Designing a sonic landscape: A practice-led approach to creating 3-D sound space for screen.

Dr Damian Candusso
Introduction
Several 3-D feature films that the author has recently been engaged on, including *The Great Gatsby* (Luhrmann, 2013) and *The LEGO Movie* (Lord and Miller, 2014), have followed strict schedules and followed a known formula for producing a high quality soundtrack that satisfies the studio, and equally importantly, the cinema audience. With such high profile and financially successful films, it is difficult to argue that the formula for producing these high quality soundtracks needs to be further investigated. *The LEGO Movie* opened in the US with the box office taking in excess of US$69 million, well above the estimated entire film budget of US$60 million in a single weekend ([Box Office Mojo](http://www.boxofficemojo.com), 2014).

Many 3-D feature films continue to adhere to traditional 2-D soundtrack methodologies and practices, with the film industry following a steady path of innovation in relation to developments of the soundtrack and in particular soundtrack exhibition. Generally speaking, advancements within the industry have seen an increase in the number of speaker channels over time, for example; mono to stereo to quad surround to 5.1 surround to 7.1 surround and now the introduction of sixty-four channel immersive sound technologies. Although several formats have explored additional sound channels throughout this linear evolution of sound on film including Fantasound, a general trend can be observed. However, does increasing speaker channels increase the immersive experience of film, or can an alternative soundtrack be designed to better immerse an audience?

Previous format limitations were a result of how many sound channels could be accommodated within the available film format of the time; for example, the physical space limitations on 35mm film and 70mm film. However, the introduction of digital film has drastically increased the capacity for sound channels through the DCP film format. Although developments in cinema sound have led to increased sound channels, I question if changing the sound environment can increase the spatial homogeneity between the 3-D image and the soundtrack.
An initial line of enquiry into 3-D sound actually resulted from a single problem that was identified whilst the author was working on the first Australian 3-D animated feature film, *Legend of The Guardians: The Owls of Ga’Hoole* (Snyder, 2010). The problem presented itself through the inability to position sound accurately within z-space (Figure 1).

![Figure 1: Z-Space](image)

An object could be panned around the room, from left to right of screen or from the front to the rear of the cinema; however, the sound object could not be located within the centre of the room. No matter how the various sound effects were designed and no matter how the objects were panned throughout the 5.1 surround sound space, limitations of the 5.1 format did not allow for the accurate positioning of the sounds along the z-axis. In particular, this was highlighted with a sword tip passing slowly in front of the audience members’ noses as the 3-D imagery showed a blade converged off the screen in strong negative parallax (Figure 2).
The objective was to create the illusion that the sound had volumetric depth that complemented the imagery.

The tip of the blade originated from the neutral parallax (on the screen plane) before extending into negative parallax as it travelled from the screen into the audience. Problems arose when trying to replicate this with various layers of sounds using the parameters of 5.1 panning. Although able to pan the base of the blade as it was anchored on the screen plane, the tip of the blade in strong negative parallax proved extremely difficult. When the sound was panned from the screen into the surround speakers, it became diffused throughout the theatre. To achieve an acceptable outcome, several sound elements were panned independently to create the aural illusion that the overall sound was attached to the blade. This however was a compromise.

Through practice-led investigation, this paper questions the potential immersive environment of the soundtrack through alternatives to loudspeaker playback through the use and portrayal of sound space.
Defining an aural spatial landscape

Gravity (Cuarón, 2013), a film set in space, provides an example of defining sound for space. It is the detail in establishing the environmental vastness that enables the soundtrack to punctuate and accentuate this infinite perception of space portrayed by the visuals. Shot in 3-D, Gravity has explored not only the visual medium, but also the sonic medium so that both elements work in unison. As Smalley suggests (Smalley 37), ‘A listener needs time to progress from an initial listening encounter with the soundscape to a state of engaging actively and fully in scanning and exploring the spectromorphological and spatial properties on offer.’ Gravity provides an example of this.

