A Morphological Analysis of Audio Objects and their Control Methods for 3D Audio

Justin Mathew, Stéphane Huot, Alan Blum

A Morphological Analysis of Audio Objects and their Control Methods for 3D Audio

Justin Mathew
Université Paris-Sud & CNRS
(LRI) – DMS Cinema
91405 Orsay Cedex France
justin.mathew@lri.fr

Stéphane Huot
Université Paris-Sud & CNRS
(LRI), Inria
91405 Orsay Cedex France
stephane.huot@lri.fr

Alan Blum
ENS Louis Lumière
93213 La Plaine Saint-Denis
Cedex France
al.blum@ens-louis-lumiere.fr

ABSTRACT

Recent technological improvements in audio reproduction systems increased the possibilities to spatialize sources in a listening environment. The spatialization of reproduced audio is highly dependent on the recording technique, the rendering method, and the loudspeaker configuration. While object-based audio production reduces this dependency on loudspeaker configurations, related authoring tools are still difficult to interact with. In this paper, we investigate the issues of spatialization techniques for object-based audio production and introduce the *Spatial Audio Design Spaces* framework (SpADS), which describes the spatial manipulation of object-based audio. Based on interviews with professional sound engineers, our morphological analysis clarifies the relationships between recording and rendering techniques that define audio objects for 3D speaker configurations, allowing the analysis and the design of advanced object-based controllers as well.

Keywords

Object-Based 3D Mixing, Design Space, Control Methods.

1. INTRODUCTION

Multi-dimensional audio recording and rendering techniques make it possible to capture sound scenes and to position sound sources in a realistic manner in 3D speakers configurations. However, content creators need to account for the diversity of speaker configuration methods in order to make the listener experience consistent across different audio spatialization formats. Technically, in current multi-channel audio production systems, achieving this spatial consistency requires unique audio mixes and output files for each speaker configuration. As an alternative, many higher-level object-based file formats have been introduced to reduce the complexity of creating several consistent audio mixes across different speaker systems [6, 9, 11]. Full support for object-based mixing is however still limited in current 3D audio authoring systems, thus reducing the possibilities of manipulating audio objects in 3D spaces and making complex 3D auditory designs difficult to achieve.

The goal of our study is to identify the limiting factors of current object-based 3D audio authoring techniques from a user's point of view. In fact, while many types of audio objects are well specified at the technical level [9, 13], they are vaguely defined at the user level. As a result, low-level channel-based production systems are still the norm. We believe that by better understanding object-based production outside of the constraining channel-based approach, we can clarify "what 3D audio objects are" at a higher-level, for users, and alleviate current limitations with new methods for designing appropriate controllers.

We first conducted informal interviews of sound engineers in order to get insights into the way they interact with audio objects, as well as the issues they encounter. We found that they interact with audio objects along three main stages of 3D audio production: *Recording/Rendering*, *Mixing*, and *Monitoring*.

1. The recording/rendering aspect focuses on how they capture audio content through different microphone techniques and how they are rendered to a speaker configuration, mostly relying on the channel-based approach;
2. The mixing aspect describes how they manipulate and control the recorded content in a 3D space;
3. Finally, monitoring is the method used to listen and analyze the mix.

From the relationships between these steps and the corresponding interaction methods, we define the **Sp**atial **A**udio **D**esign **S**paces (SpADS) conceptual framework that includes a morphological definition of audio objects as well as two design spaces for analyzing both the audio objects and their control methods. Beyond its descriptive power, we expect SpADS to help the design of new 3D object-based mixing methods that better suit the needs of professional sound engineers.

In the next section, we review related work before we discuss the findings from our interviews of sound engineers. We describe our morphological analysis of current 3D audio techniques and define audio objects. We then introduce the *SpADS framework* made of two design spaces: *Audio Object Design Space* and *Audio Object Controller Design Space*. Finally, we discuss the use of SpADS for analyzing an existing system before to conclude.

2. RELATED WORK

Object-based audio mixing came into focus with the growth of multi-dimensional diffusion systems. Very few commercial systems integrate support for this approach, which is still mostly built on top of traditional channel-based solutions (e.g. Auro 11.1[1] and Dolby Atmos[2]). As for research, audio objects were mainly investigated from a descriptive point of view (i.e. file format) [3, 9, 11, 20] rather than on how to interact with them. There are however some notable exceptions.

