Audio-Augmented Museum Experiences using Wearable Visual-Inertial Odometry

Jing Yang
Department of Computer Science
ETH Zurich, Switzerland
jing.yang@inf.ethz.ch

Gábor Sörös
Nokia Bell Labs
Budapest, Hungary
gabor.soros@nokia-bell-labs.com

ABSTRACT

The auditory sense is an intuitive and immersive channel to experience our surroundings, which motivates us to augment our perception of the real world with digital auditory content. We present a wearable audio augmented reality prototype that tracks the user with six degrees of freedom in a known environment, synthesizes 3D sounds, and plays spatialized audio from arbitrary objects to the user. Our prototype is built using head-mounted visual-inertial odometry, a sound simulation engine on a laptop, and off-the-shelf headphones. We demonstrate an application in a gallery scenario in which visitors can hear objects and scenes drawn in the paintings, feeling audio-visually engaged in the depicted surroundings. In a user study involving 26 participants, we observed that the audio-enhanced exhibition improved people's experience, as well as helped them remember more lively details of the artworks.

CCS CONCEPTS

• **Human-centered computing** → **Mixed / augmented reality**; **Auditory feedback**.

KEYWORDS

Wearable, Audio augmented reality, Visual-inertial odometry, Museum exhibition, User experience

ACM Reference Format:
Jing Yang and Gábor Sörös. 2019. Audio-Augmented Museum Experiences using Wearable Visual-Inertial Odometry. In *MUM 2019: 18th International Conference on Mobile and Ubiquitous Multimedia (MUM 2019), November 26–29, 2019, Pisa, Italy.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3365610.3365616

1 INTRODUCTION

Sound plays a significant role in many everyday situations. A sharp honk warns us about a car hurtling from behind; a piece of ringtone navigates us to a misplaced smartphone; a doorbell ring tells us the guest is waiting at the door. Usually, we not only perceive the semantic information from sounds, but also experience the sense of space which helps us locate sound sources [4]. This feature of spatial perception makes the auditory sense a potential channel for immersive human-object interactions. While equipping arbitrary everyday objects with a loudspeaker is not feasible nor acceptable, what if we could still hear sounds from them?

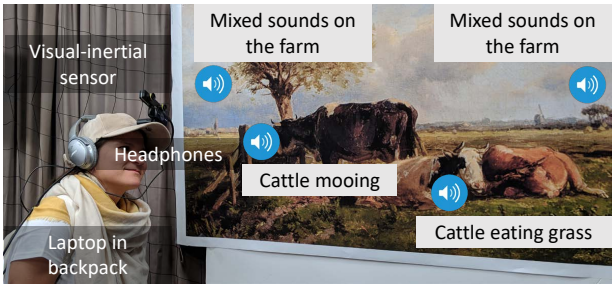

Figure 1: With our prototype, users can hear sounds coming from the painting while viewing it from arbitrary perspectives. The blue audio icons annotate the virtual sounds that are spatialized with authentic distances and directions relative to the user.

In this work, we explore audio augmented reality (AAR), a technology that augments objects with virtually spatialized sounds. The concept of AAR was proposed a few decades ago [3], but it has remained less explored in the field of AR research, of which a majority has focused on visual augmentation [14]. With AAR, users can perceive the virtually synthesized sounds as originating from a specific location in the space, by hearing them via normal headphones. AAR can provide authentic audio effects that enhance people's perception of the surroundings, and can serve as a new modality that facilitates intuitive interactions with everyday objects.

We present a wearable AAR prototype that synthesizes spatial sounds based on object-user poses in real time. When a user walks around in the space, we run visual-inertial odometry (VIO) with a head-mounted stereo camera pair to estimate the user's head pose with respect to the environment and the objects to be augmented with 3D audio. A laptop in the backpack updates the object-user pose and synthesizes 3D sound signals that are perceived by the user via headphones.

