Spatial Audio Mixing in Virtual Reality

Bargum, Anders Riddersholm; Kristjánsson, Oddur Ingi ; Mosen, Simon Rostami; Serafin, Stefania

Published in:
Proceedings of the 19th Sound and Music Computing Conference

DOI (link to publication from Publisher):
10.5281/zenodo.6797618

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Spatial Audio Mixing in Virtual Reality

Anders Bargum, Oddur Ingi Kristjánsson, Simon Rostami Mosen and Stefania Serafin
Aalborg University Copenhagen
[abargu17, okrist17, srosta17]@student.aau.dk, sts@create.aau.dk

ABSTRACT

The development of Virtual Reality (VR) systems and multimodal simulations presents possibilities in spatial music mixing, be it in virtual spaces, for orchestral compositions, or for surround sound in film and music. In this paper, we present design aspects for mixing audio in VR. By the use of interaction design principles and an examination of related research, we create a framework from which a virtual spatial-audio mixing tool is implemented and coupled with a digital audio workstation (DAW). The tool is tested against a similar computer version to examine whether the sensory benefits and palpable spatial proportions of VR can improve the process of mixing 3D and binaural sound.

1. INTRODUCTION

While there has been a lot of investigation in the world of designing virtual reality (VR) applications for computer music, especially within the field of Virtual Music Instruments (VMIs) and New Interfaces for Musical Expression (NIME) [1], little focus has been on new ways of graphically representing mixing, mastering, and audio effects processing. Traditionally, music mixing consoles are divided into functional sections, which constitute different metaphors in a stereo context [2]. As an example, each track has a channel strip, used as the main way to adjust the volume, with slide potentiometers seen as a universal metaphor for amplitude. The panning potentiometer represents a track's placement and spatial position, mapping the left and right position of a knob to the left and right location of a sound. These universal metaphors have been the standard way of representing sound sources in a stereo field and have, thus, been implemented in music mixing interfaces in different Digital Audio Workstations (DAWs). However, their figurative representations are harder to map to 3D sound and one could, therefore, look at non-traditional paradigms like the 'stage metaphor' when representing spatial audio. In the stage metaphor, also called a 'virtual mixer' [3], the level and stereo position (and possibly other parameters) are modified using the position of a movable icon on a 2D or 3D image of a stage [4]. As seen in figure 1, each source in the 'stage metaphor' is graphically visualised in space. It, thus, affects and utilises human visual localisation cues such as the ventriloquism effect, which makes a person perceive the sound coming from the location determined by the human visual system [5]. Based on the idea of the 'virtual mixer' paradigm, this paper presents a design framework for visually representing spatial and binaural music mixing using the perceptual, visual, and spatial dimensions of VR. The paper will examine how VR can facilitate sonic interaction and if VR's inclusion of multidimensional space and free rotation/movement, can enable a composer to visually place, move and mix sound sources intuitively, in a 3D space. We implement a virtual environment (VE) that can be linked with the DAW 'Ableton Live' allowing musicians, producers etc. to intuitively, effectively and accurately sketch ideas for spatial mixing. The aim of this study thus is to create an immersive and creative environment that produces a set of spatial coordinates and objects, which can be transferred to a traditional DAW. The environment is tested against a similar computer version to investigate if the process of mixing 3D sound in VR actually can be an improvement to the producer, whereas the VE as a concept is evaluated using a focus group of experts.

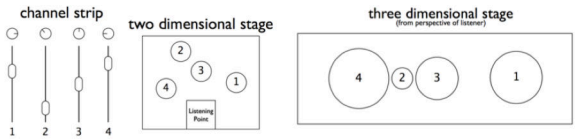

Figure 1. *Channel Strip vs. Stage Metaphor.* [6]

2. SONIC INTERACTION IN VR

There are multiple aspects to consider when designing virtual spaces and applications in VR, such as ensuring smooth interaction through minimum latency, preventing cybersickness, and facilitating presence. To fulfill smooth and successful interaction in computer music interfaces, Wang suggests that the system among other things should [7]:

- Be real-time if possible.
- Design sound and graphics in tandem and seek salient mappings.
- Hide technology and focus on substance.
- Introduce arbitrary constraints.