2.1 New Mixing Interfaces

Geier et al. proposed the SoundScape Renderer authoring method [10] as a modular rendering system for localized sound sources, with its associated file format: the Audio Scene Description Format [11]. While they compared the capabilities of their format to

[1] http://www.barco.com/en/auro11-1

[2] http://www.dolby.com/us/en/professional/technology/cinema/dolby-atmos-creators.html

J. Mathew, S. Huot, A. Blum.
A Morphological Analysis of Audio Objects
and their Control Methods for 3D Audio.
In NIME'14 : Proceedings of the International Conference on New
Interfaces for Musical Expression, May 2014.
Authors Version

others, they did not report on any study of the interface and interaction techniques. Jang et al. [14] introduced a method to interact with multiple types of audio objects through geometrical features to represent sound sources (i.e. point, line, plane, or volume). They also proposed several examples of audio-visual applications, but did not conduct any evaluation of these techniques.

2.2 Alternative Input Methods

Alternative input methods such as multi-touch or haptic feedback have also been investigated. Carrascal and Jordà introduced a multi-touch mixing interface [5] that they compared with a standard mixing console to achieve a predefined mixing task. Their results showed that non-expert participants completed the tasks faster with their design, suggesting that it eases the accessibility of object-based mixing to casual users. However, there is no evidence that it would be beneficial to expert users in a professional working environment. Melchior et al. compared the use of a haptic feedback mixing technique against a standard mouse in a qualitative study with experts [18]. The tasks consisted of controlling the 3D position, translation and trajectory of a sound source. While the haptic device was preferred by the participants, the mouse was assessed as being better for predictability and manageability. Gelineck et al. also investigated the use of smart tangible objects as controllers on a 2D surface [12]. The results of a qualitative study showed that tangibles were preferred to typical controllers, but the participants felt the system did not scale in terms of functionality.

Overall, several new object-based mixing interfaces were investigated but focused on very specific and low-level mixing tasks, without a thorough analysis of 3D audio production as a whole. A notable exception is the study conducted by Peters et al., which consisted in an on-line questionnaire about the use of tools for sound spatialization [21]. They identified issues in existing mixing tools which, as we will see in more detail later, are in-line with our analyses and also ground our study.

2.3 HCI Methodologies

Our work also relates to the adaptation of Human-Computer Interaction methodologies for the design and evaluation of new musical instruments [17, 22]. Our approach was inspired by Wanderley and Orio's use of HCI tools for evaluating input devices for musical expression [24]. They analyzed common contexts found in scenarios of interactive computer music to investigate the use of input devices from a higher level of user interaction. While we are not focusing on the formal evaluation of 3D audio production systems, such methodologies can also help to better characterize these systems and to inform their design. In particular, we adopted Card et. al.'s morphological analysis approach [4] in order to better define audio objects and characterize the related control methods within our SpADS framework. Beyond its descriptive power, our objective is also to provide designers with a tool with generative power – i.e. "*the ability to help designers create new designs*" [2] – for the exploration of new 3D object-based mixing methods.

3. INTERVIEWS WITH PROFESSIONAL SOUND ENGINEERS

Previous studies highlighted general issues with tools for spatialization, and especially the use of low-level channel-based methods to manipulate the higher-level concept of audio object. As reported in Peters et al.'s study [21], the bus architecture in Digital Audio Workstations is limiting and requires "[to develop] *input devices that are tailored to the specific needs of controlling spatialization*" [21]. Similar concerns were raised at the FISM 2013 conference where invited speakers from a round table discussion about 3D audio in cinema and broadcasting highlighted the need to capture 3D content in a correct manner, but also that there is no convenient data structure or tool for such content [19].

These general observations provided us with directions into the areas of current production methods to focus on in our study. To gather more details about the actual use of audio objects, we conducted informal interviews with two experienced sound engineers, *E1* and *E2*, at their place of work. E1 works at a concert venue and has more than 10 years experience in 3D audio and acoustics. We also observed one of his students working on a project. E2 works in a radio station and has been working in spatial audio for over 5 years, but started working specifically on 3D audio within the past year. We asked them what type of projects they were currently working on and what tools they use. Our goal was also to observe them in situ and to assess what parts of channel-based systems are difficult to use when working with 3D audio techniques. The key points extracted from these discussions highlighted details that were missing in previous work to ground our analysis of object-based tools in channel-based systems.

E1 introduced us to his work in a concert venue that broadcasts live shows with multi-channel and binaural playback. In their current projects, they have been experimenting with several types of *High Order Ambisonic* (HOA) microphones and comparing their use while mixing for multiple speaker configurations. They use custom-built plug-ins within a traditional DAW to properly decode the recordings for comparison. E1 presented this work with a student's project where the recordings were decoded and mixed into different buses to be played on four different loudspeaker configurations. Individual mixes are done in software with a standard mouse and keyboard setup, and a MIDI controller to switch between speaker configurations for comparison.

