



Video Game Audio in the Metaverse and Beyond

Iain McGregor

Computing, Engineering, and the Built Environment, Edinburgh Napier University, Edinburgh Lothian Scotland. i.mcgregor@napier.ac.uk

Sound has arguably been a key component in almost every video game since their first inception, and whilst it is rarely the reason to choose a particular title, it can have an impact upon the level of engagement. Whilst some gamers choose to mute audio and provide their own musical selection, others utilize complex audio cues to improve their gameplay. Irrespective of where a player falls upon this scale, sound plays a key role, and there are many ways within the metaverse that it can more naturally integrate into a truly involving experience.

The Metaverse at first sight might seem like short-term hype. Looking beyond cumbersome Head Mounted Displays there are many opportunities for truly immersive video game audio. The trick is to incorporate the metaverse into the physical world. With the plethora of affordable off the shelf technologies it is entirely possible for developers to create extensive auditory environments within a gaming multiverse, without the expense and installation issues of complex audio hardware. Gamers will be able to experience a new sonic world that truly extends beyond an abstract metaverse, blending seamlessly with the physical experience of gaming.

Sound can be split across a surround system, controller, and open ear headphones to provide three distinct streams, with the Haas effect applied to compensate for any missing frequencies from the low-quality speaker in the pad. The player's avatar sounds come from open ear headphones, whilst their weapon emanates from the controller, with all of the other auditory cues transmitted through whatever speaker format is being utilized, from Dolby's Atmos, right down to built-in TV speakers.

Using audio watermarking and time alignment it is possible to configure ad-hoc networks where each Internet of Things (IoT) device that has a microphone and speaker becomes a sonic node to represent a point in the metaverse, decreasing the reliance on problematic HRTFs. Everything can sonically correspond in their physical world, irrespective of whether gamers are using a console connected to the TV, HMD, or smartphone as they move around a virtual environment.

Input from each mic can be easily transformed so that inhabitants not only choose what to look like, but sound like, in whatever language they wish. Transformation of audio has the added bonus of filtering for those who would prefer not to be exposed to the trash talk or worse, associated with some game franchises.

The metaverse facilitates selective auditory attention, where unlike the physical world everything becomes potentially audible, and extreme care needs to prevent either cognitive overload, or hearing damage. All of a gamer's movements can be tracked in order to predict what they are attending to. Each element of interest can then be sonified in a meaningful manner, from the macro, right through to the micro. User's distinct hearing abilities can also be compensated for, along with their listening preferences.

With a wide range of open ear headphone designs now available, from traditional open backed headphones, bone conduction, projection or even cartilage conduction, it is easy to augment any auditory environment so that sonically gamers can concurrently inhabit physical and virtual worlds. Friends can be remotely represented by companion robots or toys in the physical world, as well as having the ability to jump character, location or just perspective in the metaverse.

Much like the physical world, where each person provides their own sonic contribution, in the metaverse, players will bring more of their personalities with them. The increased level of visual customization, abilities and props will each require a sonic equivalent, that can either be generated procedurally, captured from the user's physical environment, or selected from

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2023 Copyright held by the owner/author(s).

2832-5516/2023/1-ART1 \$15.00

<https://doi.org/10.1145/3594534>

an extensive library. Tracking can be used to identify which sounds were considered successful and related auditory cues added to create a full representation, where the sonic reach is also a choice, both in terms of transmission and reception.

Whilst gamers will still be confined to what can be heard within the typical 20 Hz – 15 kHz, transformation of normally inaudible content will be expected. Super humanism will be presumed, an array of hydrophones across the Atlantic Ocean will be as easy to interpret as the single microphone currently on Mars. Audification, where waveforms are brought into the human audible range is already common practice in the sciences but will become mainstream when players start to appreciate how much it provides them with a tactical advantage. All of the sonic techniques employed over decades by spies and film crews can be adopted in the Metaverse, to such an extent that considerably more will be expected of technologies and experiences in the physical world.

Skill levels will easily be reflected in the sonic design, those that are novices will hear everything, players with intermediate abilities will experience a more selective auditory environment, whilst experts will inhabit the zone, only hearing what they really need. Whether an opponent's sword is dull or sharp will be sonically evident to an experienced sword master, whereas a novice will experience the cliched clang and swish. This approach follows through to every conceivable activity where gamers can choose to tag along with the virtuosi or transform their own actions into something well beyond their normal abilities.

The truly interesting factor is that each gamer will be able to inhabit their own unique soundscape, something even more intricate than the highly complex game sound designs which companies successfully strive to iterate upon each year. A new form of sonic design will be required that is psychoacoustic centric, and not necessarily acoustic based. Whilst physical, spatial accuracy is essential to begin with, in a technology where everyone is actively encouraged to choose their own avatars, there is absolutely no reason why anything should sound the same to any two listeners. The Builder culture so popular with younger gamers will become the norm with sound, whether it is intentionally borrowing from another source within the metaverse, or from accidental experience does not matter, it is the ability to explore and choose their own representation that will win out.

Navigating smoothly within the metaverse requires more sensors monitoring the player than commonly associated with video games, and it is this aspect that allows the smooth transition from the past, through the present and into the future. The gaming experience has moved from an object that you carried or visited (arcades), to something that you share your life and space with, to an experience that you can truly inhabit.

There will be inherent problems to address as developers experiment in this space, which are typical in periods of transition. There will be big leaps forward and some really big mistakes, such as veering too close to reality, with its emotional and sometimes physical consequences (hearing damage, vestibular balance issues...). But many developers already understand what some of those challenges are, and what the potential leaps that they will bring.

Sound designers will need to provide a considerably broader palette for gamers, with its inherent risk of cacophony, a situation which auditory interface designers have often struggled with. There will be users who wish to focus on the results of their actions, how extensive the damage is during an explosion, whilst others will concentrate on more intimate experiences with those that they are immediately interacting with. Fortunately, the amount of data captured by microphones and other sensors will provide more than sufficient information to moderate the auditory content into something much more manageable for listeners. Tracking what sonically attracts and repels gamers within the metaverse can facilitate a high level of inherent auditory customization without the need for seemingly endless menus, which could emphasize the artificial nature of the medium, but also provide needed reassurance that everything experienced has fewer real-world, physical consequences.