This means that interaction with sound sources should be easy, quick, streamlined, and noticeable. Virtual objects need to match the location and motion of auditory objects, and the user should not be confronted with technology or implementation to increase excitement and interest. These aspects can additionally be supported using feedback and various studies state that especially haptic and tactile feedback allows a user to develop musical skills and understanding of controls [8]. Gelineck et al. as an example investigate the inclusion of external controls that allow for touch or vibrational feedback, by comparing the stage metaphor (in form of an iPad app visualising a stage) to the channel strip metaphor (normal faders and panning) when completing a stereo mix [9]. While the study does not find any significant difference in terms of performance between the two cases, the iPad application was preferred by the participants user experience-wise due to its "intuitiveness", "enjoyability" and its "ability to reveal the spaciousness of the mix" [9]. In terms of the effect of aesthetics, when interacting sonically in VR, Wang proposes different principles for graphical and visual representation of objects and environments in a 3D world [7]:

- Simplify: Identify core elements, trim the rest.
- Animate, create smoothness, imply motion: It is not just about how things look, but how they move.
- Be whimsical, organic: glow, flow, pulsate, breathe and imbue visual elements with personality.
- Aesthetic: have one; never be satisfied with 'functional'.

Gale et al. additionally suggest that one should avoid visual clutter, meaning too many objects potentially overlapping and/or occluding each other on the screen [10]. As a part of object cluster, Serafin et al. state that general control additionally can be enhanced by visually representing the player's body [8]. People cannot see their own body in VR and the frustration/confusion this may cause can be overcome by generating a visual substitution of a person's real body seen from first-person perspective [8]. All of these topics, be it feedback, real-time interaction, animation, or the 'virtual body ownership' elicited by representing a virtual body, can be facilitated by a VR system and there is no doubt that a VE thus has the potential of fulfilling successful sonic interaction.

Looking at the auditory facets of sonic interaction in VR and VEs, a main attribute is the space/room in which the sound is played. The sound itself will in different physical spaces be shaped by the room's spectral characteristics and modified by properties of the room. The perception of room acoustics is highly important for both the feeling of presence but also to elicit localisation capabilities and out-of-head localisations from a potential user [11]. One can choose different methods when employing models of spatialisation, acoustics, and reverberation to virtual rooms. Robert Hamilton distinguishes between two main models: the user-centric perspective and the space-centric perspective [12]. In the user-centric perspective, the sound will be manipulated from a 'first-person' point of view, where sounds in the virtual world will correspond to the real-world based model of hearing: they will be placed in a general aural spectrum known from every day, with corresponding depth cues implemented through filtering and delay components. This can be done by tracking the coordinate distance between event locations and the user's in-game avatar and matching given head-related transfer functions (HRTFs), or similar spatialisation algorithms, to the position of the sound [12]. The space-centric perspective, on the other hand, shifts the focus to the sound itself correlated between the virtual and physical world. In this model, sounds are no longer contextualised based on their proximity and relationship to a given user [12]. Instead, they are processed concerning both the virtual and physical world, meaning the placement in each environment will affect it. A spatialised speaker system allowing for multiple users and a communal experience is an example of this [12].

3. DESIGN OF VIRTUAL ENVIRONMENT

The conceptual idea of the virtual mixing environment is pictured in figure 2. As illustrated the user is placed in a 3D VR environment where different tracks from an arbitrary DAW, in the case of this project being *Ableton Live*, are represented as spherical sound sources in space with labels matching those of the DAW. A ray is cast from the controllers to signify which sound source is interacted with. The controller responds with vibration to signify contact between the ray and a sound source of choice. After selecting a sound source, the user can now move it in space from which data on position, distance, and angles to the head of the user will be collected. The data is passed into a binaural rendering system made in Max MSP [1], where the spatialisation is processed. This results in a match between visual and auditory locations of the sound sources and gives an audiovisual experience in space.