E2 discussed the same type of mixing setups for multiple outputs when he works on broadcasting shows in multi-channel and binaural playback. He uses a traditional DAW with specific plug-ins and a standard channel mixer. This engineer only focused on the mixing process, but said that he works closely with recording engineers who provide him with detailed notes on how the audio was recorded. One interesting observation he made is that there is "*no relationship of audio objects with output file-formats*", highlighting the inconsistent mapping between the recording methods and output file-formats that are specific to speaker configurations.

From these answers and previous work, summarized in Figure 1, we identified common issues in 3D audio production: a need to understand how the audio was recorded and will be rendered, the lack of mixing controllers that are tailored to specific microphone techniques, and a complicated bussing system to monitor. These issues and needs are directly related to well-identified stages of audio production: *Recording/Rendering*, *Mixing*, and *Monitoring*.

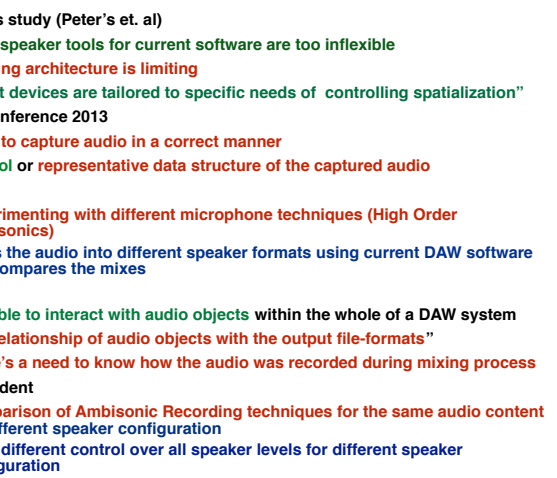

Figure 1: The major issues of 3D object-based production in channel based systems, from related work and our interviews. (green for *Mixing*, red for *Recording/Rendering*, blue for *Monitoring*)

3.1 Rendering/Recording

Understanding how the audio was recorded, and thus understanding "what is the audio object", greatly influences how it will be mixed. However, characterizations of audio objects are too vague because of their numerous but limited definitions, as well as the variety of recordings and rendering methods. This was especially obvious in E1's project where different types of HOA recordings are used in the channel-based system. Each recording needs to be considered individually and independently in the bussing architecture to be rendered correctly. Having clear details on the recorded objects was also pointed out by E2 as a crucial part of the process.

3.2 Mixing

Mixing is done through plug-ins in DAWs that are typically controlled by a mouse (and sometimes standard faders and knobs). In this channel-based approach, each mix is specific to a loudspeaker configuration and the user has to restart the mixing process each time he has to consider another type of output. We observed this constraint in the need to decode the HOA recordings several times for each speaker configuration in E1's student's project.

3.3 Monitoring

Mixing for different speaker configurations is a common concern for both E1 and E2, and this obviously implies that monitoring the mix for each speaker configuration is important. For each speaker configuration, outputs of the channel-based system must be accurately mapped to the speakers and need to be re-mapped when changing among systems. Consistent knowledge of each speaker and its relative position in each configuration is needed for the engineers to correctly create bussing paths and to switch among them when monitoring.

These interviews revealed some previously unidentified issues in 3D audio manipulation techniques within channel-based systems that cover the three stages of audio production. Analyzing their relationships from an object-based point of view shows how the *Recording/Rendering* stage dictates the spatial capabilities of an audio object, that are in turn controlled through the *Mixing* stage. The *Monitoring* stage provides information about the speaker locations that influences the spatial capabilities of audio objects when the speaker configuration changes. Figure 2 summarizes how audio objects are related to each stage, as well as their relationships. Additionally, the variety of techniques in the *Recording/Rendering* stage leads to question what audio objects are, making it difficult to assess which control methods are appropriate. In the next section, we address this issue of clearly defining audio objects by conducting a morphological analysis of current production tools for 3D audio in the *Recording/Rendering* and *Mixing* stages from an object-based point of view.

