



HAL
open science

Emily's World: behind the scenes of a binaural synthesis production

David Poirier-Quinot, Brian F. G. Katz

► **To cite this version:**

David Poirier-Quinot, Brian F. G. Katz. Emily's World: behind the scenes of a binaural synthesis production. Audio Eng Soc Conv 148, May 2020, Vienna, Austria. hal-02869648

HAL Id: hal-02869648

<https://hal.science/hal-02869648v1>

Submitted on 16 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Audio Engineering Society

Convention e-Brief 611

Presented at the 148th Convention
2020 June 2–5, Online

This Engineering Brief was selected on the basis of a submitted synopsis. The author is solely responsible for its presentation, and the AES takes no responsibility for its contents. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Audio Engineering Society.

Emily's World: behind the scenes of a binaural synthesis production

David Poirier-Quinot ¹, Lucie Hardoin ², and Brian FG Katz ¹

¹ Sorbonne Université, CNRS, Institut Jean Le Rond d'Alembert, Lutheries - Acoustique - Musique, Paris, France

² Lucie Hardoin, www.sonbinaural.com, Finistère, France

Correspondence should be addressed to Brian FG Katz (brian.katz@sorbonne-universite.fr)

ABSTRACT

Three main workflows exist to create binaural content: binaural recording, binaural decoding of Ambisonic recordings, and binaural synthesis. Binaural synthesis, as an object-oriented approach, offers the highest flexibility throughout the creation process (composition, mixing, etc.). Conversely, it typically requires separate takes for each audio source and animations to construct the final spatial mix. “Binaural Steps - Emily's World” is an 8 min short radio-fiction commissioned to highlight the possibilities of binaural synthesis using studio and Foley recordings. Synthesis was achieved using the Anaglyph binaural audio engine, designed to facilitate the transition of spatial hearing research from laboratory to industry.

1 Introduction

There exists three main workflows to create binaural content: binaural recording, binaural decoding of Ambisonic recording, and binaural synthesis. As an object-oriented approach, binaural synthesis offers the greatest flexibility during all the stages of the creation process (composition, mixing, etc.). On the flip side, binaural synthesis often requires a separate take for each audio source in a scene as well as a non-negligible degree of animation work on the part of the sound engineer in constructing the final spatial audio mix.

To help promote the use of binaural synthesis, and to provide an overview of the steps involved, we present a content creation case study, using object-oriented binaural synthesis applied to studio and Foley recordings specifically made for the project. The study is based on “Binaural Steps - Emily's World” ^{1,2}, an 8 min short radio-fiction project

commissioned to highlight the possibilities of binaural synthesis.

This e-Brief highlights the various elements and steps involved in the creation of this short binaural audio piece, with input from the sound engineer, author, programmer, and director.

2 Production brief

The brief was formulated simply: “highlight the possibilities of binaural synthesis”. It was the decision of the producer to leave the artistic aspects open to the production team, leaving the artistic freedom in the hands of the content creators directly. Originally imagined as a series of separate situation clips, the production team then embarked on a series of propositions trying to incorporate a true story, a sound fiction in its own right. While one target audience is the sound engineering community, it was very much desired that the general listening public would also be attracted to the production, and would

¹ “Emily's World” : youtu.be/7Xs-Tqzd96o

² “Emily's World” : [soundcloud.com emilys-world](https://soundcloud.com/emilys-world)

listen to it in its entirety, thereby spreading the good word...

Given a rough production budget of 5k€, an audio-only production seemed more appropriate, with the addition of visual imagery reserved for a later release if desired. This reflects a similar binaural project carried out in 2014, with the production "Station 508"³, a student radio sci-fi suspense short. One version of this production included artistic still image renderings of key points in the story offered to help the audience along.

3 Script development

The project launched with the selection of the sound engineer, a specialist in native binaural recordings, to lead the production team. This project was an opportunity to see how a seasoned *native binaural* recording engineer could take that knowledge and apply it to a *binaural synthesis* production. Initially unsure if binaural synthesis would allow the same artistic freedom and the ability to make a totally immersive binaural creation, she jumped into the project with both feet. It is not often that the sound engineer gets to write the script.

