Human Factors in Virtual Environments and Game Design, Vol. 137, 2024, 88–94
https://doi.org/10.54941/ahfe1004990

Articulated Spatial Audio for Minimally Invasive Surgery Training

Yang Cai and Joshua Paik

University of California San Diego, La Jolla, CA, USA

ABSTRACT

Contemporary spatial sound recording and reconstruction systems enable audiences to experience realistic 3D soundscapes from multiple speakers or binaural headsets. Spatial audio is especially useful in minimally invasive surgery training which represents the sound sources, dynamic patterns, and verbal communications. In our study, we recorded and reconstructed 3D sound from a laparoscopic surgery room with multiple cases. The spatial sound data contains ambient sounds, equipment sounds, and music in the OR environment. We then articulated the ECG sound based on the simulated patient's conditions. For example, when the patient feels pain, the heart rate increases. Our experiments show that the articulated spatial audio helps to narrow the gap between the abstract training boxes and the actual OR environments. It is one step forward to advance Extended Reality (XR) with physical and physiological variables.

Keywords: Spatial audio, 3D sound, Minimally invasive surgery, MIS, Surgery training, Laparoscopic surgery, Articulated sound, Ambient sound, Surgical theater, Augmented reality, Virtual reality, Extended reality, Human factors

INTRODUCTION

Extended Reality (XR) technologies are designed to bring hyper-realistic experience from the mixed virtual and physical world, including realistic soundscapes. Typically, the soundscapes are in two channels to form a stereo sound. With affordable head tracking devices, spatial audio, or 3D sound, becomes reality. Spatial audio has more than two channels, enabling the user to locate the sound sources from multiple channels in a 3D immersive environment. The auditory feedback empowers users to identify, track, and reach for objects with a designed accuracy. Acoustic touch has a potential to offer a wearable and effective method of sensory augmentation along with visual, haptic, and motion perceptions in the Extended Reality. We are wondering if we can record the spatial audio from a surgical operation room (OR) and reconstruct the sound to make a surgical simulation more realistic?

Let us look into the minimally invasive surgery (MIS) environment where surgeons' perceptions are limited because they have to see and feel through the holes on the patient's abdominal wall. Laparoscopic surgery is one of typical minimally invasive surgeries. It requires a high surgical skill to operate through an endoscope. It is extremely difficult for novice surgeons because of impairments in spatial and haptic perceptions, in the ability to perceive depth, to sense the difference in tissues, to develop mental models of the anatomical environments in perceptual-motor coordination, and to make decisions in response to adverse situations (Lin and Chen, 2013; Matern et al., 2005).

Contemporary spatial sound recording and reconstruction systems enable audiences to experience realistic 3D soundscapes from multiple speakers or binaural headsets. Spatial audio is especially useful in minimally invasive surgery training which represents the sound sources, dynamic patterns, and verbal communications. In our study, we recorded and reconstructed 3D sound from the animal training lab and a laparoscopic surgery room with multiple cases. The spatial sound data contains ambient sounds, equipment sounds, and music in the OR environment. We then articulated the ECG sound based on the simulated patient's conditions. For example, when the patient feels pain, the heart rate increases. We also articulated the pulsed sound from the laparoscopic electrical scissors, which reflects the surgeon's action in real time. Our experiments show that the articulated spatial audio helps to narrow the gap between the abstract training boxes and the actual OR environments. It is a one step forward to adavance Extended Reality (XR) with physical and physiological variables.

SPATIAL AUDIO RECORDING IN OPERATING ROOM

An operating room for laparoscopic surgery is a noisy environment: there are many sound sources from humans, instruments, operations, and ambient sound bounced from the floor, walls, and the ceiling. Figure 1 shows the layout diagram where our team collected the spatial audio data. There were at least four medical professionals in the room: the surgeon, two assistants, and the anesthetist. There were at least three critical instrumental racks: the laparoscopic tower, the ventilation machine, and vital signs, including an ECG device. The operational sound sources include the regular CO_2 pump and ECG, and the irregular talking between medical professionals, and the occasional buzzing sound from electrical knife cutting. In order to record the realistic spatial audio, we placed the spatial audio microphone on the top of the laparoscopic instrument tower.

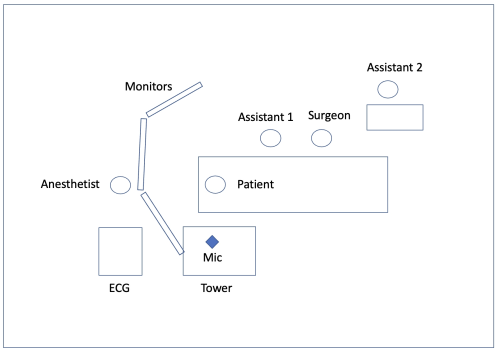

Figure 1: Overview of sound sources in the laparoscopic surgery operating room and the spatial audio recording microphone in the blue diamond symbol.

