| | | |
|---|---|---|
| **Institution:** University of Huddersfield | | |
| **Unit of Assessment:** UoA12 (Engineering) | | |
| **Title of case study:** Innovative 3D Microphone Array Designs for Improved Immersive Listening Experiences | | |
| **Period when the underpinning research was undertaken:** May 2011–present | | |
| **Details of staff conducting the underpinning research from the submitting unit:** | | |
| **Name(s):** | **Role(s) (e.g. job title):** | **Period(s) employed by submitting HEI:** |
| Hyunkook Lee | Reader | Sep 2010–present |
| **Period when the claimed impact occurred:** 2014 – Present | | |
| **Is this case study continued from a case study submitted in 2014?** No | | |

## 1. Summary of the impact

3D sound recording systems have often operated sub-optimally because no theoretical model of the relationship between the set-up of the microphones (the array) and the way that a listener perceived the resulting playback existed.

Research from the University of Huddersfield created such a model and established novel principles that enabled microphone arrays to have a more compact form, while improving the sense of realism and spatial impression for the listener.

The knowledge was commercially exploited as the design basis for an award-winning microphone array by Schoeps, a world-renowned microphone manufacturer based in Germany. The new array was adopted by sound engineers for major broadcast events such as the FIFA World Cup, BBC Proms and the French Open. Other microphone arrays and an array design app that was based on the research became essential tools for 3D sound recording at Abbey Road Studios (UK), Austrian Broadcasting Corporation (Austria), Arizona Public Broadcasting Services (USA) and Tianjin Juilliard School for musicians (China).

## 2. Underpinning research

Since 2011, use of three-dimensional (3D) and surround sound (such as Dolby Atmos and Auro-3D), has grown rapidly in cinemas, music and broadcasting. The techniques use additional loudspeakers (or they manipulate the stereo outputs) to give the illusion that the listener is ‘within’ the soundscape. Changes in practice for both recording and playback were developed by sound engineers. To record 3D sound they positioned groups of microphones (arrays) that captured the directional information on the sound in a way that meant it could be manipulated to give the listener a realistic experience. On playback additional speakers in the height dimension were employed to give a ‘vertical’ element to the sound. Sound engineering has always been a ‘black art’, with best practice resulting from experience mixed with trial and error. The relationship between how the microphones were positioned and the resultant aural experience for a listener was not scientifically understood. The research described in this case study explored the relationship between the physics of sound recording and reproduction and how humans perceive what they hear – a branch of psychophysics called psychoacoustics. The findings identified principles and algorithms that could be used to configure a microphone array to give consistent and reproducible results, whilst helping the sound engineer to mould their desired sound world.

The research described in this case study was carried out at the University of Huddersfield (UoH) by Dr Hyunkook Lee (Reader at UoH since 2010). In 2013, he established the Applied Psychoacoustics Lab (APL), which is now recognized as one of the world-leading groups in this field. Others who contributed to the research include Lee's previous PhD students Dr Chris Gribben (Acoustic Engineer at B&O, Denmark), Dr Dale Johnson (Research Fellow at UoH) and Dr Rory Wallis (Analytics Manager at AIMS Smarter, UK).

All the research described below used a similar methodology. Each hypothesis was tested using a panel of 15–20 volunteers (the typical number for hearing-based research since the panellists must have sufficient aural acuity to give reliable feedback), who provided their subjective view of the quality of each aural experience based on a questionnaire. Their responses were statistically analysed. Physical measurements were collected in parallel that were used to define physical parameters and ensure reproducibility.

To test the variables associated with 3D recording and reproduction, the effect of altering a number of elements of common practice were modelled. Typical sound engineer practice was compared to see how it measured up against the optimum listener experience.

(i) Experience had taught sound engineers that, for the most natural-sounding stereo reproduction, two microphones had to be separated in the horizontal plane. By analogy, they had assumed that pairs of microphones needed to be positioned at different heights to give the best results in the vertical plane. Lee conducted the first piece of formal research on this topic [R1] (2014), by testing multiple microphone heights and separations.

He found that having a space between vertically arranged microphones provided no perceptual benefit to the listener – the gap could be as low as zero. This resulted in a paradigm shift in 3D microphone array design. Conventional arrays had a cube-like form factor, but the research showed that a more compact, horizontally spaced and vertically 'flat' array, gave listeners a more accurate impression of the actual sound. Psychoacoustically, this can be explained by understanding that a pair of human ears are in the same horizontal plane.