To demonstrate our prototype's applicability, we simulate a museum scenario where four landscape and genre paintings are mounted on the walls. Using spatialized sounds to enhance museum experience has been explored in the past [7, 26], but their focuses were on navigation or exhibit introduction. In our work, to provide visitors with an immersive experience, we spatialize corresponding sounds for the objects and the scenes drawn in the paintings (e.g. moo for cattle, as shown in Figure 1). Users are free to move around while viewing the paintings. They can hear the

MUM 2019, November 26–29, 2019, Pisa, Italy

ACM ISBN 978-1-4503-7624-2/19/11…$15.00
https://doi.org/10.1145/3365610.3365616

sound effects in 3D and experience an authentic atmosphere, feeling audio-visually engaged in the depicted scenes. In our user study involving 26 participants, people in general reported a better sense of engagement. Furthermore, they became more interested in and were more impressed by an artwork with such audio augmentation.

Beyond the gallery scenario, we also anticipate the potential to apply such an AAR system in everyday situations such as receiving notifications, sharing audio experience over space and time, home entertainment, etc. Our main contributions are:

(1) A wearable AAR prototype that can in real time spatialize 3D sounds from arbitrary objects to users;

(2) The demonstration of our prototype in a museum scenario;

(3) Qualitative and quantitative evaluation of how AAR can enhance the museum visitors' experience.

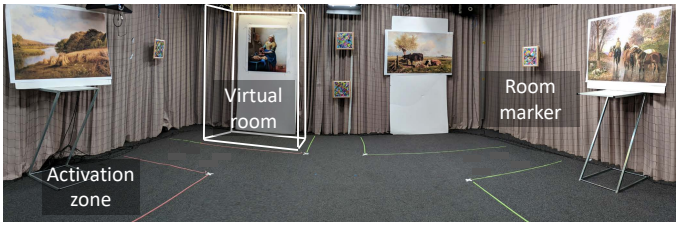

Figure 2: Paintings were placed at three sides of the room with a clockwise order from painting 1 to painting 4. The white cube represents the simulated indoor space for painting 2. On the floor we marked an activation zone for each painting. Four room markers were used to anchor and calibrate the location of the environment model.

2 RELATED WORK

Previous research has already shown the potential of AAR to redirect a user's attention, to help visually impaired people, to enhance visual experiences, and to understand surroundings. Typical AAR applications include navigation [1, 5, 12, 21], notification [2, 10, 11, 22, 23], and audio content creation [17]. Researchers also integrated AAR with other human senses such as visual and haptic senses to help a user understand an urban environment [13, 15].

The core to create authentic 3D audio experience is to precisely track the user-object pose with six degrees of freedom. In the past, tracking was usually implemented with environment cameras (outside-in), head-mounted cameras (inside-out), magnetic or radio-frequency modules, from which inside-out tracking fits our wearable purpose best. The egomotion of a camera can be estimated via simultaneous localization and mapping (SLAM) or visual odometry (VO) algorithms, such as DTAM [19], LSD-SLAM [8], ORB-SLAM [18], SVO [9], etc. In our initial exploration, we found their general problem of significant and frequent drifting when dealing with fast and abrupt head movements, but such head movements are common and subconscious when people walk around or respond to unexpected sounds. Therefore, we shifted our attention to visual-inertial odometry (VIO) approaches that are more robust against quick and sudden movements. When evaluating several VIO approaches, we experienced camera-IMU synchronization problems using the device[1] introduced by the authors of ROVIO [6] and OKVIS [16], and found the VINS-Mono [20] method not sufficiently stable on IOS mobile implementation[2]. Therefore, we decided to use an off-the-shelf VIO sensor that produced the best results in our implementation. More details follow in the next section.

Museums and galleries have been actively used to integrate and stimulate human senses, in order to explore novel ways of representing artworks and improve visitors' perception and interest. For example, Vi et al. [25] enhanced visitors' experience of paintings with sound, smell, and haptic feelings that were designed to reflect the artists' intentions. Spatial audio was applied in several related projects but it was mainly used to play navigation or exhibit introductions. Wakkary and Hatala [26] presented a navigation system that played sounds from left, right, and in front of the visitor. The mobile audio guide proposed by de Borba Campos et al. [7] could notify blind people about their orientation. Vazquez-Alvarez et al. [24] introduced a system that could play the introduction and other people's reviews of an artwork from its location. Different from these projects, we focus on enhancing the visitors' perception of the painting by auralizing the drawn contents. The Microsoft Oregon Project[3] is similar to ours as they augmented landscape drawings with sounds that were recorded on site during the painting process. Their system was built with trackers and loudspeakers that were suspended from the ceiling and covered the whole exhibition room. In contrast, we present wearable tracking and sound playback via personal headphones, which enables private and even personalized 3D audio experience in a public gallery or museum.