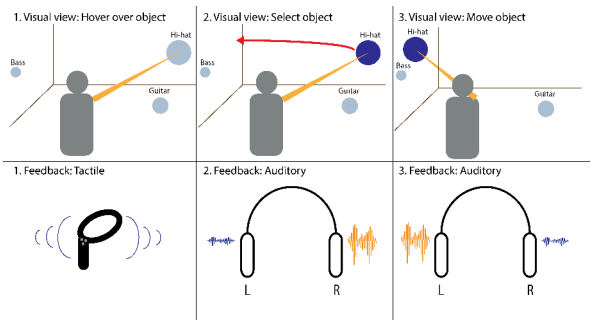

Figure 2. *Illustration of the concept.*

From the design principles presented earlier, the following decisions have been made for the virtual mixing environment:

1. A simple yet aesthetic environment will be created to focus on the importance of the mixing task. Com-

[1] https://cycling74.com/products/max

plex scenes such as concert halls and theater stages have been avoided as this might take focus away from listening.

2. The user-centric perspective will be utilised as it matches the 3D audio use-case and elicits the possibility of matching experienced sound to out-of-head localisation cues supporting the virtual body-ownership.

3. Spheres/sound sources are depicted around a stage to utilise benefits of the virtual mixer/stage metaphor such as 'intuitiveness', 'enjoyability' and 'ability to reveal the spaciousness of the mix', as earlier stated by Gelineck et. Al (section 2).

For the spatialisation of audio, dynamic binaural synthesis was implemented. For the synthesis, head-related impulse responses (HRIRs) from the MIT Media Lab [2] were used. The pack includes IRs ranging from -40 to +90 degrees on the vertical axis, where each elevation angle had corresponding IRs for 360 degrees on the azimuth in 5-degree intervals. The values of each IR were stored in text files readable as matrices in Max MSP and convolved with incoming audio tracks real-time, depending on data sent by the VE. Linear interpolation was implemented for each IR on the azimuth. The convolution of the incoming signal and the different HRIRs was done in frequency domain allowing for faster processing, whereas distance simulation and real-world spectral cues were simulated using the inverse-square law in conjunction with subtle coloration from low-pass filters.

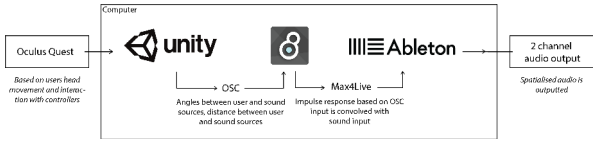

Figure 3. *Illustration of the system.*

The pipeline of the interactive VR environment was as follows:

1. Firstly, a combination of the Oculus Quest system and the game engine "Unity" was used to create a 3D environment that allowed the user to manipulate and position objects within a virtual space.

2. Secondly, object coordinates, angles, and user's head rotation, were sent through Open Sound Control (OSC) to Max MSP via the User Datagram Protocol (UDP) connection. The VE itself is thus not producing any audio output, rather it sends information to a DAW, which can be recorded for a binaural mixdown.

3. Max MSP and Ableton Live executed real-time sound rendering and binaural synthesis.

[2] `https://sound.media.mit.edu/resources/KEMAR.html`

An overview of the different stages, systems and the software used, can be seen in figure 3. The final VE used for evaluation purposes, is pictured below in figure [3] 4.