To assist, an author was brought in by the engineer to collaborate with early phases of writing the script, starting from the various constraints in the brief.

The engineer wanted the story to start in mono and then slide towards synthetic binaural sound. This first point was considered important, so that the ear can compare and appreciate the change of space.

The budget only really allowed the hiring of one principal actress, and the story had to integrate this significant constraint. Any other parts were to be played by members of the production team. Budget was also allocated for sound effect recordings to construct the scenes.

Final Synopsis: While Emily, an actress, and Sam, a technician, are working in a recording studio, an argument breaks out. Emily leaves the building, keeping the live wireless tie-clip microphone that Sam has equipped her with still active. She then starts a journey through naturalistic yet fantastic soundscapes, trying to convince Sam of the richness of binaural perception, and that there is nothing more

beautiful than listening to the world as we naturally hear it....

The Journey: Emily takes the listener to various places and events in her memory, through the binaural recording she transmits via the wireless mic, giving the listener a first person perspective in these different scenes. The scenes are designed to experiment with various types of soundscapes: a busy environment, an outside space with distance, an enclosed space, and a setting exploring proximity and depth.

4 Recording approach

Unable to go and record all scenes from scratch, the project used some sounds and ambiences collected from recordings made by the sound engineer for personal projects and previous film shoots. The ones that inspired the most and for which quadraphonic recordings were in-hand: a big city (New-York), a beach (in Brittany, where the engineer resides), a forest, and the interior of an apartment.

Inside these soundscapes, the main character tells her memories, shares her impressions and transports the listener from one place to another via the notion of a linking sound object in each scene that connects the changing of space, changing from one scene to another: a cigarette that she lights in the streets of New-York to blowing smoke on a deserted beach, a church that rings on the beach and resounds in her apartment, the pouring of a cup of tea that transports us to the forest...

Three types of audio content were used in the production: studio recordings (for the dialog), Foley recordings (for the majority of events), and the above mentioned field recordings.

The Foley sequences, the most prevalent of the audio content, were recorded within the anechoic chamber of Sorbonne University. The Foley sound effect sequences included horse trotting, walking, leaves rustling, tea pouring, page turning, and cigarette smoking.

Due to availability of the chamber, some dialog and Foley recordings were made in acoustic experimental booths at the Institut Jean Le Rond d'Alembert, Sorbonne University.

³ "Station 508" : youtu.be/V6N-6DxAVQw



Figure 1. Foley and voice recording sessions: pouring tea, studio dialog, horse trot in the sand, and steps in the forest.

The jazz quartet musical extract came from the AVAD-VR⁴ anechoic audio-visual database [1], also recorded at Sorbonne University.

5 Binaural rendering engine

Binaural synthesis was achieved using the Anaglyph binaural audio engine⁵, a research based plugin conceived to facilitate the transition of spatial hearing research from laboratory to industry. Anaglyph features a personalizable morphological ITD (Interaural Time Difference) model, near-field ILD (Interaural Level Difference) and HRIR (Head Related Impulse Response) parallax corrections, a Localization Enhancer, an Externalization Booster, and SOFA HRIR file support [3]. Details of the workings of the rendering engine can be found in [2].

Anaglyph is being distributed as a VST (Virtual Studio Technology) plugin, compatible with most Digital Audio Workstations (DAW). Its primary function is the real-time rendering of a dry mono audio input buffer to a 2-channel binaural stereo audio output. A cross between a research platform and a professional solution, Anaglyph is conceived as a proof of concept test for state-of-the-art spatialization techniques to be compared against production environment expectations.

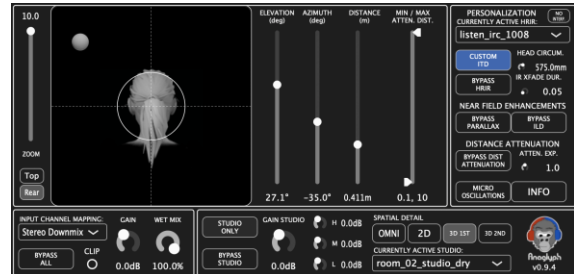


Figure 2. Anaglyph VST interface.