3 AUDIO-AUGMENTED EXHIBITION

3.1 The Real Environment

In a room of size $6\,m \times 6.5\,m \times 3.4\,m$, we distributed four paintings and four room markers on three sides as shown in Figure 2. 3D sounds were synthesized from the paintings. The markers were used to anchor and calibrate the visitor's location with respect to the room. The visitors wore a cap mounted with the visual-inertial sensor, a laptop in a backpack running the sound simulation engine, and a pair of unmodified headphones (see Figure 1), and they could walk around freely in the room.

We selected four landscape and genre paintings as shown in Figure 3, of which the sizes were around $1\,m^2$. These four paintings were selected because the scenes were depicted in a large depth range so that we could attach virtual sounds to distributed objects in the environment (see Figure 3). Such a sound distribution fit our goal to enhance the overall engagement with 3D soundscapes.

3.2 The Simulated Environment

Ideally, in order to create real audio experience that matches the artwork, the added spatialized sounds should be recorded on site. However, since for the selected paintings there exist no real-time recordings, we gathered audio clips online from YouTube and Freesound, and spatialized them from appropriate locations and distances. We

[1] https://github.com/ethz-asl/mav_tools_public/
[2] https://github.com/HKUST-Aerial-Robotics/VINS-Mobile
[3] https://www.microsoft.com/en-us/research/project/the-oregon-project/

Painting 1: *Harvesting* by Henry H. Parker

Painting 2: *De Melkmeid* by Johannes Vermeer

Painting 3: *Resting Cows* by Anton Mauve

Painting 4: *Homeward Bound* by John Frederick Herring Jr

Figure 3: The four paintings used in our application. Virtual sounds are illustrated with blue icons at corresponding positions. We spatialized the sounds with proper depth (1-200 m). Painting 2 is 0.82 m×0.92 m and the others are 1.4 m×0.85 m. The images are from `http://www.paintinghere.com`.

carefully selected sounds according to the painting contents in order to create authentic soundscape as it might exist in real life.

To model the sound source locations and acoustics effects that can further enhance the sense of presence (e.g. reverberation), we first made an environment model offline. In the game engine Unity3D, we made a digital copy of this room, with paintings and the sound sources registered at the corresponding locations. Painting 2 (*De Melkmeid*) depicts an indoor scenario, therefore, we simulated a virtual room (illustrated by the white cube in Figure 2) with coarse concrete as the surface material in order to produce proper indoor acoustics effects. Inside the white cube area, the sound of pouring milk would be clearer and the cock crowing and the dog barking could be perceived as coming from outside.

For the user study in this work, we intended to avoid the sound disturbance from the other paintings when the user was focusing on one of them. Therefore, we marked an activation zone for each painting on the floor (see Figure 2), and the sounds could only be heard inside the corresponding zone. In reality, we suppose that such zones are not necessary, especially if the paintings can group together as a complete scene, or if the sounds are intentionally utilized to attract the visitors.

3.3 User Tracking

In our work, we leveraged inside-out tracking to have a wearable form and to allow the freedom to operate in uninstrumented environments.

First, to anchor and activate the environment model with little restriction, we utilized room markers recognized by the Vuforia SDK. Since the marker positions were pre-defined, the model could be placed at the correct location with respect to the user from an arbitrary viewing position in the room. When walking around to view the paintings, the markers were also used to re-position the room model if the user tracking drifted over time.

