ARMixer: Live Stage Monitor Mixing through Gestural Interaction in Augmented Reality

Weihan Huang
Stephanie Bourgeois
yuiwong@kmd.keio.ac.jp
sbourgeois@kmd.keio.ac.jp
Keio University Graduate School of Media Design
Yokohama, Japan

Yun Suen Pai
Kouta Minamizawa
pai@kmd.keio.ac.jp
kouta@kmd.keio.ac.jp
Keio University Graduate School of Media Design
Yokohama, Japan

ABSTRACT

Existing stage monitor mixing systems are inefficient and cannot accommodate the communication between the musicians and sound engineers. We introduce ARMixer, which allows musicians to perform self-stage monitor mixing through gestures in augmented reality to provide an intuitive mixing experience. We performed two usability tests and found that ARMixer is acceptable to the user and has excellent psychoacoustic intuitiveness in terms of mixing parameter controls by gestures and identifying mixing target.

ACM Reference Format:
Weihan Huang, Stephanie Bourgeois, Yun Suen Pai, and Kouta Minamizawa. 2021. ARMixer: Live Stage Monitor Mixing through Gestural Interaction in Augmented Reality. In *SIGGRAPH Asia 2021 Posters (SA '21), December 14–17, 2021, Tokyo, Japan.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3476124.3488632

1 INTRODUCTION

Audio mixing plays an important role in a music production, and using the visual metaphor of an audio mixing interface (AMI) can help users familiarise with the interface layout, recall mixing knowledge, and complete a mixing task. Compared to a conventional channel strip metaphor which is used with several knobs and faders, the stage metaphor is based on the concept of "deep mixing" to create a virtual stage that presents each audio channel through spheres, providing an intuitive mixing experience [2019]. However, stage metaphor is not widely used in the AMI field because the AMI becomes cluttered and difficult to be manipulated when multi-channels (spheres) need to be processed [2017; 2016]. Additionally, current stage monitor mixing systems have various problems, for instance, the sound from wedge monitors would be "collected" by microphones, once it is too loud, there is a risk of whistling. Also, it is a high cost for small venues to use in-ear monitors. More importantly, the communication efficiency of mixing between musicians and sound engineers is also a challenge [2012]. [2015] proposed a mobile application to allow performers to mix monitors, but switching the interface between the instrument and phone screen may interfere with the performers' ongoing task. Therefore, we introduce ARMixer, a system which uses the stage metaphor for its interface, allows musicians to perform an in-situ self-stage monitor mixing through gestures in augmented reality (AR) and provides the intuitive mixing experience.

2 DESIGN & IMPLEMENTATION

Each musician with an instrument on stage can actually be seen as a single audio channel. We designed ARMixer so that the virtual AMI corresponds to the position and placement of other surrounding musicians. Only one audio channel is processed per mix, meaning the interface is simple and the stage metaphor is appropriate to be used in this monitor mixing scenario. Furthermore, compared to a tangible interface that may affect the spectacle of the stage, the interface leveraging AR is personalized because the stage monitor serves the individual musician.

2.1 Hardware & Software

As shown in Fig 1, we simulated the video see-through AR HMD by assembling the *Oculus Quest 2* VR headset with the *Zed Mini* stereo camera. The *Leap Motion* hand controller was also attached in front of the VR headset for compatibility with natural gesture interaction. In addition, we used an audio interface (*Focusrite Scarlett 2i2*) with a microphone or other instruments connected as the real-time audio input and output devices in the program. Finally, all gesture interactions and interface were edited in *Unity*.

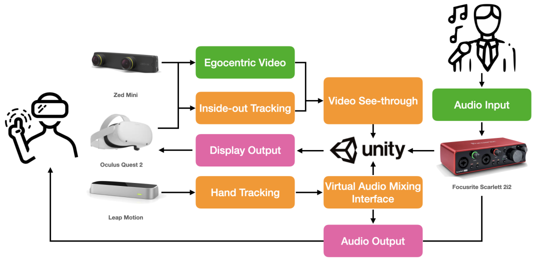


Figure 1: ARMixer System Overview

2.2 User Interface

For the AMI elements, the single audio channel is represented by a blue sphere following the stage metaphor paradigm. ARMixer provides four mixing parameters: volume, pan, reverb, and equalizer (high 12kHz, mid 2.5kHz, and low 80Hz frequencies). As shown

SA '21, December 14–17, 2021, Tokyo, Japan
© 2021 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-8687-6/21/12.
https://doi.org/10.1145/3476124.3488632