The film begins with a musical crescendo before abruptly stopping to complete silence, reflecting the science of space. In many ways Gravity is mixed quite unconventionally, in that the dialogues do in fact pan across screen and beyond, with the characters even panning into the surround speakers. Notably the music is also composed for surround sound with different elements panning throughout the various speakers, at times spinning around the room alongside the camera and dialogues. In an interview for the Soundworks Collection, director Alfonso Cuarón makes particular mention of how the panning of the dialogues and the music helps provide a superior immersive experience (Coleman, 2013). In combination with the atmospheric soundscape, Gravity is successful in providing an aural illusion of infinite space by allowing the music and dialogue to pan and make use of all speakers. This was a brave move by Cuarón as it goes against convention. Cuarón made a conscious decision to allow the sound and the speakers to work for the narrative of the film, rather than following traditional mixing practices.

The art of mixing a film is knowing where to direct the audience’s attention in relation to the story. In many instances and as history has shown, the direction is always forward towards the screen. Gravity is an example of these rules having been broken deliberately. By complementing the unique story and location of the film, this is done
deliberately to orientate the audience with the characters as they are often off screen, floating in space.

**Practice-led research**

As an exception to norm, *Gravity* provides a remarkable example of suturing the audience with the film landscape (or lack thereof). So how can this work when we are not in space? Through practice-led research various alternative approaches for creating an immersive and homogenous 3-D film soundtrack have been investigated. This includes manipulating the spatial environment of the atmospheres, dialogue and foley, recording sound with localisation considerations in ambisonic and binaural formats, and where recording was not an option, explore digitally synthesised 3-D spatial sound options.

Two films have been selected for this enquiry. These are:

- *Carwash* (Candusso, 2013)
- *Foxed!* (Stewart and Bezaire, 2013)

Through practice-led research I am afforded the luxury of not having to work to any preconceived convention as I investigate the creation of homogenous 3-D spatialisation. This investigation questions whether it is possible to achieve a more immersive, realistic, and homogenous 3-D audio-visual experience through creatively designing a soundtrack beyond existing technologies and methodologies.

**Carwash (3-D)**
Carwash is an original film created by the author. It is a single shot, first person point of view (POV) scene recorded from inside an actual carwash.

**Aim**
The aim of Carwash was to investigate the accurate reproduction of 3-D sound and vision from capture through to exhibition. The limitations of the 5.1 sound format were considered in the conception of the film. Contemporary cinema sound formats provide a canopy effect and Carwash exploits this. Carwash was purposely created to allow for the simultaneous filming in stereoscopic 3-D whilst capturing the location sound in various formats including binaural stereo and ambisonic B-format, simulating a virtual head. I hypothesise that the virtual head would record vision and sound in 3-D, and when played back, would allow the viewer to see and hear in 3-D.

**Method**
Filming Carwash was a reversal of the industry practice of capturing the best imagery first and considering the sound later. Recording the best sound was prioritised, with the image a secondary consideration. For the purpose of this exercise, my priority was to capture sound recordings that could best replicate 3-D sound whilst maintaining homogeneity with the 3-D imagery. Location sound for 3-D
films is almost never recorded in a 3-D format. Carwash embarked on creating a short film-clip that challenged this notion.

![Figure 4: Binaural head with 3-D GoPro camera attached and Soundfield microphone](image)

This film was intended to situate the spectator within the environment with a first person perspective. To create the POV (point of view) experience for the viewer, a 3-D camera rig was mounted to a binaural head. This ‘virtual head’ allowed the capture of simultaneous sound and vision, akin to our own two eyes and two ears. There was no post-production or processing made to either the sound or vision, allowing the original recordings to be unadulterated. An additional binaural version of the film soundtrack was created using the software plug-in Auro-Headphones. Three versions of Carwash were created.