Second, to estimate the visitor's head pose in real time, we utilized the ZED mini stereo camera with VIO implemented in the Stereolabs ZED SDK. Equipped with motion sensors, this camera could run VIO with a depth range of 0.15-12 m and a pose update frequency up to 100 Hz. Then we simultaneously approximated the visitor's real-time head pose in the Unity3D simulation.

Note that one can also track paintings and estimate visitors' relative pose using Vuforia, but this hardly works when walking between paintings without plenty markers in the camera's field of view. Plus, it can also be an issue if a painting is low on visual features or the painting's visual features are similar to the marker patterns. By using VIO, we can smoothly track the visitors, and therefore synthesize 3D sounds continuously from correct positions even if they move around a lot or view paintings at extreme angles. Multiple users can use such wearable tracking at the same time in the same space to enjoy their own AAR experience (even personalized, if they wish so). Note, however, that good lighting conditions and an adequate number of static visual features in the space are necessary for robust VIO.

For this prototype we used four room markers that were placed at around visitors' eye level. This was sufficient for our user study in this environment. In a real museum with a larger space and multiple users, one can distribute more markers at different heights. This can help avoid occluding the line of sight to the markers.

3.4 Sound Simulation and Delivery

Given the environment model and the visitor's head pose in the Unity3D scene, we utilized the Google Resonance Audio SDK to simulate the sound propagation and model the room acoustics. Finally, the spatialized sounds were played to the user via off-the-shelf headphones. Please refer to the accompanying video[4] to get an impression of the user experience.

Note that this system with offline environment acquisition and with online pose estimation as well as audio spatialization can also be easily extended to other exhibition scenarios.

4 EXPERIMENTS & EVALUATION

We intended to explore how the virtually spatialized audio could improve people's museum experience. We supposed that such 3D sound would make it more immersive and interesting to view a painting. In addition, the sound could help people remember more details of an artwork. We conducted a user study to verify these ideas and to understand users' experience.

4.1 Experiment Procedure

To design the experiment, we first invited five people to our preliminary test, in which they were asked to view the paintings in their

[4] https://youtu.be/TigYa-9VCYM

Table 1: The items for each painting in the multiple-choice question. In addition to the listed options, we also provided a choice of "other" where participants could add unlisted objects.

	multiple-choice items
painting 1	wheat field, river, sheep, birds, trees
painting 2	a milkmaid pouring milk, bread, pots&jars
painting 3	cattle, tree, fence, farmyard
painting 4	horses, people riding horses, trees, ducks, farmhouse, stream/brook

preferred order and experienced the audio effects of each painting. We found that four out of five participants viewed the paintings in a clockwise (1→2→3→4) or counterclockwise (4→3→2→1) order, while only one person followed a random order (2→3→1→4). In general, these five people gave us positive feedback, saying that the viewing experience was pretty interesting and immersive with the virtually added 3D sounds. Besides, a couple of them clearly recalled several sounds and their associated objects in the paintings.

Based on these initial findings, we invited 26 participants (age $\in [18, 42]$, average = 27.31, standard deviation = 5.62, 11 female) to our formal experiment that consisted of three parts: painting viewing, short interview, and questionnaire.

All 26 participants were asked to view all paintings in their preferred order wearing our prototype. However, during the viewing process, 13 people (group 1) only heard painting 1 and painting 3 while painting 2 and painting 4 were muted, while for the other 13 people (group 2) it was the other way round. Such a design was based on the following four reasons:

(1) By muting two paintings, participants could experience the difference between with sound and without sound.

(2) Since painting 1 and painting 4 had more audio-augmented objects than painting 2 and painting 3, making 1+3 and 2+4 as two groups could keep a roughly balanced perception load for each participant.

(3) As most people would view the paintings in the clockwise or counterclockwise order, it was more fair to group this way so to make the viewing process "with sound→without sound→with sound→without sound" or backwards.

(4) Having two groups could alleviate the influence on people's memory that was mainly caused by the contents of the painting but not by the existence of virtual sounds.

At the beginning of the study, participants were only informed that they would hear sounds when viewing two of the paintings, but they had no clue about the sound contents. This way, participants could immediately report if sounds were abnormally played but they would not have biased expectation that might influence their user experience. All participants started from the room center. In case the tracking drifted off significantly during the process (as monitored by the experiment investigator), they would be asked to re-anchor the environment by looking at the room markers.

