On the potential of spatial audio in enhancing virtual user experiences

Lauri Wuolio & Elina Moreira Kares

Digital transformation has altered the means of communication and use of digital tools both in our everyday lives and at work environments. This growing change took a massive leap further during the global Covid-19 pandemic, when many companies and people had to adapt to the new normal. This change also meant increased screen time for many as team meetings and business conferences had to be organised virtually.

Simultaneously, terms such as virtual and zoom fatigue took place in conversations. The change from physical to digital came with a toll, and many described a variety of negative physical and mental effects that seemed to be caused by the prolonged screen time. Research on the phenomena rose, and many scholars suggested that increased cognitive effort (e.g. Bailenson 2021; Riedl 2021; Nadler 2020), could be one major contributor to the phenomena.

Our House of Experiences-project is funded by the European Regional Fund as a part of recovery acts against the pandemic. It is run in collaboration with three other Finnish Universities of Applied Sciences; Laurea, Turku, and XAMK. The project aims to find new solutions and to develop consumer experiences and services in digital and physical spaces, that have been hit by the pandemic. Several focus areas for research and development were recognised through a wide partner network, and as one of them, the project grasped at the need of developing conference and meeting services for the changing needs put forward by digitalisation.

Hosting services for conferences are offered by several companies such as event companies, but also by many hotels and restaurants. The change in customer behaviour during the pandemic impacted these services directly, and when the customers moved online, their specific needs altered as well. To keep the standards of good customer experience, the companies had to adapt their services to fit the new needs and demands.

So how develop the customer experience further for the digital participants, and how make the digital meetings as user friendly as possible, and at the same time tackling the negative effects such as virtual fatigue? The House of Experiences -project took an approach to study these matters from the perspective of audio design, especially from the context of how utilising spatial audio could improve user experiences. An extensive literature review and a small experimental case study were built to examine the phenomena.

What is virtual spatial audio?

In 2007, when YouTube had been running for only two years, someone uploaded a video that contained only audio. In the video description people were instructed to use headphones and listen with their eyes closed. A wonderfully vivid scene followed with two

Italian sounding men speaking in English and guiding the listener through a haircut at a barbershop.

The video went viral and has gathered nearly 40 million streams on YouTube — quite a feat for a video that consists of text instructions only and no moving video footage! Yet, the viewers could clearly witness the events in the video, such as one of the men picking up a guitar and playing on it a couple of meters to the left of the barber chair, or the other man placing a plastic bag over the listener's head, before performing an aurally impressive haircut with scissors and an electric razor around the listener's head.

In 2023, a YouTube video from 2007 might seem ancient, but by the time of upload, the [Virtual Barber Shop](#) audio was already more than a decade old. It was originally created in 1996 by QSound Labs for a client and later uploaded to their website as a technology demo. Even though spatial audio today is often presented as a recent technological invention, listening back to the classic virtual barber shop audio clearly indicates that "virtual spatial audio" has been around for a good while.

In the case of Virtual Barber Shop, the audio is edited for binaural stereo — "binaural" meaning "two ears". The audio is mixed and processed to mimic how sounds appear in our ears in real life. Imagine you hear someone talking to you on your right side. The sound reaches your right ear a tiny fraction of a second earlier than the left ear. The sound that reaches your left ear is also muffled and altered by the mass of your head. There are also tiny reverberations of the sound that are reflected from the room walls and objects. From all this information your brain constructs a spatial imaging of the sound, telling you the location of the sound source. Thanks to evolution, all of this happens instantly without any conscious effort.

The history of virtual spatial audio

The Virtual Barber Shop audio was created in mid 1990s, but it might come as an even bigger surprise that the first binaural spatial audio experiments were made more than a century earlier, in the late 1800s. The "Théâtrophone", or the "theatre phone", was introduced in France in 1881 by **Clément Adler** and it allowed people to follow opera and theatre performances over telephone lines. From an array of eighty microphones a binaural stereo mix was created for the listener, who needed two earpieces (left and right channel) to enjoy the show in full stereo.