4. A MORPHOLOGICAL ANALYSIS OF AUDIO OBJECTS

As we already pointed out, spatial audio recordings highly depend on the recording techniques that were used and will thus be perceived differently by the listener. Sound engineers must be aware of that fact when they spatially mix these recordings in order to have a consistent perception of their spatial properties over different rendering systems. Audio objects could help to bridge this gap between recording and rendering methods by providing a higher-level and consistent abstraction, which the engineer would manipulate independently of the technologies that are used upstream and downstream (see Figure 2). It would require however a better definition of these audio objects and of their possible control methods due to the variety of current techniques. To this end, we conducted a morphological analysis of the different tools in the *Recording/Rendering* stage and how they are controlled in the *Mixing* stage. By analyzing the functionalities of these tools, our

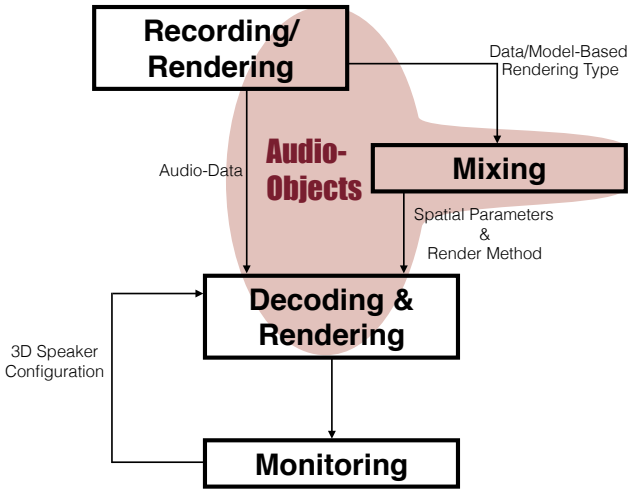

Figure 2: Block diagram showing the relationships between three stages of object-based production (Recording/Rendering, Mixing, Monitoring) .

goal is to define the relationships between audio data and rendering methods, thus defining audio objects, their corresponding manipulable properties and their control methods.

4.1 Defining Audio Objects

To connect recording techniques with rendering methods, Geier et al. distinguished between *Data-Based* and *Model-Based* audio objects [11]. Data-Based rendered objects are audio captured with a microphone technique that encodes spatial information into the audio data. In model-based rendering, the recording technique is modeled as virtual sound sources within the scene. This higher-level representation is unlike many current systems whose audio objects' properties are closer to the technical aspects of the system.

4.1.1 Data-Based

One method to capture an acoustic space is to place microphones in the same configuration as the speaker playback configuration, resulting in a one-to-one mapping between microphones and speakers that we call *Direct Rendering*. This method is commonly used with microphone arrays to recreate the sound field that is traveling based on Huygens' Principle [1]. However, it is inflexible to other speaker configurations and requires many microphones to achieve high spatial resolution. A typical example is the *Holophone H3-D* microphone, which captures a 2D space that is directly rendered to a 5.1 Surround Sound speaker format [23]. Conversely, the *Spatially Encoded* method uses the directivity and orientation of multiple microphone capsules to encode a space within the audio data. Decoding audio data with different methods allows the captured space to be rendered on any configuration. The *Double M/S* microphone for instance is a spatially encoded technique that can be decoded to Surround Sound or B-format for rendering on any speaker configuration. This technique enables the control of the characteristics of the microphone's polar patterns, giving users the ability to rotate the captured sound scene [25]. Another common spatially encoded technique is 1st order Ambisonic recording, which captures a 3D sound scene that is decoded to B-format for rendering on any speaker configuration [11]. The Ambisonic Studio plug-in allows manipulation of the spatial parameters of Ambisonic type recordings and renderings, such as orientation and directivity [7]. From a user's point of view, these types of recordings need to be treated collectively as one unit, to ensure spatial accuracy of the captured scene. Any type of manipulation must be done with knowledge of how it will affect the scene during playback.

4.1.2 Model-Based

Model-Based rendering methods include *VBAP*, *WFS*, and *Ambisonic Panning* [1, 11, 10]. Typical recordings with these techniques are mono, as in in the *Dolby Atmos Authoring* system [9]. Iosono's *Spatial Audio WorkStation 2* plug-in takes a different approach with two basic types of audio objects: Channel Objects and Event Objects. Channel Objects are "*sources of audio that may contain one or more sounds mixed together [...] coming from one location that may be moved around*", and Event Objects " *are singular sounds [...] with specific start and end times in certain location*" [13]. From an objective point of view, the audio object in these tools is a mono audio signal and spatial manipulations are only 3D panning capabilities. This is limiting when considering the palette of possible spatial properties. An exception is the IRCAM *Spatialisateur* [15], which considers audio objects as localized sound sources as well, but defines their spatial parameters from their interaction within a room, thus allowing more spatial manipulations such as orientation. However, Model-Based rendered audio does not have to be restricted to just mono recordings. Linear or geometrical microphone arrays could be rendered this way, each microphone being a point in a line or shape, creating the same contexts that Jang et al. referred to [14]. While these audio objects are more flexible in terms of the type of rendering to use, the spatial capabilities of most tools are still restricted to 3D positioning only.