- 5.1 version created from the original ambisonic recording. [click here to watch in anaglyphic 3-D – requires 5.1 sound system]
- Binaural version recorded through the dummy head. [click here to watch in anaglyphic 3D- requires headphones]
- Auro-Headphone (synthesized binaural) version created from folding down the 5.1 version. [click here to watch in anaglyphic 3D- requires headphones]

**Results and discussion**
Having experimented with binaural and ambisonic sound formats prior to this research, I already had experience and an understanding of the capabilities and limitations of each format. However, I had never recorded these audio formats simultaneously with 3-D image capture. Although binaural sound has previously been simultaneously recorded with visuals on films including *Bad Boy Bubby* (de Heer, 1993) and *Souviens-Moi* (Derobe, 2013) I am unaware of any films that have been recorded and released with an unchanged or unedited original on-set recorded soundtrack. Ambisonic recordings are often used to record atmospheres, but they are then combined with the overall soundtrack.

*Carwash* was created to synchronously capture 3-D imagery and various sound formats in a simplistic controlled environment. The aim was to accurately explore the reproduction of 3-D sound and vision from capture through to exhibition and this was successfully achieved. The film represents actuality as it was filmed and recorded from a single point of view in a single take. One of the shortcomings of conventional speaker sound systems is that they are unable to render sound accurately within z-space. As 3-D imagery converges off the screen into negative parallax, this creates dislocation as the sound emanates from around the audience, creating a cocoon effect. *Carwash* deliberately capitalises on this by using an environment that is a cocoon. That is, the car represents a cocoon, with the sound of the carwash emanating outside it.

The film contains no dialogue or any additional sound effects, with only the sounds that took place at the location being exhibited. *Carwash* provides an example of successfully creating a homogenous soundtrack and 3-D imagery in 5.1 and two versions of binaural stereo headphone formats.

Although successful in creating a homogenous audio-visual experience, applying the methodology used on *Carwash* to a feature film would be impractical, but not impossible. Having dialogue and/or camera angles with varied use of parallax would
make the capturing of sound very time consuming. This would be most noticeable in the editing process. However, it would be successful, as proven here, for short films.

**Foxed!**

**Aim**

*Foxed!* is a 3-D Canadian production that had a pre-existing 5.1 soundtrack prior to me being granted permission to use the film for the purpose of this investigation. The original 5.1 mix was created conventionally using common film mixing methodologies, in the same manner as a feature film. As noted by Kerins, 'most dialogue continues to reside in the front centre channel, and many filmmakers still have some reservations about using the surround channels too aggressively.' (Kerins 71) The original sound mix demonstrated many of the shortcomings of contemporary surround sound formats used with 3-D film. [(Click here to view the original stereo version of the film)](#)

The practice-led research objectives of considering this film were to create a binaural headphone version of an immersive sound mix. The aim was to:

- Create a more immersive film through manipulating the existing soundtrack.
- Create a soundtrack that is homogenous with the 3-D imagery.
- Support the point-of-view shots aurally to better situate and immerse the audience in the first person.

**Method**

In the original approach to the remix of *Foxed!*, I intended to create all pan data in a 5.1 version, and then through the use of plug-ins, fold down the mix into a binaural version. It was intended that all 3-D spatial information from the 5.1 version would be compatible, thus saving time and providing an efficient workflow. After several early mixing tests, this methodology proved to be unsuccessful, as the z-axis information did not translate into the headphone mix. This necessitated creating a mix natively in the binaural format. Through the creation of a binaural stereo headphone mix, the
film intended to emulate sound within a 3-D space, with panning in the x, y and z-axes as appropriate with the 3-D imagery

**Results and discussion**

All raw-sounds were supplied from the studio that created the original soundtrack. However, all mix data including volume, reverb, equalisation, compression and all effects were not supplied. This allowed the film to be remixed without knowing any original settings, or having any preconceived judgment.

*Foxed! (Binaural - 3-D remix)*

*Foxed!* was never intended to have a binaural soundtrack release and as such, the supplied sound recordings were not in the binaural format. All supplied sounds were a combination of standard monophonic and stereo files. Although conventional for loudspeaker formats, the binaural format requires that sounds be recorded (encoded) using a specific binaural stereo recording configuration. The challenge was to create a binaural soundtrack from non-binaural encoded audio.