However, at the time a much simpler sound solution was needed, and one-channel mono sound became the technical standard for many years. Modern stereo sound was invented in the 1930s by **Alan Blumlein**, but it took until the 1970s before stereo sound took over the mono standard in the cinema and home entertainment systems.

In 1976, Dolby Laboratories created their Dolby Stereo standard, and it was later developed further into multi-channel surround systems, such as Dolby Surround, and most recently, Dolby Atmos. As a physical installation, with its sixteen loudspeakers, a fully-fledged Dolby Atmos system is almost beyond the reach of an individual consumer. However, in 2020 Apple announced that they will support Dolby Atmos in their newest mobile devices. Instead of cramming sixteen tiny speakers into their newest mobile

phones, Apple added a feature that allows the users to enjoy multi-channel Dolby Atmos mixes in stereo using headphones. While wildly different than the technology QSound Labs used to create their Virtual Barber Shop demo, Apple's Spatial Audio technology is based on the same binaural principles.

Spatiality offers a more immersive listening experience

Spatial audio is a term that is often used to describe various surround sound systems, where a setup of multiple speakers creates an illusion of a space or movement of sound within a space. In a movie theatre, for example, the voice of the main character can follow her even as she steps outside of the picture frame. By taking advantage of the multiple speakers placed around all sides of the movie theatre, the sound designers can move the film sound around the audience and control its movement and directionality, to create immersive and realistic sonic illusions.

While multi-speaker setups might offer a great solution for more immersive sound at movie theatres and even at living rooms of hobbyists, the price tag and technical complexity of these systems have created a need for much simpler solutions. Some of the early digital stereo systems in 1980's and 1990's added some sense of virtual spatiality in form of digital reverb effects, that made the music sound like they were being played in a different space. This approach, however, coloured the sound in undesired ways, often resulting in a sound, that was defined by the reverb itself — instead of producing a convincing feeling of "being there", the sounds just appeared as if they had been recorded in a reverberating space.

The sound of imaginary worlds

In the 90s, most of the development of the hi-fi audio systems focused on the digital products and software. The entertainment industry was waking up to the massive economical potential hiding in gaming, and great efforts were made to make great sound quality more affordable. With the advent of powerful consumer level computer audio cards, such as Creative Sound Blaster, the industry started focusing more on sound creation for games and interactive multimedia.

At the same time, the processing power of gaming consoles and computers grew rapidly. This allowed the developers to step beyond two dimensions in their games. Some of the classic 90s games, such as Doom or Super Mario 64, might have looked 3D — but they didn't sound 3D. Their fantastic 3D virtual worlds were being successfully "visualized" on the screens, which created a need for realistic "auralization" of their sound.

Contemporary game engines, such as Unity and Unreal Engine, especially when used with specialized audio engines such as Wwise and FMOD, allow realistic 3D rendering of sound in virtual spaces. Game engines, while mostly used for the creation of games, are already finding attraction outside of games. They are being explored by artists and designers for culture, education, and other uses. It is easy to imagine the benefits of realistic virtual representations of real world in the digital domain, especially in the field of education.

The interest in virtual spatial audio has grown steadily also in audio-only formats, such as music and podcasts. In 2020, Apple introduced their own Spatial Audio sound technology, that allows users to listen to content in Dolby Atmos format, in a more immersive way. Their technology works by using the sensors in some of their headphones. It creates a virtual sound field around the listener's head, while tracking the head movements. The experience comes close to listening to music in the sweet spot of a multi-speaker sound studio.

Imagine listening to a music performance where on your right you hear a sax player and on your left the bass player. As you move your head towards left, the bass moves from left closer to the centre, while sax moves even further to your right. Just like in the physical world, even the slightest head movement is translated as a tiny change in the stereo image presented in our ears. These small head movements can make the listening experience more natural and incredibly immersive.

Case study on the potential of spatial audio in video conferencing

To investigate whether a virtual spatial audio could be used to improve the user experience, an experimental case study was built up. The study examined the effects of audio design in the context of hybrid conferences, from the perspective of an online participant. A total sample of 40 volunteers (female = 52.6 %, male = 47.4 %) participated in the study, where they had to watch a series of five pre-recorded conference videos. In each of the videos four speakers discussed a chosen topic, that varied from video to video, e.g., from project management to the impacts of covid-19 on digitalisation.