We used the Zoom H3-VR Audio Recorder, which is a device that records 360 degree audio in four channels. These 4 channels include one W channel (omnidirectional) and three directional channels: *X* (front-back), *Y* (left-right), and *Z* (up-down).

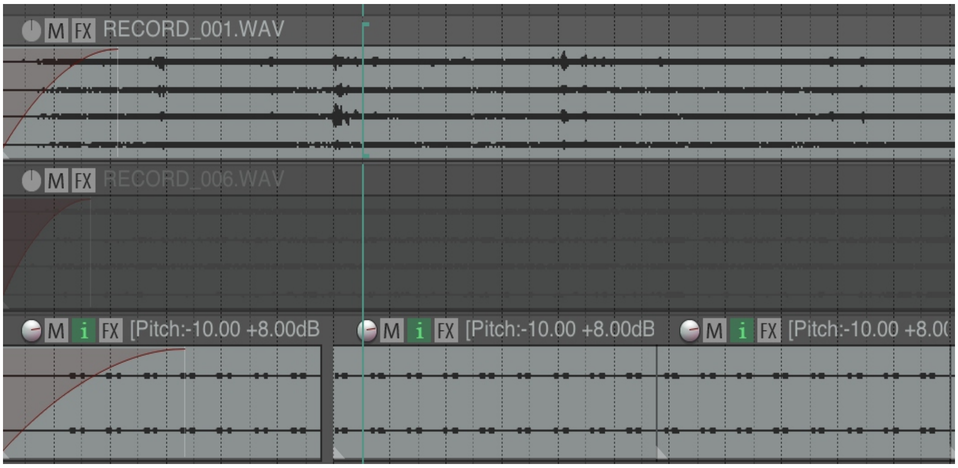

Figure 2: The recorded four-channel spatial audio waves.

RECONSTRUCTING THE SPATIAL SOUND

In order to record audio within a real-life laparoscopic surgery setting, this device first records audio in ambisonic A format which we then can extract and convert into ambisonic B format, specifically into ambiX format for the most clear audio quality.

A digital audio workstation called REAPER is used which supports multichannel audio manipulation. The plugin that is used in order to complete the conversion from ambisonic A to ambisonic B format was the "SoundField by RODE" plugin which is a top quality ambisonic processing and rendering application.

Once this audio file conversion is complete, we then use the "DearVR AMBI MICRO" plugin in order to convert the audio file into binaural audio to best fit the headphone or earphone format we intended to use with the Third Generation Airpods. On top of the ambisonic conversion processing, we also processed the audio to have the clearest mixdown using equalization plugins such as Fabfilter's Pro-Q 3, specifically targeting the lower end frequencies which contained distortive elements. How many channels do we need for reconstructing spatial audio? It depends on the available recording and playback devices. Since we have the 4-channel spatial audio recording microphone as input, we would recommend at least 4 output channels. The common limit for the output would be 24 channels, which depends on the spatial audio resolution, available playback channels, and the physical space. Figure 3 shows the graphical user interface for the 3D reconstruction of the spatial sound in a 5.1 setting which creates the effect of surround sound with the correct speaker system in place.