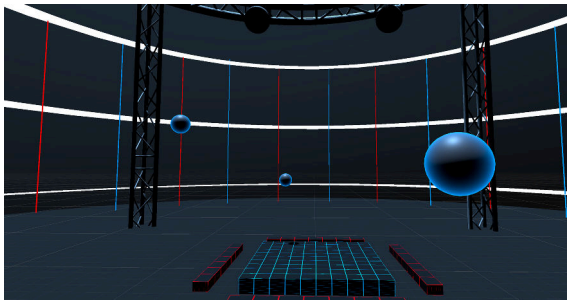

Figure 4. *Final design of the environment.*

4. METHODOLOGY

The VE and its communication of sound spatialisation was evaluated in two different ways. Firstly, mixing tasks with amateur music producers were carried out, assessing mixing precision and time spent on recreating a mix, pre-made using the developed spatial mixing tool. The mixing task was done comparatively, comparing the VE with the same environment implemented as a computer screen version e.g. the user would move icons on a screen with a mouse and keyboard rather than using a head-mounted display (HMD) and controllers. The computer version was meant to act as a test condition representing a traditional workflow that did not include the use of spatial interaction and perceptual cues. To quantitatively assess and isolate potential improvements of the VR version, differences in mix *precision* and general time *efficiency*, were measured. Precision was calculated based on relative distances between sound sources in a target mix and the participants' recreation thereof. Secondly, a focus group interview with music production students at the "Royal Rhythmic Academy of Music" in Copenhagen was carried out to qualitatively research general expert opinions on the product and its potential use-cases.

The bank of impulse responses contained 145 files for the azimuth, whereas 1434 files would be included for a full spherical binaural experience. As part of optimising the convolution process with the HRIRs, which is a computationally heavy process, and in order to simplify the information of the matrix containing the HRIR information, it therefore was examined whether elevation cues, in form of elevation HRIRs only, actually were needed from a perceptual perspective. The perceptual experiment was conducted (n = 14) at Aalborg University Copenhagen. The participants of the study were informed of the research question *"Do you feel like the sound is matching the position of the object?*" before the test started and asked to answer either 'yes' or 'no', with the option to hear the sound

[3] For video demonstration see `https://vimeo.com/674596681`

again. Hereafter, a series of visual 'sound sources' happening simultaneously as the spatialised auditory stimuli at different locations, either matching or not matching the visual sources, was presented to the participants. It was found that 95.2% felt the audio matched the visual position when the visual stimuli were elevated while the auditory stimuli still were kept at 0 degrees. On the other hand, 91.4% of the participants felt that the audio matched the position of the visual stimuli when both were at zero elevation. The results from this evaluation show that having visual cues for a given audio source, made the participants interpret the sound to originate from a visible object. Related research additionally shows that individual azimuth localisation is resistant to both elevation and reverberation [13], and it was therefore decided that HRIR convolution for vertical movement safely could be excluded in favor of processing power and general system complexity. The downside of this decision inevitably limits the systems potential and realism especially for sources placed at extreme vertical position where azimuth is irrelevant. However, only using azimuth IRs should be enough for a spatial audio mixing proof of concept.

5. PARTICIPANTS AND PROCEDURE

Both the focus group interview and the mixing task evaluation took place at Aalborg University in Copenhagen. 24 participants with different musical backgrounds and mixing experience, took part in the mixing task evaluation. Convenience sampling was utilised as all participants were part of the researchers' network. All the participants had experience mixing music where 54.2% of the participants had 3+ years of experience and the remaining had 1-2 years of experience. Regarding experience mixing spatial audio (binaural, surround, ambisonic) the majority (54.2%) had not tried it before. The majority of the participants were additionally familiar with VR (83.3%). The mixing evaluation asked the participants to recreate a mix and allowed them to switch back and forth between the target mix and their mix. Sessions lasted between 25-45 minutes for each participant. 12 students and a professor from the 'Music Production Bachelor' of the 'Rhythmic Conservatory of Copenhagen' additionally attended a focus group interview, after having tried out and played with the system. The interview lasted for 90 minutes, was transcribed, and coded into important tags and labels.

6. EVALUATION

6.1 Mixing Evaluation

The VR mixing tool was evaluated through a mixing task against a computer version with similar functionality in order to shed light on eventual differences in precision across the two media. In the computer version the user can simulate head rotation by right-clicking and dragging the mouse. A 'track' is 'grabbed' by hovering the mouse cursor over a 'track-object' and holding the left mouse button, whereafter it is possible to move and place the grabbed object using the keyboard keys 'W-A-S-D'. Both mouse actions can be performed simultaneously to allow for moving objects and rotating around the scene at the same time. The reference mix of the mixing task has been realized with the proposed VR system. The means and standard deviations of the collected data for both time (in seconds) and sums of relative precision compared to the reference mix (in Unity units) are shown below.