In the study, the participants were divided into two groups. One received a usual monaural audio condition of the videos, that is the standard audio quality in video conferencing platforms. The other received a spatialised audio condition, which was edited to imitate audio in a naturalistic setting as the participant would have been in the room with the speakers, dividing sound sources into the respective physical locations of the speakers.

Besides traditional data collection methods, the study utilised Haaga-Helia's Service Experience Laboratory LAB8's biometric tools to analyse the impacts of spatialised audio. Traditional methods were survey questions, that were used to measure comprehension of speech, identification of speakers, and the perceived difficulty level of previous in each video. The biometric methods, including eye tracking (ET) and electrodermal activity (EDA) measurements, were used to analyse the cognitive load and arousal levels, respectively, that were theorised e.g. by Riedl (2021), as the root causes for the development of virtual fatigue.

The preliminary results revealed no significant major results between the spatial test group and monaural control group on the effects of audio design. However, some significant results were found with further analysis. Cognitive load was clearly impacted negatively by the darkness of video in comparison to videos filmed in a sunny daylight conditions, and the difference was greater in the monaural group.

In the comprehension of speech and the perceived difficulty of it, some group differences existed. When the test group performed well in the comprehension task, the perceived

difficulty was low. Vice versa the same trend could not be observed from the control group which tended to score the difficulty high nevertheless how the comprehension task went. More thorough analysis will be done and the results published later.

Conclusions

Even though the hypothesis on cognitive load as a premorbid factor of virtual fatigue could not be confirmed, the preliminary results of the case study already revealed some interesting insights on the potential benefits of using virtual spatial audio in improving user experiences. In the context of virtual conferences, spatial audio could potentially enhance the experience especially in cases of multiple simultaneous speakers, multiple speakers with similar voice characteristics, or when visual stimuli are less supportive for the perception. However, the potential benefits don't end there, as spatial audio could also be used to create more immersive virtual environments. With the coming of metaverse and other more engaging and realistic virtual spaces, why not take the audio design to the same level from the traditional monaural and stereo setups to spatial design, that mimics the acoustics of the real world.

Some software developers have already implemented virtual spatial audio in their products. Clubhouse is a social live radio application, where the listeners can join the live broadcast with their own phones. Virtual spatial audio was added to the application to allow easier distinction between speakers. Apple's Facetime calls are already utilizing their Spatial Audio technology and Verizon's BlueJeans is a video conferencing software, that uses a similar virtual speaker auralization technology.

For a long time, education and communication has emphasized the power of visual information. However, too much visual information might lead to increased cognitive load, so we need to develop better understanding when and what to communicate visually and when to rely on audio.

In the future, virtual spatial audio might open interesting doors in various fields, such as language learning, communications, virtual geography, and tourism. The best part is, we don't need to wait another ten years for further technological advancements. We already carry the needed technology in our pockets — and on both sides of our head.

References

Bailenson, J. N. 2021. Nonverbal Overload: A Theoretical Argument for the Causes of Zoom Fatigue. Technology, Mind, and Behavior, 2, 1.

Filimowicz, M. (Ed.). 2019. Foundations in Sound Design for Interactive Media: A Multidisciplinary Approach (1st ed.). Routledge.

Fırat, H.B., Maffei, L. & Masullo, M. 2022. 3D sound spatialization with game engines: the virtual acoustics performance of a game engine and a middleware for interactive audio design. Virtual Reality 26, 539–558.

Nadler, R. 2020. Understanding "Zoom fatigue": Theorizing spatial dynamics as third skins in computer-mediated communication. Computers and Composition, 58.

Riedl, R. 2021. On the stress potential of videoconferencing: definition and root causes of Zoom fatigue. Electronic Markets, 1-25.

Roginska, A., & Geluso, P. (Eds.). 2017. Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio (1st ed.). Routledge.