After viewing the paintings, we first spent 2-3 minutes talking to the participants about their experience and feedback. After that they filled in a questionnaire that consisted of two parts. In the first part, they were asked to recall the painting contents by checking multiple choices for each painting. As listed in Table 1, these choices included 3-6 main objects and scenes such as "river" in painting 1 and "cattle" in painting 3. We designed the choices in a way that reflected the painting contents as well as corresponded to the added sounds. After the multiple choices, they answered the following three questions on a 5-point Likert scale from "strongly disagree" (1) to "strongly agree" (5): (Q1) I think I can remember a painting better with such 3D content-related sounds. (Q2) The sense of engagement was better with such 3D sounds than without them. (Q3) With such 3D sounds I will be more interested to view a painting even if it is not my preferred genre.

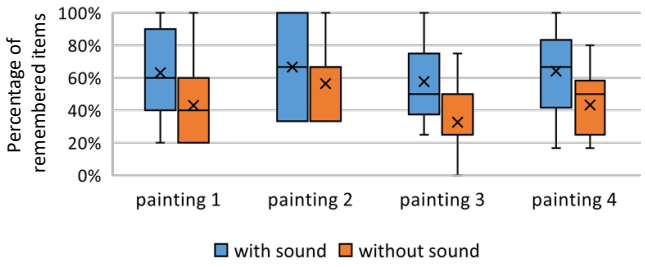

Figure 4: Results of the multiple choice questions, shown with percentage of items remembered in each painting. For painting 2, the median line for "without sound" overlaps the upper limit of the box while for painting 3 it overlaps the lower limit of the box. Note that the results are not comparable across paintings due to the different numbers of choices for each painting. It shows that people tended to remember a painting better with 3D content-related sounds.

4.2 Experience Improvement with 3D Audio

During the short interview, participants reported that they could experience *"very obvious"* difference between with sound and without sound. Some participants felt that the 3D audio gave them another channel to experience the painting — they generally commented the viewing experience with 3D audio as *"interesting"*, *"amazing"*, and *"new"*, and they *"enjoyed it very much"*. Some participants clearly stated that they could feel the sounds coming from different locations, which engaged them in the painted scenarios more than the ones without audio augmentation. During the user study, we also observed that several participants were very excited about the 3D audio effects and they turned head a lot or walked around to experience the sounds at different angles. Participants also shared their ideas for improvement. Four participants suggested to tune the sounds based on the user's viewing behavior. For example, the sound volume can be adjusted if a user focuses on a specific object. It would be interesting to conduct a future study and investigate users' experience with the integration of gaze tracking.

Regarding the multiple choice questions, Figure 4 compares how much the participants could recall about each painting in percentage. It shows that participants generally remembered a painting better if it was augmented with content-related spatial sounds.

On average, participants could recall 62.9% of objects from audio-augmented paintings, but only 43.9% without sounds (mean difference 19%). However, the memory performance varied largely among individuals. Seven participants remembered more about the silent paintings, but the differences (mean 11.7%) were smaller than the overall average difference (19%). Note that not all the given choices were able to be augmented with sounds, e.g. the bread in painting 2. Some people just overlooked such choices although they heard sounds from the painting. However, several participants reported that the virtual sounds attracted them to *"view the painting more carefully"* and then they could remember such silent objects.

Another finding is that the added sounds could make their associated objects more memorable, even if the objects were not remarkable enough. For example, the sheep in painting 1 were small as being on the field across the river. With the added 3D bleating, eight participants from group 1 *"remembered the sheep sound"* thus remembered their presence in the painting, while only one person from group 2 recalled them. In our everyday life there also exist similar situations, where we notice inconspicuous objects by their sound.

In general, the results verify our assumption that 3D sounds can help people remember more details of a painting, which also corresponds to their self-evaluation in Q1 (average = 4.01, 95% confidence interval = [3.75, 4.26]). Regarding the score of Q1, we further conducted a Mann-Whitney U Test and found no significant difference between two groups of participants (U = 74.5, p = 0.571).