The latest version (0.9.4) introduces several new features compared to the previous reported release:

- Available as a Unity plugin
- Windows 32 bit support
- Optimised convolution engine, based on the JUCE dsp::Convolution class
- User defined HRIR crossfade duration, to control the trade-off between responsiveness and zipper noise free rendering.
- Threaded SOFA file loading (keeping the user interface responsive)

A complete release note is available on the Anaglyph website.

Anaglyph is used actively in the research lab for studies on spatial hearing and virtual reality [4,5].

6 Mixing

The Anaglyph VST offers a variety of controls for adjusting elements of the binaural rendering algorithms. Dynamic control of source position, including distance, can be automated via the DAW host.

The externalisation booster, which adds the simulated acoustics of a listening studio equipped with a higher-order Ambisonic 3D speaker array, can be modified with respect to its size/reverberation via presets and the EQ and level can be dynamically adjusted, offering tailored solutions for the needs of different scenes.

The audio mixing was carried out in 2 phases, due primarily to preferences of tool familiarity of the sound engineers and the fact that Anaglyph is not yet available for all platforms. The result, binaural

⁴ www.lam.jussieu.fr/Projets/AVAD-VR.html

⁵ anaglyph.dalembert.upmc.fr

rendering was carried out in Reaper, while final composition editing of the rendered binaural tracks was carried out using ProTools. This situation was not ideal, and is the motivation for additional investment in development of an AAX version.

7 Observations

Here are some first-person accounts from the production team concerning the project.

7.1 Sound Engineer / Author

The bulk of the work was the post-production: editing/mixing/spatialization. I had to train myself very briefly on Reaper, which I used to spatialize all the sound objects and then editing in ProTools. The back and forth was not very practical but there was not enough time to learn how to edit and mix efficiently in Reaper.

I first spatialized the ambiances of the 4 scenes, having the actress listen to them in order to give her context during the studio session. The voices were then added, followed by the sound effects related to her, such as footsteps, presence, and some props. Finally, I added the remaining sound effects.

Initially, the placement of the sound objects was not always easy and the rendering felt sometimes inconsistent. I realized, for example, as I got more familiar with Anaglyph that the externalization of the sources worked better by activating one of the 3 room effects proposed. However, the scenario had 3 outside scenes and only 1 inside scene, but it was too late to turn the script upside down...

I had a hard time trying to recreate artificially the head movements (varying voice directivity) of the actor when she talks to Sam, who represents the listener's listening point. While this realistic effect comes relatively free in native binaural recording, I did not really succeed in creating this perception of movement in the synthesis and finally abandoned the idea due to the time required.

I also found one of the known pitfalls of the binaural, the lack of stability for frontal sources. These points could still be improved in the tool.

One of the strong points of the plug-in is the possibility to integrate external HRIR files. I took advantage of it to mix/spatialize the sources with my own HRIR, measured at Orange Lab, Lannion.

Another strong point is the "micro oscillation" option, which puts the sound object in slight continual movement around the defined position, recreating natural minor head movements.

Finally, I wanted to take up this challenge and I am today rather satisfied with the result, even if some defects remain due to the lack of time given to the project. I enjoyed working on it.

7.2 Developer / Researcher

Feedback from the sound engineer helped us understand the strength and weaknesses of the plugin. Future versions of Anaglyph should first and foremost reliably work on every operating system and DAW commonly used in compositing workflows.

Some work is likely required regarding the graphic user interface, to provide various levels of detail on input parameters depending on the user's familiarity with the plugin. Finally, the studio reverberation module (used as an externalization booster) may be extended with more flexibility to better match the reverberation characteristics of the anechoic binaural synthesis to pre-recorded ambiances, resulting in a combined externalisation booster and parametric spatial reverberator, though it has not been the intention of the plugin to function as a 3D-reverberator. Similarly, design of externalization booster settings better suited to outdoor scenarios should also be explored.

7.3 Director / Producer

This experience was born from a desire to highlight to the community what the current state-of-the-art is in terms of binaural synthesis, to go beyond both the commercial hype and the general large industrial solutions that can often lag behind research developments.