Re-recording every sound again in binaural was not an option. Instead, I manipulated the original recordings using computer software processing. In some isolated instances, I re-recorded a few key sounds where absolutely necessary, including some of the original dialogue, using the custom-built binaural head microphone (*Figure 5*). The recordings were necessary as the binaural plug-in failed to reproduce much of the spatial positioning accurately within z-space.

*Figure 5: Custom made binaural head using mannequin head, silicon ears and condenser microphones inside ear canals.*
Creating the binaural mix required a complete remix from the original source files and also necessitated the wearing of headphones for the entire mixing process. A binaural mix must factor in all pan and spatial data necessary to resemble 3-D positioning of sound using only two channels. This posed many challenges including track routing, reverb usage, plug-in efficiencies and other restrictions.

**Creating 3-D dialogues for headphones.**

In cinema it is a relatively simple process to down-mix from 5.1 (or any other surround format) to a stereo version, as all of the tracks are ‘folded down’ – that is, the left, centre, right, surround and LFE tracks are folded down to create a single stereo mix. Although the headphone mix of this film is technically stereo, it is also binaurally encoded and thus the fold down from 5.1 does not translate as the y and z-axes 3-D spatial information is lost. A stereo signal contains only the left and right information (the single x-axis); however stereo binaural is encoded with spatial information for all three axes.

The dialogues of *Foxed!* provided the greatest challenge for the binaural soundtrack. In addition to conveying the narrative, the dialogue was mixed to provide a spatially accurate relationship with the 3-D imagery. This is evident when the two foxes are looking for Emily (1 min 5 sec), when Emily hears the discussion between her mother and the fox that has replaced her (2 min 30 sec), and also during the opening of the film. During the opening, the audience is introduced to Emily through hearing her breaths. These were mixed spatially to mirror the localisation of the imagery. One of the greatest challenges is highlighted through the title sequence and again when the audience experiences Emily’s point of view through vision and sound as she runs up the stairs (1 min. 34 sec.). Because the dialogue was originally recorded in Canada, there was no option but to use the original breaths, and artificially create the effect of the breaths coming from the viewer’s point of view. A combination of mixing techniques, including reverberation manipulation and also a third party binaural plug-in (H3-D Binaural Spatializer –v2.1.2 by Longcat.), was used to create this illusion.
The dialogue for this scene was duplicated onto two separate mono tracks. One track was reserved for the POV shots, and the other track for all other shots of Emily. This enabled the freedom to easily dedicate separate pan, reverb and binaural spatial settings independently of each track.

As Emily arrives at her home (1 min. 38 sec), she is separated from her mother by a glass window. From inside the house, her mother is unable to see or hear Emily as the glass is characteristic of a one-way mirror. The approach to this scene was to not only use sound localisation, but also to portray and replicate the acoustic spaces based on the image and the narrative. The original dialogues from the film were re-recorded using a combination of convolution and binaural recording techniques with the dummy head microphone. Figure 6 below shows the binaural head in a room that is separated by glass doors from the speaker replaying the sound.

![Figure 6: Convolution binaural recording of dialogue through glass – dummy head in distance](image)

The convolved dialogues for the mix include a combination of synthesised 3-D spatialisation and 3-D binaural recordings captured in a room resembling the acoustic space as prescribed by the visuals. Gierlich mentions that through binaural post-processing, sound engineers can play a creative role by inventing new sound
situations above and beyond merely recording the original sound situation as authentically as possible. (Gierlich 227) This practice-led research demonstrated that processing beyond binaural post-processing was necessary. All of the spatialisation of Emily’s dialogue during the opening scenes was synthesised using plug-ins; however, once Emily arrives back at her home and is peering through the window, the dialogue is a binaurally recreated re-recording from the original film. Not having access to the original cast necessitated that I consider my available options in creating dialogue within the 3-D space. Using a combination of binaural synthesis and capturing the dialogue through the binaural head provided the required spatial results.

Notably, the primary factor that enables the dialogues to be presented in a 3-D space is the labour-intensive heavy manipulation of post-production effects. This includes the use of filtering, equalisation and reverb qualities, including spread and tonality.