As noticed in the interview, some participants had a clear audio memory after viewing the paintings. Some sounds belonged to invisible things, but they fit well and enhanced the participants' overall perception. For instance, to painting 2 we attached cock crowing and dog barking outside, which, as commented by some participants, immersed them into a *"lively countryside morning"* scenario. For painting 1 and painting 4 we added the sound of wind, and for painting 3 we added mixed sounds on a farm (birds chirping, farm work, etc.). In addition to the sounds for visible objects, these environment sounds also contributed to a better sense of engagement for our participants, as also indicated by their answers for Q2 (average = 3.88, 95% confidence interval = [3.56, 4.19]). No significant difference was found through a Mann-Whitney U Test (U = 79.5, p = 0.780).

Finally, the participants leaned towards the opinion that with such 3D sounds, they would be more interested to view a painting, even if it is not of their favorite genre (Q3, average = 3.62, 95% confidence interval = [3.32, 3.89]). Like before, there existed no significant difference between two groups (Mann-Whitney U Test, U = 78, p = 0.717).

In summary, according to the participants' verbal feedback and their questionnaire answers, we believe that AAR in museum can improve visitors' sense of engagement and facilitate a clearer memory of art pieces.

4.3 Tracking Accuracy

In addition to the appropriate audio design and spatialization, the accurate tracking also made important contribution to the positive user experience. During the user study, all 26 participants walked at a normal speed and viewed the paintings with natural head movements, but their poses varied significantly from each other. Their viewing processes lasted roughly 1-3 minutes. 20 of them experienced a rather stable tracking, in which the sounds were perceived continuously from expected orientations. The zones marked on the floor also matched the audio activation well. However, another six participants did not immediately hear any sound when stepping into the activation zones due to some drift in the horizontal directions, so they re-calibrated the location with the markers on the wall. Upon hearing 3D sounds, some participants viewed the paintings at different angles and turned their heads a lot to experience the audio effects, and our prototype ran robustly to handle this.

To further evaluate the tracking performance, we compared 10 movement trajectories captured by our prototype with the ground truth captured by a state-of-the-art tracking system Vicon[5]. In order to do a simultaneous tracking by both systems, the investigator walked in the room wearing the VIO-cap together with the helmet tracked by Vicon. Each movement lasts around 20 seconds. Figure 5 shows an example comparison in head position and rotation. Regarding the position, despite some errors in the x and y directions, our prototype tracks better in the horizontal plane than the vertical direction z, in which it may drift around 30 cm. Fortunately, such a vertical error has little influence on participants' audio perception for two reasons. First, this error is not significant compared to our painting sizes and some sound source distances (see the caption of Figure 3). Second, the vertical direction is difficult to determine due to the lack of the differences in sound intensities and arrival times at our two ears. Regarding the rotation, our prototype can track well in all three dimensions. From the roll and the yaw plots we see a roughly constant error, which corresponds to the initial pose difference between the VIO-cap and the Vicon helmet.

4.4 Discussion

From the results we have seen that the virtual spatial audio channel could engage visitors in a depicted scene and help them remember more lively details of a painting. Note that we also concern the original intention that an artist would like to deliver in his/her work. It is possible that an artist prefers to keep a broad space of imagination for visitors and this might be offended by such an additional layer of perception. This potential offense is probably more severe for non-realistic art, such as abstract and surrealistic paintings. Therefore, our intention is to demonstrate the possibility and potential benefits by applying AAR in museums and art galleries, but not to confirm a better way for art expression. Our research results may inspire artists to consider including an auditory channel in their work. It would be interesting to further explore appropriate audio augmentation for artworks in collaboration with artists.

Beyond the museum scenario, we also see the potential to apply AAR systems in everyday applications. Users, especially visually impaired people, can use it to receive notifications from surrounding objects for navigation or understanding the environment. Considering the application of spatial sounds in VR games, we believe AAR can also be utilized for entertainment like on-site games, serving as another channel to convey information and/or enhancing immersion for players.