(ii) The accepted best-practice of separating microphones in the vertical plane led to 'interchannel crosstalk' because the upper microphone captured the same sound as the lower one, which made the resulting sound 'image' more blurry due to interference. Using a similar experimental strategy to (i), Lee [R2] (2017) showed that angling the upper microphone by a fixed amount (so that it no longer directly faced the sound source), meant that it preferentially picked up ambient sound, which led to more accurate 3D sound image reproduction. This so-called psychoacoustic threshold level difference was measured to be 7dB and became the theoretical basis for determining the vertical angle between the microphones for several arrays designed by professional sound engineers.

(iii) To try to improve the perceived 'spread' of a vertical sound image, the signal to each speaker was slightly manipulated (decorrelated) by sound engineers because it improved how 'natural' a sound appeared when using a conventional speaker set-up. Research [R3] (2017) proved that applying decorrelation in the vertical plane had a negligible impact on the perception of the listener.

(iv) The perceived quality of a recorded sound is always a trade-off between the degree of 'space' in the soundscape and accuracy with which the listener can pinpoint the position of a particular sound. This is one element of the 'art' of the sound engineer. The conventional approach for positioning the microphones in an array that delivered the most accurate 'position' in 4-channel 360° reproduction, was to separate them by 23cm. Testing, using an expert panel of listeners, showed that this was wrong with the best results coming from a 50cm separation [R4] (2019). This led Lee to develop a new microphone array, the ESMA-3D. It provided instructions on how to position existing microphones for recording and was incorporated into a new psychoacoustic model [R6].

(v) Lee further refined the model to enable it to accurately predict the position that a sound source appeared to a listener, based on the physical position of the sound sources and the positions of the microphone arrays relative to them. This is known as phantom image localization

[R5] (2013). Different notes from various musical instruments (piano, trumpet, etc.) were recorded and analysed in a real setting. The results were then modelled. This generated the data needed to locate the sound image of a source of sound at a specific position between a set of loudspeakers. The resulting localization prediction model could be used to predict the position of perceived sound images.

(vi) A software app, the Microphone Array Recording and Reproduction Simulator (MARRS) was developed by Lee and Johnson based on the data and model [R5]. It enabled the simulation and automatic configuration of a microphone array [R6] (2017) and gave a more accurate prediction of perceived sound image positions than existing tools. It also provided a more intuitive visualization for the sound engineer; based on the position of the sound sources and other input parameters, such as how 'spacious' the sound should be and how accurately it should be possible to locate the sound, the app recommends how to configure the microphone arrays and where to place them.

In summary, the research has codified the techniques that sound engineers used for 3D and surround sound recording and shown that their traditional approaches were sub-optimal. It has produced novel methods that improve the spatial impression of recorded 3D sounds and reduced the microphone array size significantly, compared to conventional methods.

## 3. References to the research

The references listed below are considered to have a 4*/3*/2* quality. [R1] to [R5] have been published in high quality journals: Applied Sciences (Scimago Q1 in Engineering) or Journal of the Audio Engineering Society (Scimago Q2 in Engineering, Q1 in Music). [R6] has been presented at an AES International Convention, which is one of the most prestigious conferences dedicated to audio.

[R1] Lee, H. and Gribben, C. (2014) 'Effect of Vertical Microphone Layer Spacing for a 3D Microphone Array', Journal of the Audio Engineering Society, 62 (12), pp. 870–884. https://doi.org/10.17743/jaes.2014.0045

[R2] Wallis, R. and Lee, H. (2017) 'The Reduction of Vertical Interchannel Crosstalk: The Analysis of Localisation Thresholds for Natural Sound Sources', Applied sciences, 7 (3), pp. 278. https://doi.org/10.3390/app7030278

[R3] Gribben, C. and Lee, H. (2017) 'A Comparison between Horizontal and Vertical Interchannel Decorrelation', Applied sciences, 7 (11), pp. 1202. https://doi.org/10.3390/app7111202

[R4] Lee, H. (2019) 'Capturing 360° audio using an Equal segment microphone array (ESMA)', Journal of the Audio Engineering Society, 65 (9), pp. 733–748 https://doi.org/10.17743/jaes.2018.0068