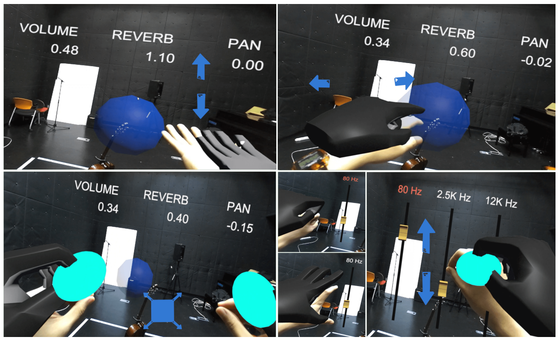


Figure 2: ARMixer Interface: Volume, Pan, Reverb and Equalizer Controls

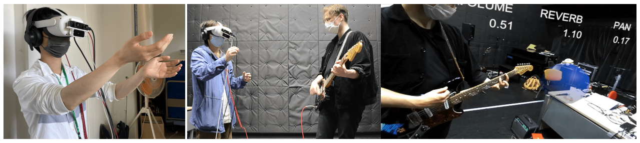


Figure 3: Left: Experiment 1; Right: Experiment 2

in Fig 2, **Volume Control:** When the user holds their right hand out with the palm facing upwards and moves it up and down along the y-axis in space ↕, the audio volume increases and decreases accordingly. When the volume increases, the opacity of the blue sphere increases at the same rate. **Pan Control:** As the user keeps their left palm facing downwards and moves it left or right along the x-axis in space in front of them ↔, the audio source pans left or right accordingly. **Reverb Control:** When the user zooms into the floating sphere, the reverb becomes stronger as the sphere gets bigger. The user can expand the sphere by pinching their two hands simultaneously. **Equalizer Control:** The user can select the fader by pinching with their right hand and moving up and down along the y-axis in space ↕, while opening and closing their left palm to switch the frequency.

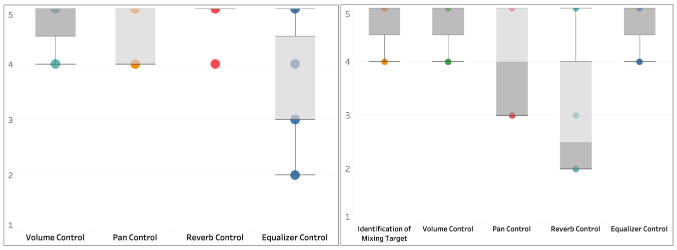
3 USABILITY EVALUATION

We performed two preliminary usability evaluations for ARMixer (Fig 3). All participants were given 10 minutes to familiarize themselves with the gestures. For the first study, 7 participants (3 males and 4 females, age mean 24.29, SD 1.38, average 4.29 years of stage performance experience) were recruited. Each participant was required to mix a vocal channel in a pre-recorded song. For the second study, we invited 4 participants (3 males and 1 female, ages mean 25.75, SD 2.36, average 4 years of stage performance experience) to simulate a stage monitor mixing scenario in a music practice room. They were divided into two two-piece bands and were asked to perform a real-time mixing of a bandmate's channel.

4 RESULTS & DISCUSSION

We used the System Usability Scale (SUS) for evaluating the acceptability. As a result, the average SUS score from *ex1* was 74.29 (SD 8.50), indicating that ARMixer was acceptable and has the potential to be a good product (A score of 70 or more means an acceptable product). On the other hand, the average SUS score from *ex2* was 64.36 (SD 5.54), the score was decreased because participants thought that they had to perform with their instruments while mixing and the physical pressure from HMD distracted them from the performance.

We evaluated the intuitiveness via a 5-point Likert Scale questionnaire which includes four intuitiveness of the metaphorical relationship between gestural interaction and parameter control and the identification of mixing target (only available in *Ex2*). In Fig 4, the lower whiskers of the volume, pan, reverb controls were above 4 from *ex1*, indicating strong agreements on all terms. Participants have expressed different opinions in the equalizer control, 1 of 7 commented that the visualization of the reverb control was not as intuitive and simple as the former three. In *ex2*, the lower whiskers of the volume, equalizer controls, and the identification of mixing target were above 4, suggesting that these three criteria had excellent intuitiveness. Pan control also performed well as it had a median of 4. Although participants from *ex2* showed diverse opinions, 2 of them agreed that reverb control was the most satisfactory parameter control. However, all participants thought that the accuracy of gesture recognition should be improved.



Left: Intuitiveness Performance in *Ex1*; Right: Intuitiveness Performance in *Ex2*

Figure 4: Plot for Intuitiveness Evaluation

5 CONCLUSION & FUTURE WORK

The users deemed it acceptable to mix in AR. For our future works, we will explore the use of optical AR glasses compatible with eyegaze tracking and use machine learning to improve the accuracy of hand tracking. Finally, we hope to conduct more tests with professional musicians to receive more accurate insight.

REFERENCES

Christopher Dewey and Jonathan Wakefield. 2017. Formal usability evaluation of audio track widget graphical representation for two-dimensional stage audio mixing interface. In *142nd Audio Engineering Society International Convention 2017, AES 2017*. Audio Engineering Society, United States.

Steven Gelineck and Anders Kirk Uhrenholt. 2016. Exploring Visualisation of Channel Activity, Levels and EQ for User Interfaces Implementing the Stage Metaphor for Music Mixing. In *Proceedings of the 2nd AES Workshop on Intelligent Music Production*, Vol. 13. Audio Engineering Society, London, United Kingdom.

David Gibson. 2019. *The art of mixing: a visual guide to recording, engineering, and production.* Routledge, New York, NY, USA.

Brad Herring. 2012. *Sound, lighting and video: a resource for worship.* Routledge, New York, NY, USA.

Andries Valstar, Min-Chieh Hsiu, Te-Yen Wu, and Mike Y. Chen. 2015. Giggler: An Intuitive, Real-Time Integrated Wireless In-Ear Monitoring and Personal Mixing System Using Mobile Devices. In *Proceedings of the 23rd ACM International Conference on Multimedia* (Brisbane, Australia) *(MM '15)*. Association for Computing Machinery, New York, NY, USA, 971–974. https://doi.org/10.1145/2733373.2806377