	Precision (Unity units)		Time (s)	
	Screen	**VR**	**Screen**	**VR**
Mean	35.74	35.60	558.38	448.04
Std. dev.	12.64	12.48	325.21	248.17

Table 1. *Mean and Std Deviation of Precision and Time across the two experimental conditions.*

QQ-plots and Anderson-Darling tests confirmed that the gathered *precision* data was normally distributed (p = 0.4381 for screen, p = 0.0693 for VR) while the *time* data was found not to be normally distributed (p = 0.0005 for screen and p = 0.0422 for VR). Since only the *precision* data were normally distributed (see figure 5), t-tests were used to test for the null-hypothesis *"mixes made by the participants in the VR version has no difference in, or less, relative precision to the reference mix, compared to the computer version"*. No significant difference was found between the means of the two conditions and the null hypothesis could thus not be rejected (p = 0.9531). However, due to dynamic and stereo errors in one of the audio clips used in the target mix, several participants stated that this specific audio track was exceptionally difficult to localize in both test conditions. If the position of this track is left out, the t-test rejects the null hypothesis (p = 0.0015). Thus it points towards a tendency of mixes made in the VR version having a higher relative precision to the reference mix, than the mixes made in the computer version as long as audio sources are dynamically static and mono.

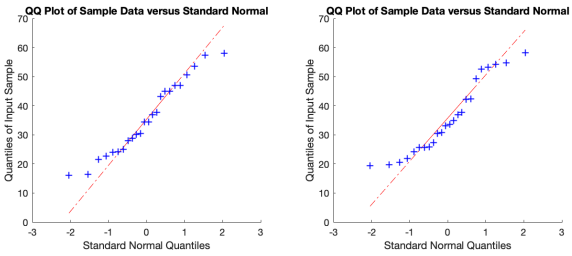

Figure 5. *Q-Q plots to check for normality of distribution of precision data. Left: screen version, right: VR version.*

6.2 Focus Group Interview

The second part of the evaluation consisted of a focus group interview and is below divided and categorised into themes and main topics, derived from a coding process. Table 2 highlights different quotes from the interview supporting the opinions on each topic.

6.2.1 Efficiency and Precision

The participants were asked about their initial thoughts regarding VR as a mixing tool and were all concerned of how efficient it potentially could be. Some participants felt that VR might be used as a quick sketching tool and they compared it to a big brush painting on a canvas. Besides the concept of this project, it was stated that it could be used as a more creative tool, rather than something one would use for precision. There was a general consensus that a DAW was expected to be more precise than the implemented VR program. It was mentioned by one of the participants that determining the program's efficiency and precision was difficult when they have not spent more time using the VR program, but that the program in general, together with the binaural algorithm, was experienced as something quick rather than precise. All participants agreed that the program gave enough information to make a judicious mix, but that the controllers made it hard to fiddle around and go into small details position-wise.

6.2.2 Spatial Sound Algorithm and Features

One participant described the panning as being "underdimensioned", though in general the spatial algorithm was found to be satisfying. A few participants noticed the exclusion of elevation, whereas most participants felt the match between sound and source movement realistic. Multiple participants described how they could imagine using this tool to create automation. Having different visual representations of the different sound sources as well as having a visual representation of sound activity on each track as a VU-meter on mixers was also mentioned.

6.2.3 Environment and Concept

It was stated that the virtual room could set a mood for the production by having different abstract elements. One participant pointed out that if the room should set a mood, it should be in a visually abstract way and not by looking realistic since this was the way they mixed mentally. It was, furthermore, stated that the decision of keeping the environment relatively neutral made sense in order not to influence the mix in an undesired direction and that the visuals used were pleasant and made sense in a mixing situation.

A participant pointed out that they found the prototype to be useless for them since it was designed for spatial mixing and not directly suitable for exporting a stereo mix. They pointed out that the prototype seemed to be designed for the producer to have a good experience instead of the final listener to have a good experience. It was mentioned by one participant that the application would be more relevant to use if it at least included the functions of a large-format console channel strip, for each audio track.

6.2.4 Comfort in VR

One participant explained how they felt dizzy after using the prototype, while another participant imagined that they could not spend more than 5-10 minutes in VR. It was furthermore discussed by several participants whether switching between headset and screen was better than staying in VR. Both ideas were supported by different participants.