4.1.3 Spatial Parameters

Overall, model and data-based software only support subsets of all the possible spatial parameters, which we define as *Translation*, *Orientation*, and *Area/Volume*. Translation (*1D*, *2D*, and *3D*) refers to how many dimensions an audio object can move in. The Orientation parameters (*Pitch*, *Yaw*, and *Roll*) refer to the direction of orientation of sound sources or sound scenes, depending on the type of the object. Area/Volume parameters describe the shapes an audio object can have, with *Width*, *Height*, and *Depth*. These parameters are however perceived and thus manipulated differently by sound engineers, depending on whether they use a data-based or model-based approach in the *Monitoring* stage. Common data-based audio objects immerse listeners in the center of a sound scene, whereas typical model-based audio objects are localized as spatial sound sources [16]. This difference in perception creates different contexts for manipulating the spatial parameters, which are thus dependent on the type of audio object.

4.2 Spatial Audio Design Spaces (SpADS)

From this analysis of the nature of audio objects and their spatial parameters in the *Recording/Rendering* and *Mixing* stages, we define the dimensions of two design spaces for SpADS: *Design Space of Audio Objects* (SpADS-A) and *Design Space of Control Methods* (SpADS-C).

4.2.1 SpADS-A (Audio Objects)

SpADS-A defines and classifies audio objects as audio data along two dimensions: their *Rendering* method and their *Spatial Parameters*. Various audio objects are then made of different combinations of audio data and rendering methods. As presented in Figure 3, SpADS-A helps gain a better understanding of how audio objects of similar or different natures compare to each other.

SpADS-A allows quick identification of how recording techniques can be rendered, and the corresponding 3D mixing capabilities with spatial parameters. For example, a mono audio signal is only used in Model-Based renderers, which can be translated up to three dimensions and oriented in three directions. The Stereo A/B recording is considered to be two point sources that are linked, and differs slightly from mono recording. Stereo recording is a two point source that can be translated together, oriented from an anchor point, and increased/decreased in width between the two

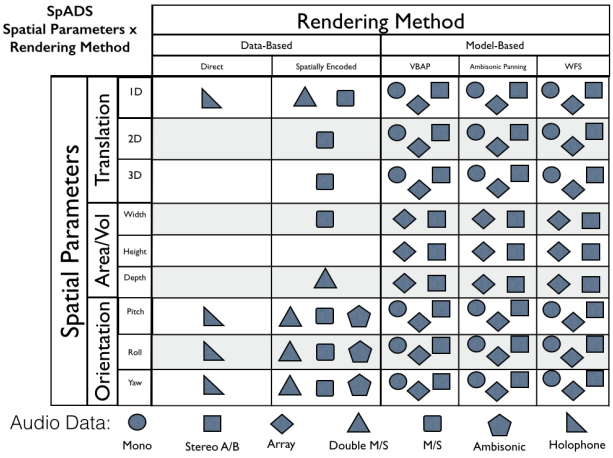

Figure 3: An example of an analysis comparing different types of audio objects using SpADS-A.

points. Increasing the number of point sources creates microphone arrays, adding more capabilities in the spatial manipulations. It seems that only model-based audio objects provide this ability, but some encoded data-based renderers can provide the same spatial capabilities. For instance, the M/S recording technique is a two-channel microphone technique that is decoded into stereo, allowing it to have the same parameters as the Stereo A/B method. However, the need to decode the M/S technique categorizes it as a data-based object rather than a model-based object [25]. Double M/S is an improvement over the M/S technique that captures audio with 2D spatial qualities [25], and can only be translated in one dimension, oriented, and area manipulated. This increase of dimensionality provides different spatial capabilities from the perspective of the listener, which is situated at the center of the recreated area. Ambisonic recordings capture a 3D space and locate the listener similarly but can only be oriented and not translated.

Overall, SpADS-A shows how audio objects can be rendered and manipulated, highlighting similarities and differences among them. Now, we can use these spatial parameters to explore how different input devices can be used to manipulate these objects during the *Mixing* stage.

4.2.2 SpADS-C (Controllers)

Our study of controllers for audio objects is inspired by the use of traditional mixing consoles in the three stages of production. In this context, the audio data used in the *Recording/Rendering* stage is a mono signal, and the typical *Monitoring* speaker configuration is stereo. The resulting audio object is rendered with left/right equal power panning that positions it on the line between the speakers. A simple one-dimensional potentiometer is normally used as positioning input device, and its physical position gives visual feedback of the audio object's location. Additionally, multiple potentiometers can be controlled at once with two hands. Our *SpADS-C* design space extends this analysis to 3D speaker configurations and classifies input devices according to their capabilities for controlling the spatial parameters of audio objects (see Figure 4). The *Input* and *Visual Feedback* dimensions respectively describe how the user controls the device and the produced visual feedback. Input can be *Single* or *Multiple*, whether the user can control one or several controllers at once, and have defined degrees of freedom (i.e. how many parameters each input can control). The Visual Feedback[3] can be provided by the *Physical* position of the controller or a *Virtual* representation within the software GUI, which can have several dimensions (e.g. 2D vs 3D graphics).