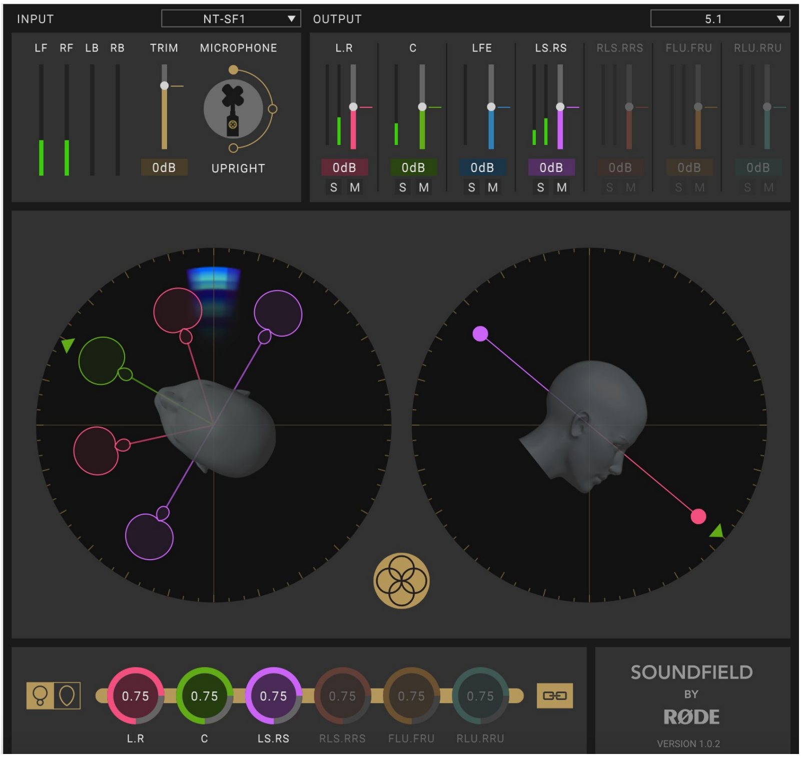

Figure 3: 3D reconstruction of the spatial sound.

PLAYBACK SPATIAL AUDIO

There are many ways in the methods of spatial audio playback. The easiest way is perhaps to use multiple speakers. We used five speakers to simulate the soundscape of a surgical operating room. Figure 4 shows the layout of the speakers in the lab space. The blue dots indicate the locations of the speakers and the green triangle indicates the listener's position. There is no need to implement any body or head tracking. Instead, the listener may move around and identify the individual locations of sound sources.

The second method is to use the binaural headphones. However, the headsets could have a space conflict with the regular virtual reality (VR) or augmented reality (AR) headsets if the user wanted to wear both headsets at the same time. So, we move on to the third method: a pair of earphones. We can incorporate the head-tracking technology with Apple's Airpods Generation 3 so that the ambisonic audio would also dynamically interact with the movement of the head just like in reality. For instance, as one turns their head while listening, an ECG machine can be heard in one corner of the room while ventilators and the sound of operating tools can be heard in another area of the room or in the center. Figure 4 shows the earphone-based spatial audio playback demo where the user's head is tracked.

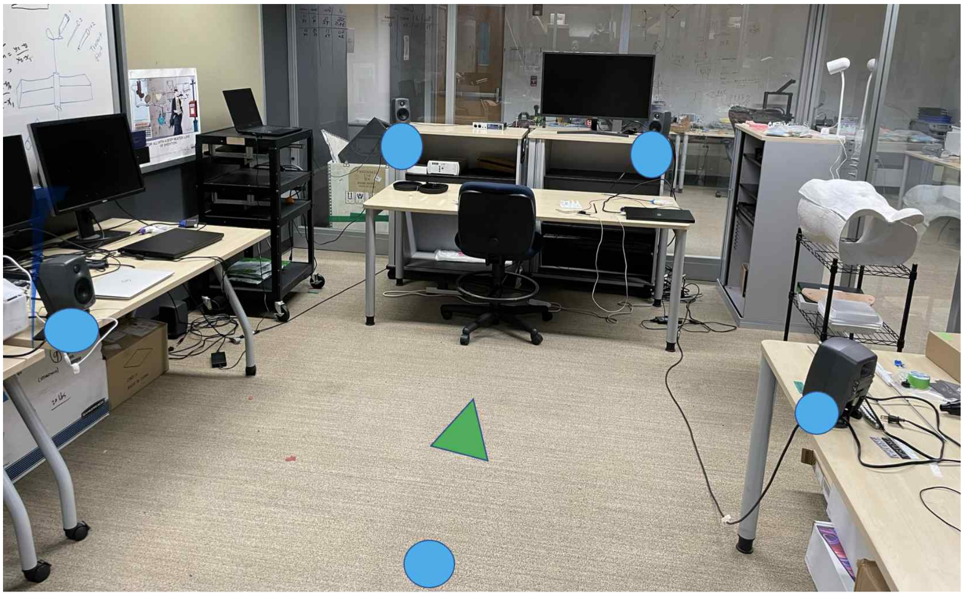

Figure 4: The spatial audio reconstruction with 5 speakers in the lab space. The blue circles are the locations of speakers. The green triangle is the location of the listener.

Figure 5: The earphone-based head-tracking spatial audio playback for the OR soundscape of laparoscopic surgery.

ARTICULATED SPATIAL AUDIO

Laparoscopic surgeons usually can detect anomalous situations from a patient's vital signs, for example, the pulse rate of the ECG sound. When a patient feels pain, the patient's pulse rate would increase significantly. The surgeon would inform the anesthetist to administer more anesthesia. In light of this, we want to articulate the ECG sound pattern to simulate the surgical scenarios. The same method could be applied to other sound sources.