7. DISCUSSION

While it is evident from the mixing task evaluation that recreating a mix in VR in some instances, and for some instruments, is more precise than doing it in an equivalent computer version, it is hard to draw any definite conclusions. Firstly, the mixing task evaluation was based on comparison, and it can be argued that the competition between the two conditions was slightly unfair due to the PC version being operated using a mouse and keyboard. Since the VE was designed to facilitate the spatial nature of VR, it could have been beneficial to compare the VR mixing tool against conventional spatial mixing tools for DAWs instead. Additionally, a quantitative measure in the form of *precision* was used to evaluate the mixing tool. While this allowed for quantitative comparison, including statistical hypothesis tests between interfaces, it is questionable whether the most *precise* tool is the best tool, as the task of mixing audio might be considered a subjective process and an artform. *Precision* as a valid evaluation metric in this instance, simply is debatable. The precision measure from the mixing task evaluation should therefore rather be used to support the qualitative data from the focus group interview, to establish a full picture. Lastly, it would have been desirable to have had multiple reference mixes, created on both conditions. This would have given a better understanding of navigation in the two environments and avoided any bias.

Concerning the focus group interview, two main findings are clear: 1. The participants saw the product as a quick sketching tool to test ideas and outline mixes rather than a tool to control precision and finer spatial details within the sound, and D. The participants were overall positive about the interaction with the product and its visual appearance, sensory benefits, and intuitive controls. In relation to the first finding, several participants stated that time and intuitiveness, in general, was an important aspect for them in a mixing tool/device and that the program indeed seemed to facilitate this. A participant expressed, among other things, that they "could imagine that you would get done faster with some things", and that the program seemed "very effective". Furthermore, participants agreed that the VR program definitely could be used as a quick sketching tool for swift ideas and testing of audio placement in a given space. This could, among other things, have been a result of the intuitive way of placing sound sources as well as the quick dynamic sound feedback and the possibility to link it up with Ableton Live. This could also have been due to the simplicity of the environment and the fact that only fundamental controls, pre-made audio effects, and interaction possibilities were included, giving it a 'to the bone' concept. Besides allowing positioning of sound sources (panning and volume), the program simply did not offer state-of-the-art possibilities such as the potential to manipulate the sound in finer detail, thus forcing the participant to use more time in the environment. This was moreover seen in the 'features' discussion of the interview, where participants emphasised a need for interactive dB meters, mute buttons, and the possibility to

Efficiency and Precision	”I could imagine that you would get done faster with some things. It seems very effective.”
Spatial Sound Algorithm and Features	”I found it slightly under-dimensioned so when you panned things to the side it was not as much as you would imagine. Front and back made good sense.”
	”It could also be used to do automation in a mix [...] You would have a much bigger area to draw on. I think that would be extremely useful.”
	”I think it is necessary to know that there is activity on the track”
Environment and Concept	”I think as the program is right now it might work even better for people who do not have experience making music and have to learn to visualise music in an extremely intuitive way.”
	”I cannot accept that I have not decided what it is this movement does. [...] I do not have any emotional connection to this.”
	”When I mix it is definitely something visual happening in front of me, I see the elements in front of me. It is not necessarily that I see the orchestra in front of me, it is much more abstract. A sprinkle over here, the sub-frequencies being another shape.”
Comfort in VR	”(In the environment, I could spend) 5-10 minutes or something like that”
	”I felt a bit sick. When I took off the glasses I felt really dizzy, but I think it is something you maybe have to get used to.”

Table 2. *Selected quotes from focus group interview.*

”do automation in a mix”.