[3]While promising for 3D audio object manipulation, we do not consider advanced feedback technology such as force-feedback since visual feedback is still the norm in professional audio production.

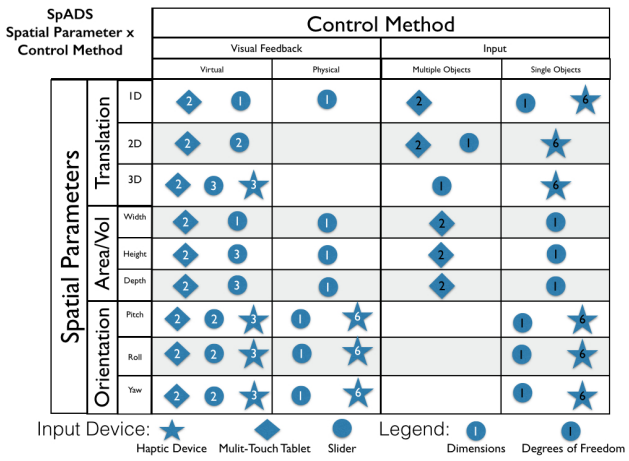

Figure 4: An example of SpADS-C with some input devices for controlling spatial parameters with different visual feedback.

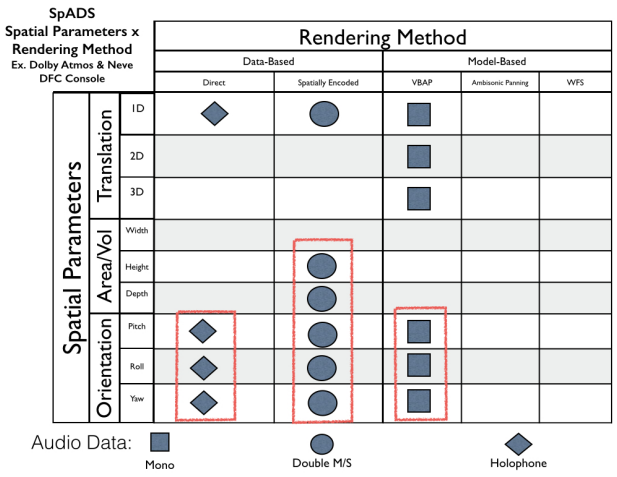

Figure 5: SpADS-A analysis of possible audio objects that can be used in the Dolby Atmos with the Neve DFC console.

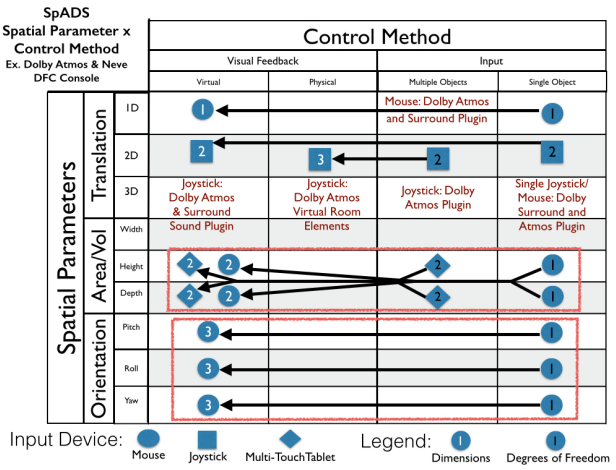

Figure 6: SpADS-C analysis for Dolby Atmos with the Neve DFC console which highlights missing controls for audio objects parameters.

As presented in Figure 4, SpADS-C helps to compare how some input devices can be used to control spatial parameters. For example, a Haptic Device with 6 degrees of freedom can provide integral control of multiple parameters, but can only control one audio object at a time. However, a multi-touch tablet can provide multiple inputs, but has only two degrees of freedom for each input. Also, virtual feedback can be coupled with different kinds of controllers, but only some of them provide physical feedback: In a typical mixing console for instance, faders and knobs provide physical visualizations augmented with level meters that can be considered as virtual feedback. For input devices without physical feedback, such as a tablet, virtual visualization is often mandatory.