Our articulated ECG machine sound is created through the manipulation of equalization filters and pitch shifting in order to create a distinct yet harmonically sound noise within the surgical setting. This sound is then processed using the ambisonic plugin "SoundField by RODE" and placed at a specific point within the surrounding soundscape in order to demonstrate an authentic spatial audio experience. Figure 6 shows the comparison of the ECG sound wave before and after surgical equalization where certain frequencies are reduced and enhanced.

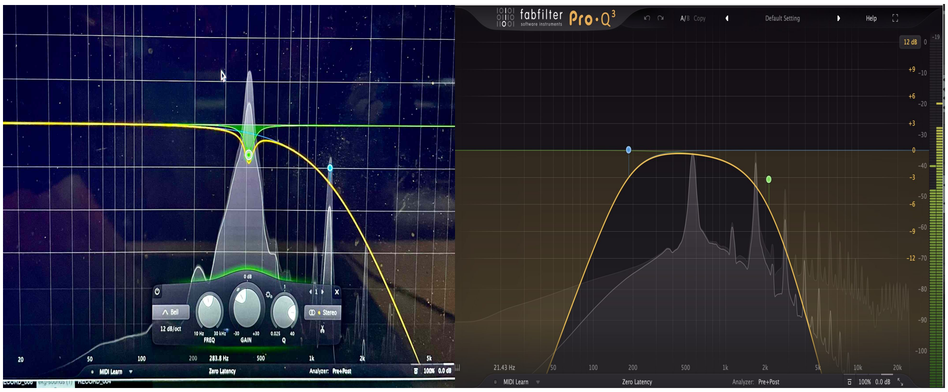

Figure 6: The frequency of the ECG sound is reduced to mimic the actual sound effect.

CONCLUSION

In this study, we explore how to record the spatial audio from an operating room, how to reconstruct the spatial audio, and playback in multiple speakers or the head-tracking earphones. We can also articulate the sound patterns of vital signs, specifically the ECG sound for simulating surgical scenarios. We found the sound articulation is significant to the minimally invasive surgery training because it can generate many surgical scenarios beyond the collected data.

ACKNOWLEDGMENT

The authors would like to thank the support of the Wellcome Leap SAVE Program and the NIST PSCR Program. The authors would like to thank Professor Shahrokh Yadegari and Autumn Hope for their support on spatial audio technology and data collection, and Professor Ryan Broderick for the observations and data collection in the operating room.

REFERENCES

Beattie, K. L., Hill, A., Horswill, M. S. *et al.* (2021). Laparoscopic skills training: the effects of viewing mode (2D vs. 3D) on skill acquisition and transfer. *Surg Endosc* 35, 4332–4344 (2021). https://doi.org/10.1007/s00464-020-07923-8

Chung, J. Y. and Sackier, J. M. (1998) A method of objectively evaluating improvements in laparoscopic skills," *Surgical Endoscopy*, vol. 12, no. 9, pp. 1111–1116, 1998.

Du, C., Li, J., Zhang, B. *et al.* (2022) Intraoperative navigation system with a multi-modality fusion of 3D virtual model and laparoscopic real-time images in laparoscopic pancreatic surgery: a preclinical study. *BMC Surg* 22, 139 (2022). https://doi.org/10.1186/s12893-022-01585-0

Kramer, G. (1994). Auditory display. Addison-Wesley Publishing Company, 1994

Lin, C. J. Chen, H. J. (2013). The Investigation of Laparoscopic Instrument Movement Control and Learning Effect, *BioMed Research International*, vol. 2013, Article ID 349825, 16 pages, 2013. https://doi.org/10.1155/2013/349825

Matern, U. Koneczny, S. Tedeus, M. Dietz, K. and Buess, G. (2005) Ergonomic testing of two different types of handles via virtual reality simulation, *Surgical Endoscopy and Other Interventional Techniques*, vol. 19, no. 8, pp. 1147–1150, 2005.

Sinha RY, Raje SR, Rao GA. (2022). Three-dimensional laparoscopy: Principles and practice. J Min Access Surg [serial online] 2017 [cited 2022 Oct 10];13: 165–9. Available from: https://www.journalofmas.com/text.asp?2017/13/3/165/181761

Wickens, C. D. (1999). Frames of reference for navigation," in *Attention and Performance XVII: Cognitive Regulation of Performance, Interaction of Theory and Application*, D. Gopher and A. Koriat, Eds., pp. 113–144, The MIT Press, Cambridge, Mass, USA, 1999.