It is worth discussing whether or not the HMD used for the project was the correct choice. The Oculus Quest was the chosen HMD due to it being wireless and thus consumer-relevant, providing the highest screen resolution compared to similar devices, as well as having a satisfactory refresh rate. However, since the hardware was built into the HMD, the computational power was limited. Limited computational power ultimately resulted in limited features in the final mixing tool. Features such as different shapes for different instruments and additional visual feedback were excluded from the implementation to accommodate low CPU load thus potentially reducing the overall usability of the tool. Additionally, the focus group agreed that adding more tracks in the VR environment would introduce clutter problems matching the suggestions in [10] in section 2. As only five tracks were part of the mix in the evaluation, having more could potentially eliminate the benefits of VR compared to PC. Related to this, it was stated: ”When we tried it here it was very manageable with five tracks, but if you have 67 tracks [...] it might hinder you more than it helps.” A suggestion for this was being able to group tracks. Concerning the concept, some participants struggled to grasp the core idea behind the product, dynamic spatial mixing. The fact that the mix changed relative to head movement confused many participants and hindered them in understanding the possibilities and functionality of it. As one participant said, ”I often ended up looking one way and then imagining that I mixed in stereo [...] this just made me feel that everything was imprecise.”

8. CONCLUSION

This paper has presented the design, implementation, and evaluation of a VR environment controlling dynamic binaural synthesis for 3D audio mixing. A real-time Max MSP patch was implemented to convolve incoming audio with HRIRs retrieved from data sent by the VE through OSC. The implementation allows for real-time sound rendering and binaural synthesis based on virtual sound location data. Both qualitative and quantitative evaluations were conducted through the form of a focus group interview and a mixing task evaluation. The result from the mixing evaluation hints toward the VR mixing tool improving mix precision for a simple audio mix, compared to a computer version. The answers from the focus group interview, furthermore, confirm the potential of a VR mixing tool and its spatial as well perceptual benefits. However, it is stated that it could better serve as a creative ’sketching’ tool to quickly try out different ideas than a precise spatial audio mixing environment, due to the lack of fine adjustment possibilities and potential visual clutter.

Acknowledgments

The participation to the conference was supported by the European Art Science and Technology Network (EASTN-DC).

9. REFERENCES

[1] A. R. Jensenius and M. J. Lyons, *A NIME Reader: Fifteen Years of New Interfaces for Musical Expression*. Springer, 2017, vol. 3.

[2] S. Gelineck, M. Büchert, and J. Andersen, “Towards a more flexible and creative music mixing interface,” 04 2013, pp. 733–738.

[3] D. Gibson and G. Petersen, *The Art of Mixing: A Visual Guide to Recording, Engineering, and Production*, ser. Mix pro audio series. MixBooks, 1997. [Online]. Available: https://books.google.dk/books?id=T34yQAAACAAJ

[4] B. De Man, N. Jillings, and R. Stables, “Comparing stage metaphor interfaces as a controller for stereo position and level,” 09 2018.

[5] S. Yantis and A. A. Richard, *Sensation and Perception*. New York, NY: Worth Publishers, 2014.

[6] J. Ratcliffe, "Hand and finger motion-controlled audio mixing interface," in *NIME*, 2014.

[7] G. Wang, "Principles of visual design for computer music," 09 2014.

[8] S. Serafin, C. Erkut, J. Kojs, N. Nilsson, and R. Nordahl, "Virtual reality musical instruments: State of the art, design principles, and future directions," *Computer Music Journal*, vol. 40, no. 3, pp. 22–40, 2016.

[9] S. Gelineck, D. Korsgaard, and M. Büchert, "Stage- vs. channel-strip metaphor: Comparing performance when adjusting volume and panning of a single channel in a stereo mix," in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME 2015)*, E. Berdahl, Ed. Louisiana State University, 6 2015, pp. 343–346.

[10] W. Gale and J. Wakefield, "Investigating the use of virtual reality to solve the underlying problems with the 3d stage paradigm."

[11] U. Zölzer, *DAFX: Digital Audio Effects*. Wiley, 2011. [Online]. Available: https://books.google.dk/books?id=jILxqgnjDKgC

[12] R. Hamilton, "Building interactive networked musical environments using q3osc," 02 2009.

[13] E. Méaux and S. Marchand, "Sound Source Localization from Interaural Cues: Estimation of the Azimuth and Effect of the Elevation," in *Forum Acusticum*, ser. Proceedings Forum Acousticum, Lyon (on line), France, Dec. 2020. [Online]. Available: https://hal.archives-ouvertes.fr/hal-03042326