SpADS-C can categorize and compare different types of controllers but does not help to assess which are better suited for the manipulation of specific audio objects and parameters. Such assessment will require further investigation and evaluation. However, this high-level design space of controllers already provides valuable insights into the matching between the spatial capabilities of audio objects and the functionalities of current systems.

5. DISCUSSION

The SpADS framework is a tool to explore the spatial capabilities of audio objects and to compare and match controllers to their spatial parameters. As an example, we used SpADS to analyze the *Neve DFC* console for the authoring of *Dolby Atmos* content [8]. We considered only the use of mono and surround sound recording techniques and subsequent rendered audio objects that we placed in SpADS-A and SpADS-C (see Figures 5 & 6).

We analyzed three examples of recording and rendering techniques that can create Surround Sound objects: *Double M/S*, *Holophone*, and *mono audio recordings*. We assumed that the mono audio recordings are rendered with *VBAP*, but another rendering technique may be used. For the Holophone recording, the audio is mapped out directly to the Surround Sound object through channel configuration, but the *Dolby Surround Sound* [9] or Schoeps' *Double M/S* plugin tools [25] could be used as well. Both plug-ins can be controlled with a mouse, but the Dolby Surround Sound plug-in can position mono audio objects with the joystick in the Neve DFC Console (see Figure 6). Feedback is always virtual within the plug-in, but the joystick also provides an additional virtual 2D grid. It also gives physical visual feedback, but only during interaction, and it has to be re-aligned when changing between controlled audio objects [8]. Finally, the Dolby Atmos plug-in must be used to manipulate 3D localized mono audio objects. They can be controlled with the mouse or the two joysticks on the console, providing the user with both physical and virtual feedback [8].

The SpADS-A analysis highlights that some spatial properties cannot be controlled with this production system (boxed out in red in Figure 5). Once transposed to SpADS-C (boxed out in red in Figure 6), it gives the opportunity to explore and increase the functionality of the system in terms of spatial mixing capabilities. For instance, a mouse with simple GUI controls could increase the missing spatial capabilities, or a tablet could be used to control the Area/Volume parameters for the Surround Sound object. These examples illustrate how SpADS can help to describe existing systems for spatial mixing from an audio object based point of view, and potentially improve them. Beyond its descriptive and generative powers, we believe that SpADS can be extended to support the evaluation and comparison of control methods and input devices for the manipulation of audio objects [24, 4].

At a higher level, SpADS addresses the issue pointed out by E2 in our preliminary study: there was "*no relationship of audio objects with output file-formats*". SpADS' object-based approach clearly defines the necessary relationship between the recording and rendering techniques that define audio objects in a user-centered approach. Using this relationship, designers can explore control

methods for the spatial parameters of an audio object or a combination of audio objects, and design new authoring systems that can render and mix a variety of audio objects together for 3D audio.

6. CONCLUSION

The development of multi-channel mixing has provided sound engineers with the ability to spatially manipulate audio in 3D, but from our interviews of content creators, we highlighted several issues of current object-based production systems. The vague definition of audio objects, the lack of well-suited controllers, and the need to listen to multiple mixes led us to define the three stages of 3D audio production to focus on: *Recording/Rendering*, *Mixing*, and *Monitoring*. Our morphological analysis of audio objects allowed us to better define them as well as their spatial capabilities at a higher-level. The resulting **Sp**atial **A**udio **D**esign **S**paces (SpADS) conceptual framework introduces two design spaces for analyzing the spatial parameters of audio objects (SpADS-A) and the spatial parameters that an input device can control (SpADS-C).

This initial proposal of SpADS focuses on the *Recording / Rendering* and *Mixing* stages in audio production, and does not yet include *Monitoring*. Extending SpADS to account for the *Monitoring* stage would enable the analysis of complete audio production systems. We also plan to conduct experiments to better analyze the possible matchings between the spatial properties of audio objects and several types of controllers/input devices. This will complement the descriptive power of SpADS with comparative and generative capabilities [2], and ultimately make it possible to explore and evaluate new audio production software focused on the flexibility of fully object-based *Recording/Rendering*, *Mixing*, and *Monitoring*.

7. ACKNOWLEDGEMENTS

We thank the sound engineers who participated in our study for their time and fruitful discussions. Thanks to Halla Olafsdottir, Joseph Malloch and Michel Beaudouin-Lafon for helpful comments on this paper. This research is partially supported by the French Agency for Technical Research (ANRT) and DMS-Cinema.

8. REFERENCES

[1] M. Baalman and D. Plewe. WONDER - a software interface for the application of Wave Field Synthesis in electronic music and interactive sound installations. In *Proc. of the International Computer Music Association*, 2004.