[R5] Lee, H. and Rumsey, F. (2013) 'Level and Time Panning of Phantom Images for Musical Sources', Journal of the Audio Engineering Society, 61 (12), pp. 978–988. http://www.aes.org/e-lib/browse.cfm?elib=17075 [can be supplied on request]

[R6] Lee, H., Johnson, D. and Mironovs, M. (2017) 'An Interactive and Intelligent Tool for Microphone Array Design', In proceedings of Audio Engineering Society 143rd Convention. http://www.aes.org/e-lib/browse.cfm?elib=19338 [can be supplied on request]

## 4. Details of the impact

The research generated novel methods to improve the spatial impression (a sense of the relative positions of sound sources and perceived spaciousness) of recorded 3D sounds, whilst reducing the microphone array size significantly compared to conventional methods. Four different 3D microphone arrays were developed and these were exploited by the sound-recording industry for improved 3D music recording (a radio station and a classical music concert), outdoor recording (a microphone manufacturer and a TV station) and for virtual reality music recording (a recording studio and a VR content production company).

The impacts can be summarized under three headings:

1. Influencing the design of a commercial 3D audio microphone array.
2. Enhancing immersive sound recording techniques.
3. Improving the realism of recorded classical music.

The research findings led to talks and workshops on 3D audio at international conferences, including Audio Engineering Society (AES) Conference on Spatial Reproduction (2018). These resulted in industry interest and subsequent adoption of the new approaches.

**Influencing the Design of a Commercial 3D Audio Microphone Array**
Schoeps is a best-in-class Germany-based manufacturer of microphones for sound professionals. Their microphone array product, the ORTF-3D, was completely redesigned in 2016 based on consultation with Lee, which applied the main findings of three papers [R1, 2, 3]. Before redesign, the array had a cube-like form factor and its large size made it impractical for sound engineers to carry and install easily. This was a particular problem for outdoor location recording and broadcasting environments, such as music concerts and sports events. The company said UoH research "helped [it] to identify crucial and non-crucial aspects of microphone array geometry and the extent to which certain design parameters influence spatial perception" [S1]. As a result, the research "significantly influenced the design of 3D audio microphone arrays at Schoeps" [S1] and led to a new version of the ORTF-3D that was much more compact. The vertical microphone spacing was reduced to zero, which reduced the size and weight of the unit without affecting its performance. This gave users benefits in terms of portability, set-up time and health and safety. ORTF-3D won the prestigious TEC Award for its innovative design during the 32nd NAMM Show (the world's largest music products trade fair) in 2017 (tecawards.org/tec-winners).

The new version of ORTF-3D was rapidly adopted and used for broadcasting a number of major musical and sports events, such as the BBC Proms (2016–2019), FIFA World Cup (2018), and the French Open (2018–2019). The BBC's Technical Producer, responsible for recording and broadcasting the BBC Proms concerts, stated that ORTF-3D provided "flexibility when mixing and better spatial impression than Ambisonics and other coincident mic arrays" [S2]. The sound engineer who was in charge of capturing the crowd sounds of the 2018 Russia World Cup football matches for HBS (a Swiss sports broadcaster) worldwide broadcasting, testified that ORTF-3D "played a major role in delivering excellent immersiveness and precise localisation for crowd sounds in 3D… which would have not been possible with a conventional microphone array like First-Order Ambisonics". He said that the microphone array provided "a far better sense of envelopment and a more accurate localisation as well as a larger listening area, which is of paramount importance for TV viewers" [S3].