**Foxed! Conclusion**

*Foxed!* was remixed in the binaural format in order to investigate the effectiveness of creating a more immersive mix; of creating sound that homogenously matches the 3-D imagery within z-space; and of situating the audience within the first person perspective for the POV shots. Although providing different results, the remix offered a more immersive mix than the original version.

The binaural remix provides a more homogenous audio-visual experience than the 5.1 version as sounds could be situated with more accuracy along the z-axis. Although certain elements were effective in the 5.1 version, the binaural version provided a more immersive and more homogenous experience with the 3-D visuals. There was not a single solution in providing an effective binaural mix. Plug-in processing and additional re-recording were needed on various shots and this added to the complexity of the mix. This complexity also meant that the binaural mix required far more time to create than a conventional 5.1 mix.
The limitations of the investigation included working with pre-recorded sounds and not having the original sounds recorded in the binaural format.

**Conclusion**

Through practice-led research the author challenged contemporary feature film mixing practices and methodologies. This approach allowed an investigation into creating a more immersive, realistic and homogenous 3-D audio-visual experience without the restrictions of commercial film practices. The remixes exploited the opportunity to pan all elements within the soundtrack, including dialogue and foley.

Remixing in the binaural format demonstrated that a more homogenous soundtrack can be obtained compared to the original releases by panning all sound elements including dialogues. Using an interpretation of Smalley’s space zones, (Figure 7) as sounds move from ‘panoramic space’ to ‘circumspace’ within the 5.1 format, the definition in the sounds’ positional rendering is lost. Sounds that are located in the ‘panoramic space’ have convincing acoustic depth in the 5.1 format, with the atmospheres from all clips providing examples of this.(Smalley 48) If a sound pans quickly through these spaces, the 5.1 format can also be effective in providing homogeneity.
The panning of sounds from panoramic in front of the listener to circumspace and then through panoramic behind the listener, is also effective in providing the illusion of sounds passing through these spaces, with *Foxed!* providing several examples. This includes the sparks at the beginning of the film and also Emily running during the opening title card. In very specific cases, the 5.1 format can work effectively in providing homogeneity for the entire soundtrack, not just individual elements, if a film is created without z-space action or dialogue, and therefore no sound in the circumspace as with *Carwash*. This takes advantage of the cocoon effect of speaker formats.

Creating the various headphone remixes required extensive individual attention. *Carwash* was an original film that allowed the sound to be recorded specifically for the 3-D medium whereas *Foxed!* contained an original soundtrack that needed a considered approach to each particular shot. Headphone technologies allow for improved homogeneity with sounds located within the ‘circumspace’ and for POV shots, as demonstrated in all mixes. However, creating headphone specific mixes...
was time consuming, with no single technique providing a simple solution. The use of binaural plug-ins was not effective in all instances, with additional binaural re-recordings required for *Foxed!*. Preparing early in the production stages, and accommodating binaural technologies throughout the recording process is of benefit as demonstrated through *Carwash*.

Overall the headphone mixes provided a homogenous and spatially accurate relationship with z-space 3-D imagery compared to the 5.1 mixes. This became increasingly apparent as sounds moved closer to the ‘circumspace’ as in Figure 7. Contemporary speaker formats provide superior bass (LFE) reproduction that enables the listener to physically feel a sound. A binaural format with a dedicated LFE channel however is impractical due to it requiring three channels.

During an online survey in 2014, it was asked if it would bother participants to have speakers within the arms of 3-D glasses; 70% of participants said no (Figure 8). This suggests that headphones may be a consideration for future film exhibition, even if used in a hybridised situation with an LFE track.

![Figure 8: Would it bother participants to have speakers within the arms of 3-D glasses](image)
Through practice-led research it became apparent that the accurate positioning of sound in any format is time consuming, as every sound needs to be positioned across all three axes and in some instances with automated movement. This highlights that it is often difficult to accurately provide positional data and have accurate positional rendering within z-space in the 5.1 and binaural formats. The BBC, DTS, Barco Auro and Dolby have committed to developing headphone technologies, which signifies that headphone technologies are a serious consideration for improving the immersive sound environment.

Works Cited
Soundworks Collection.