[2] M. Beaudouin-Lafon. Designing interaction, not interfaces. In *Proc. of the Working Conference on Advanced Visual Interfaces*, AVI '04, pages 15–22. ACM, 2004.

[3] J. Breebaart, J. Engdegård, C. Falch, O. Hellmuth, J. Hilpert, A. Hoelzer, J. Koppens, W. Oomen, B. Resch, E. Schuijers, and L. Terentiev. Spatial Audio Object Coding (SAOC) - the upcoming MPEG Standard on Parametric Object Based Audio Coding. In *Proc. of the 124th Audio Engineering Society Convention*, 2008.

[4] S. K. Card, J. D. Mackinlay, and G. G. Robertson. A morphological analysis of the design space of input devices. *ACM Trans. Inf. Syst.*, 9(2):99–122, 1991.

[5] J. Carrascal and S. Jordà. Multitouch interface for audio mixing. In *Proc. of the International Conference on New Interfaces for Musical Expression*, NIME '11, 2011.

[6] B. Claypool, W. Van Baelen, and B. Van Daele. Auro 11.1 versus object-based sound in 3d. Technical report.

[7] D. Coureville. Ambisonic Studio B2X Plug-in Suite for Mac OSX, 2013.

[8] D. Critchley. *Working with the Dolby Atmos format on the AMS-NEVE DFC console*. AMS-NEVE.

[9] I. Dolby Laboratories. *Authoring for Dolby Atmos Cinema Sound Manual*. Dolby Laboratories Licensing Corporation, 2013.

[10] M. Geier, J. Ahrens, and S. Spors. The SoundScape Renderer: A Unified Spatial Audio Reproduction Framework for Arbitrary Rendering Methods. In *Proc. of the 124th Audio Engineering Society Convention*, 2008.

[11] M. Geier, J. Ahrens, and S. Spors. Object-based audio reproduction and the audio scene description format. *Organised Sound*, 15:219–227, 2010.

[12] S. Gelineck, D. Overholt, M. Büchert, and J. Andersen. Towards an interface for music mixing based on smart tangibles and multitouch. In *Proc. of the International Conference on New Interfaces for Musical Expression*, NIME '13. ACM, 2013.

[13] IOSONO. Spatial audio workstation 2 operation manual.

[14] D. Y. Jang, J. Seo, K. Kang, and H. K. Jung. Object-based 3d audio scene representation. In *Proc. of the 115th Audio Engineering Society Convention*, 2003.

[15] J. M. Jot. *Spatialisateur Ircam/Espaces Nouveaux User Manual*. IRCAM, 2012. updated by Rama Gottfried.

[16] J. M. Jot, V. Larcher, and J. M. Pernaux. A comparative study of 3-d audio encoding and rendering techniques. In *Proc. of 16th AES International Conference: Spatial Sound Reproduction*, 1999.

[17] C. Kiefer, N. Collins, and G. Fitzpatrick. HCI Methodology For Evaluating Musical Controllers : A Case Study. In *Proc. of the International Conference on New Interfaces for Musical Expression*, NIME '08, pages 87–90, 2008.

[18] F. Melchior, C. Pike, M. Brooks, and S. Grace. On the use of a haptic feedback device for sound source control in spatial audio systems. In *Proc. of the 134th Audio Engineering Society Convention*, 2013.

[19] R. Nicol, E. Corteel, F. Camerer, H. Déjardin, W. Bleisteiner, and A. Churnside. Round table discussion: 3D sound and its application. 16th Forum International Du Son Multicanal, 2013.

[20] N. Peters, S. Ferguson, and S. McAdams. Towards a spatial sound description interchange format (SPATDIF). *Canadian Acoustics*, 35(3), 2007.

[21] N. Peters, G. Marentakis, and S. McAdams. Current technologies and compositional practices for spatialization: A qualitative and quantitative analysis. *Computer Music Journal*, 35(1):10–27, Spring 2011.

[22] D. Stowell, A. Robertson, N. Bryan-Kinns, and M. D. Plumbley. Evaluation of live human-computer music-making: Quantitative and qualitative approaches. *Int. J. Hum.-Comput. Stud.*, 67(11):960–975, 2009.

[23] H. M. Systems. *Holophone H3-D Surround Sound Microphone User Guide*. Rising Sun Productions Ltd.

[24] M. M. Wanderley and N. Orio. Evaluation of input devices for musical expression: Borrowing tools from hci. *Comput. Music J.*, 26(3):62–76, 2002.

[25] H. Wittek, C. Haut, and D. Keinath. Double m/s - a surround recording technique put to test. Technical report, 2010.