This was the first time I have contracted a professional team to create a production, and it was a very enjoyable experience to take such a project outside of the academic realm, working with experts in their field and observing how they used the tools at hand. This was also my first experience with a professional Foley artist, and element that I very much wanted to explore in this project, offering the most flexibility and control over the creation sound scenes, limiting the use of recorded ambiances to predominantly background layers.

About a decade ago, the consensus of many discussions and roundtables with developers, researchers, and production houses was that while the tools for binaural audio were well advanced, their integration and use in content creation was lagging, with the industry failing to explore and exploit the technical possibilities. This was, in part, due to the limit of production workflow and computing time. Since then, there have been significant strides in bringing quality binaural to the mainstream, thanks in part to joint research-industry projects like BiLi⁶ and S3A⁷ with the active involvement of French and English television and radio, and actions by major industrial players such as Facebook and YouTube. Now, partially due to the readily available tools, 3D microphones, and VR devices providing easily accessible head-tracking, spatial audio content is significantly on the rise.

However, there is still much confusion for those non-specialists, with the “advent” of new spatial audio, like 8D, which are really a move backwards from the state-of-the-art. The conflation of binaural audio and psychological effects like ASMR⁸ and “binaural beats”⁹ continue to add confusion to consumers and engineers entering the field.

It is in my view only through the production of quality content will such confusions and false claims fade. In parallel, we continue to work on advancing rendering technology as we pursue extended research on human spatial hearing in complex situations.

While I was quite satisfied with the current production, the limitations, predominantly associated with the tight budget available for this production are now clearer to me. Such knowledge will surely influence the planning of our next production.

8 Conclusions

The production of “Binaural Steps - Emily’s World” is the first in a series of experimental binaural works used to highlight the possibilities of binaural audio and features of the Anaglyph engine as it develops. The release of “Emily’s World” coincides with the latest release update of the Anaglyph plugin version 0.9.4. The plugin is freely available for evaluation,

academic, and non-commercial use. A commercial licence is also available.

Acknowledgements

We would like to acknowledge the technology transfer agent *SATT Lutech* for contributing funding for this production. We would also like to acknowledge the following production contributors: Script - Tony Hayère, Voice talent - Juliette Aver, Sound effects, Foley artist - Eléonore Mallo, and Administrative Manager - Ronan Cabon (Tourne de Travers Production)

References

- [1] D. Thery and B.F.G. Katz, “Anechoic audio and 3D-video content database of small ensemble performances for virtual concerts,” in ICA Conf, 2019, pub.dega-akustik.de/ICA2019/data/articles/000471.pdf
- [2] D. Poirier-Quinot and B.F.G. Katz, “The Anaglyph binaural audio engine,” in Audio Eng Soc Conv 144, pp. EB431:1–4, 2018, www.aes.org/e-lib/browse.cfm?elib=19544.
- [3] P. Majdak, Y. Iwaya, T. Carpentier, R. Nicol, M. Parmentier, A. Roginska, Y. Suzuki, K. Watanabe, H. Wierstorf, H. Ziegelwanger, and M. Noisternig, “Spatially oriented format for acoustics: A data exchange format representing head-related transfer functions,” in Audio Eng Soc Conv 134, 2013, www.aes.org/e-lib/browse.cfm?elib=16781
- [4] D. Poirier-Quinot and B.F. Katz, “Assessing the impact of Head-Related Transfer Function individualization on task performance: Case of a virtual reality shooter game,” *J. Audio Eng. Soc.*, vol. 68, no. 4, pp. 248–260, 2020, doi:10.17743/jaes.2020.0004.
- [5] F. Zagala, M. Noisternig, and B.F. Katz, “Comparison of direct and indirect perceptual Head-Related Transfer Function selection methods,” *J Acoust Soc Am*, vol. 147, no. 5, pp. 3376–3389, 2020, doi:10.1121/10.0001183.

⁶ www.bili-project.org

⁷ www.s3a-spatialaudio.org

⁸ en.wikipedia.org/wiki/ASMR

⁹ [en.wikipedia.org/wiki/Beat_\(acoustics\)#Binaural_beats](https://en.wikipedia.org/wiki/Beat_(acoustics)#Binaural_beats)