[5] https://www.vicon.com/

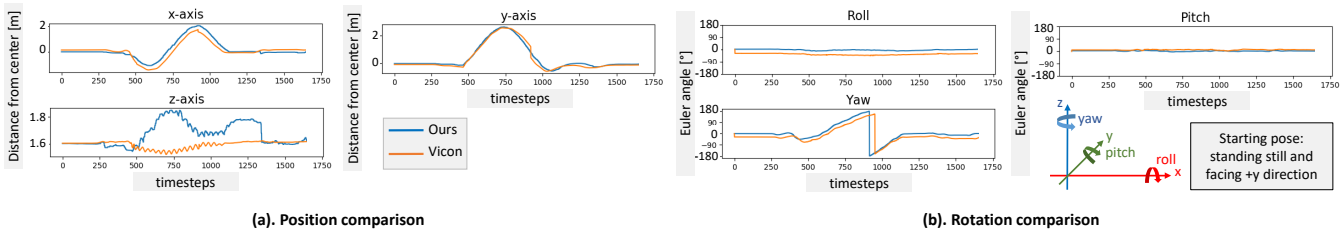

Figure 5: Comparison of the tracking accuracy between our prototype and the ground truth system Vicon. (a). Regarding the position, VIO tracks better in the horizontal directions than the vertical direction. (b). Regarding the rotation, VIO tracks well in all three dimensions.

5 CONCLUSION & FUTURE WORK

In this paper, we present a wearable AAR prototype that tracks user-object poses in real time, then synthesizes and delivers 3D sounds from objects to the user. We tested our prototype in an exhibition scenario, and showed that content-related spatialized sounds could improve visitors' sense of engagement as well as enhance their memory of a painting.

There are several interesting directions for future exploration. At the application level, as discussed in Section 4.4, we can collaborate with artists to investigate suitable audio augmentation that helps to express artwork. It is also interesting to test the application of AAR in other real-life situations, such as receiving notifications, sharing audio experiences, assisting visually impaired people, etc.

At the technique level, while the sound propagation and acoustics effects are simulated online, our prototype relies on offline environment acquisition, i.e., the objects and the room geometry and materials need to be modeled before use. This can work well in static environments such as home and offices where objects are placed at fixed locations, but dynamic environments require real-time environment modeling (geometry, materials, poses), which is our next step of exploration. Another issue to explore is how to implement such an AAR system using more portable devices like smartphones and headphones that are equipped with motion sensors. We anticipate that lightweight AAR systems can generally enhance our interaction with the surroundings in real life and eventually contribute to the vision of ubiquitous augmented reality.

REFERENCES

[1] R. Albrecht, R. Väänänen, and T. Lokki. 2016. Guided by Music: Pedestrian and Cyclist Navigation with Route and Beacon Guidance. *Personal and Ubiquitous Computing* 20, 20:1, 20(1), 121–145.

[2] A. Barde, M. Ward, W.S. Helton, M. Billinghurst, and G. Lee. 2016. Attention Redirection using Binaurally Spatialised Cues Delivered over a Bone Conduction Headset. In *Human Factors and Ergonomics Society Annual Meeting*. SAGE Publications.

[3] B. Bederson and A. Druin. 1995. Computer Augmented Environments: New Places to Learn, Work, and Play. *Advances in Human Computer Interaction* 5 (1995), 37–66.

[4] J. Blauert. 1997. *Spatial Hearing: The Psychophysics of Human Sound Localization.* MIT press.

[5] S. Blessenohl, C. Morrison, A. Criminisi, and J. Shotton. 2015. Improving Indoor mobility of The Visually Impaired with Depth-based Spatial Sound. In *IEEE ICCV*.

[6] M. Bloesch, S. Omari, M. Hutter, and R. Siegwart. 2015. Robust Visual Inertial Odometry Using A Direct EKF-based Approach. In *IEEE/RSJ IROS*.

[7] M. de Borba Campos, J. Sánchez, A.C. Martins, R.S. Santana, and M. Espinoza. 2014. Mobile Navigation through A Science Museum for Users Who Are Blind. In *UAHCI*.