**Enhancing Immersive Sound Recording Techniques**
The reputation of the UoH APL 3D audio research led to consultancy on immersive sound recording techniques with Abbey Road Studios (London), one of the most famous recording studios in the world. The UoH-developed ESMA-3D microphone array [R4] was introduced to Abbey Road through a collaborative project in 2017. Since that time it has become the studio's go-to microphone array for various types of 3D recording [S4]. One example was a large orchestral recording session for the Electronic Arts' computer game, Star Wars Jedi: Fallen Order, which was performed by London Symphony Orchestra [S4, S5]. The award-winning recording and mixing engineer who worked on the session became "a fan of the array from the initial experiment sessions for the game conducted in March 2019" [S4]. The excellent 360° imaging the array offered, led him to develop a new way to bridge the sonic gap between the strings and the woodwinds, which produced impressive results [S4]. The Head of Audio Product at Abbey Road stated that "of all the arrays I have tried over the last couple of years, ESMA-3D is the most useful and sonically pleasing microphone array with versatile end user applications from stereo to 5.1 and Ambisonics to Dolby Atmos" [S4]. ESMA-3D has also become an essential tool to "glue various performers in the room into a cohesive acoustic fingerprint" in Abbey Road's pioneering work on six-degrees-of-freedom virtual reality (VR) audio, which adds in a representation of body movement to the standard three degrees used in conventional VR [S4]. ESMA-3D was also "the backbone" of a VR experience created for Highways England (HE)

[S6] by MagicBeans, a start-up dedicated to providing highly immersive virtual and augmented reality experiences. It recorded various types of motorway traffic noise using the array and used the recordings to create an immersive training environment for roadside workers using VR (2019). The CEO of MagicBeans said that "the imaging stability and sweet spot with ESMA-3D is so much better and larger than with a comparable Ambisonic recording, and the spaciousness and timbre of the ESMA-3D recordings are far better". He added, "We would not have been able to provide this large-scale sound experience without this manner of recording being pioneered by Dr. Lee …. [it is] a significant step forward in making high quality 3D captures practical and useful" [S6].

**Improving the Realism of Recorded Classical Music**
Central Sound (CS), an award-winning audio production service, is part of Arizona PBS (Public Broadcast Service). It specializes in capturing acoustic music performances for broadcast. The PCMA-3D microphone array [R1] helped CS to reach their goal of a "quintessential sound" in their immersive recordings for the Grammy-Award-winning Phoenix Chorale and Grammy-nominated True Concord Voices and Orchestra in 2019 [S7]. Having moved away from the 'trial and error' approach, their chief recording engineer and manager said "the test methodology Dr Lee has employed gives us confidence that we are utilising proven techniques" [S7].

The Austrian Broadcasting Corporation (ORF) was the first public broadcaster in Europe to start a 5.1 surround sound service and their senior sound engineer was one of the pioneers in 5.1 surround recording. He explored new techniques for outdoor 3D audio recording for broadcast and, working with Lee, developed a new 3D microphone array (2018) [S8]. Named the Double UFIX it used the novel UoH localization prediction model [R5] and the associated iOS/Android MARRS app [R6]. The ORF engineer said that the UoH research "significantly increased sense of realism or an even more convincing illusion of the resulting sounds" [S8].

The app provided "an intuitive way to derive tailormade microphone arrangements for specific recording situations" [S8]. It helped sound recording engineers achieve their goals for accurate and spacious imaging in recording, by visualizing the virtual sound images resulting from different microphone configurations. Compared to other similar tools, MARRS helped them produce results more quickly, accurately and reproducibly. The app is currently used by 600+ users worldwide. The Director of Recording and Music Technology at The Tianjin Juilliard School for musicians in China has used MARRS as a starting point of microphone placement for many of his professional recording sessions, including Peabody Symphony Orchestra and Tianjin International Chamber Music Festival [S9]. He stated that, "it saves me a lot of time in correctly configuring a microphone array and achieving an accurate stereo imaging, thanks to the accurate sound image visualization and the powerful parameter control options it provides" [S9].

**5. Sources to corroborate the impact**

[S1] Letter from CEO, Schoeps (world-renowned microphone manufacturer
[S2] Letter from Technical Producer, BBC Radio and BBC R&D
[S3] Letter from Professor/Tonmeister, University of Applied Sciences Darmstadt
[S4] Letter from Head of Audio Products, Abbey Road Studios
[S5] Stiles, M. (2019) Composer Stephen Barton Returns for Spatial Audio Experiments [online] Available at: https://www.abbeyroad.com/news/composer-stephen-barton-returns-for-spatial-audio-experiments-2554 [Accessed: 20 Jan 2021]
[S6] Letter from CEO, MagicBeans (VR training specialist)
[S7] Letter from Manager, Central Sound at Arizona PBS
[S8] Letter from Senior Sound Engineer, Austrian Broadcasting Corporation (ORF)
[S9] Letter from The Director of Recording and Music Technology, The Tianjin Juilliard School