. I mean, part of it was like, “acoustic bass – oh we can't create anything below 500Hz.” But also, a lot of it was just based on feel, like “does this feel right.” That was done at the very early stage.

And that's an important element you would say?

Yeah, absolutely. To me, it's one of them, probably the most important element to make sure that we understand what we can do in design with the given product. Because, I mean, we always get acoustic specifications, of course, but those sometimes aren't – I'm sure they're very accurate but they don't always play that well, because they don't take into consideration the resonance, the cavity you could live in. Always, with product, the earlier the better on that front.

Was it easy to get new sounds that you would design right into the robot to hear them in context?

Again, they were doing the implementation. I did have a robot – now that i think about it – and I was able to push sounds onto the robot as well for a period of time. You know it is often the case when working as an external partner – often with a startup like Kuri or many companies that I work with, they don't have a wealth of products to be like “Oh you can have this for three months.” So, at times because they're working on all their stuff, I did have access to one for like a month and they just gave me some code to run through terminal to load files. From what I recall it was very easy.

Great. How did – if at all – the design for Kuri differ from more conventional product sound design?

Very different – much more playful, i would say, with those pitch bends modeled after the human intonation. Sounds were longer. Sounds were more – I wouldn't say more melodic because some product sounds are very melodic, but typically product sounds are more based on a single note, melody or single hits, rather than this being a consistent tone that pitch bends melodies. So it's really very different. I've never heard a product like this and I actually would say no product should actually sound like this. I have a very opinionated point of view on product sound and I think it needs to be extremely simple. Kuri's language is simple, but the way that it played out into the world is very playful and I think products generally need to be pretty reserved and almost sound more sophisticated. Obviously, it's dependent upon the brand, but i think Kuri was much more playful than a lot of products should be.

If you had unlimited time and budget for creating sound for a new robot, are there things that you would do differently or any things that you would try?

Unlimited time and unlimited budget? Hmm, what kind of speaker am i working with?

Whatever you like.

To me, the challenge with product is that typically the speakers can't go much below something like 500 Hz. They usually have small, cheap speakers, so if I had a great speaker, I would explore much more the lower pitch range. I tend to think that lower pitch ranges are softer. They just tend to be warmer and more inviting and welcoming, so I would definitely – A lot of my design work is based around what I call *considerate* or *polite design*, and about staying out of the way and being effective, but very respectful to our environments. And, I think the way we can do that is to utilize more of that low end, nothing super low end, but in the 200 Hz to 500 Hz range, that has softer attacks and is just generally more polite and more soothing. It wouldn't even really be about budget or time. It has more to do with the bandwidth of a speaker.

Yeah, so imagining like a human-sized robot on a small base with wheels, with all the space that you need to have good full-range speaker capabilities...

I think I'd also like exploring the more human side of the voice. Again, it depends on the persona of the robot. I think there's a lot of interesting capability to explore in our voice for sure. And, I don't know where I would take it, necessarily, but something that's really soft and peaceful. I'd probably explore that space and the low pitch range, something that's really warm, inviting and welcoming.

Did Kuri have a text-to-speech engine?

No.

So everything you designed was the voice of the robot?

Correct.

Talking about product sound design, it's quite interesting for me – I've now seen people from video game audio go into that space with a focus on creating a believable object and recreating an immersive reality on a robot. And now, I hear from you talking about *polite design* and clean, simple, unintrusive sound design, which is a different angle. In your opinion, how should robot sound be designed?

I think they could strike a balance, because I do think there's a place for more immersive sound, absolutely. But, I think it can be done in a way that's really tasteful and really clean still. I think there's a good balance, because the work with Kuri was very simple by design and based on not over-promising in that organic quality of it. But, I think you can take something like that and pitch it down an octave and use a sine wave and put some reverb and spatialization on it and it's a totally different thing. It's more immersive, it becomes warmer and then just more grounded. So, I think that's a good balance because there's

certainly value in the game audio approach. Game audio typically is very immersive and I'm skeptical. I think there would be people that would be really excited about that, but I think that would be a small percentage of people. I think for the majority of people, if robots are really to come into our home, that it'd be cool for a week to have that and then after a while it would just be too much. So I come from a very different angle of more user-centric design. In the product and brand world where it's really about that minimal aspect – making it understandable, making it branded, and conscious of its environment – I think that can be timeless. I think that can exist in our spaces without taking up too much space. But, I do think there's a balance there. There's value on both sides. It just has to be highly considered and very thoughtful in its execution.

I see. Did you have enough time for the initial concept phase in the design process for Kuri?

I mean, you can always use more time. I would have liked to explore more. I was always okay with where we went with Kuri, because I thought it fit the product. But, personally did I love it as a designer? Not necessarily. I wouldn't really have cared to have that voice in my home, but it was right for that product and right for that brand and I believe that. That's the hard thing as a designer, but I've learned over time to separate myself from those experiences. I do wish we had more time in the beginning phases because it was pretty quick. It was probably two months. I would have liked to explore much wider for sure.

So you design sound for something that is supposed to communicate emotions and be like a living thing with a personality. Did you ever encounter that task in any other of the design work that you did?

Yeah, actually I worked with postmates two years ago for a while on Serve. Have you heard of Serve?

I saw that, yes.

I did all the design with them and it was a very similar type of brief. It's a similar type of communication, but I mean to some extent every project has that. It's not real. It's not literal communication of emotion, but you're always trying to convey some type of emotion through a product or a brand sound, so there's always an element of that, but not quite to the level that you're speaking to here.

Okay, that's interesting! So, whether Kuri communicates delight or something like an app communicates delight – You say that there's a common thread there?

Definitely.

That's interesting. Connor, that's all i got! Thank you so much!