[8] J. Engel, T. Schöps, and D. Cremers. 2014. LSD-SLAM: Large-Scale Direct Monocular SLAM. In *ECCV*.

[9] C. Forster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza. (2017). SVO: Semidirect Visual Odometry for Monocular and Multicamera Systems. *IEEE Transactions on Robotics* 33, 2, 33(2), 249–265.

[10] F. Heller and J. Borchers. 2014. AudioTorch: Using A Smartphone as Directional Microphone in Virtual Audio Spaces. In *ACM MobileHCI*.

[11] F. Heller, J. Jevanesan, P. Dietrich, and J. Borchers. 2016. Where Are We?: Evaluating The Current Rendering Fidelity of Mobile Audio Augmented Reality Systems. In *ACM MobileHCI*.

[12] F. Heller and J. Schöning. 2018. NavigaTone: Seamlessly Embedding Navigation Cues in Mobile Music Listening. In *ACM CHI*.

[13] Y. Hsieh, V. Orso, S. Andolina, M. Canaveras, D. Cabral, A. Spagnolli, L. Gamberini, and G. Jacucci. 2018. Interweaving Visual and Audio-Haptic Augmented Reality for Urban Exploration. In *ACM DIS*.

[14] K. Kim, M. Billinghurst, G. Bruder, H.B. Duh, and G.F. Welch. 2018. Revisiting Trends in Augmented Reality Research: A Review of The 2nd Decade of ISMAR (2008–2017). *IEEE TVCG* 24, 11 (2018), 2947–2962.

[15] T.CK. Kwok, P. Kiefer, V.R. Schinazi, B. Adams, and M. Raubal. 2019. Gaze-Guided Narratives: Adapting Audio Guide Content to Gaze in Virtual and Real Environments. In *ACM CHI*.

[16] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale. (2015). Keyframe-based Visual-Inertial Odometry Using Nonlinear Optimization. *The International Journal of Robotics Research* 34, 3, 34(4), 314–334.

[17] J. Müller, M. Geier, C. Dicke, and S. Spors. 2014. The BoomRoom: Mid-air Direct Interaction with Virtual Sound Sources. In *ACM CHI*.

[18] R. Mur-Artal, J.M.M. Montiel, and J.D. Tardos. (2015). ORB-SLAM: A Versatile And Accurate Monocular SLAM System. *IEEE Transactions on Robotics* 31, 5.

[19] R.A. Newcombe, S.J. Lovegrove, and A.J. Davison. 2011. DTAM: Dense Tracking And Mapping in Real-Time. In *IEEE ICCV*.

[20] T. Qin, P. Li, and S. Shen. (2018). VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator. *IEEE Transactions on Robotics* 34, 4, 34(4), 1004–1020.

[21] S. Russell, G. Dublon, and J.A. Paradiso. 2016. HearThere: Networked Sensory ProsThetics through Auditory Augmented Reality. In *ACM AH*.

[22] E. Schoop, J. Smith, and B. Hartmann. 2018. HindSight: Enhancing Spatial Awareness by Sonifying Detected Objects in Real-Time 360-Degree Video. In *ACM CHI*.

[23] T.J. Tang and W.H. Li. 2014. An Assistive Eyewear Prototype That Interactively Converts 3D Object Locations into Spatial Audio. In *ACM ISWC*.

[24] Y. Vazquez-Alvarez, M.P. Aylett, S.A. Brewster, R. von Jungenfeld, and A. Virolainen. 2014. Multilevel Auditory Displays for Mobile Eyes-free Location-based Interaction. In *ACM CHI*.

[25] C.T. Vi, D. Ablart, E. Gatti, C. Velasco, and M. Obrist. 2017. Not Just Seeing, But Also Feeling Art: Mid-air Haptic Experiences Integrated in a Multisensory Art Exhibition. *International Journal of Human-Computer Studies* 108 (2017), 1–14.

[26] R. Wakkary and M. Hatala. (2007). Situated Play in A Tangible Interface and Adaptive Audio Museum Guide. *Personal and Ubiquitous Computing* 11, 3, 11(3), 171–191.