

*An experimental study of spatial sound
usefulness in searching and navigating
through AR environments*

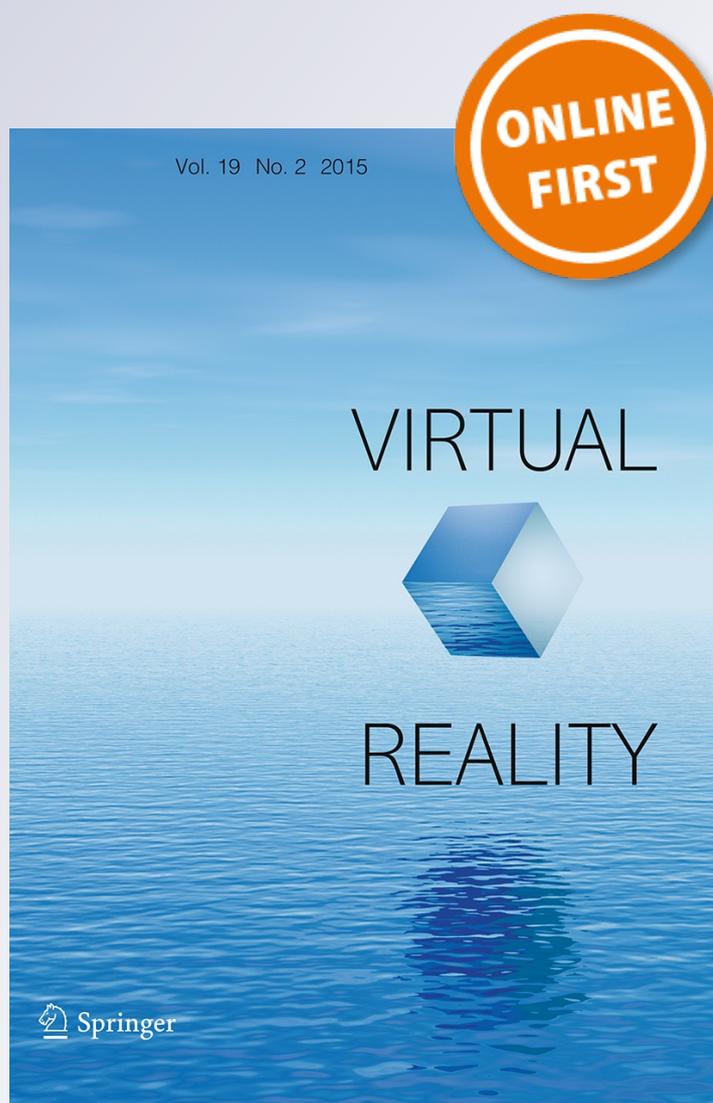
Dariusz Rumiński

Virtual Reality

ISSN 1359-4338

Virtual Reality

DOI 10.1007/s10055-015-0274-4



Your article is published under the Creative Commons Attribution license which allows users to read, copy, distribute and make derivative works, as long as the author of the original work is cited. You may self-archive this article on your own website, an institutional repository or funder's repository and make it publicly available immediately.

An experimental study of spatial sound usefulness in searching and navigating through AR environments

Dariusz Rumiński¹ 

Received: 5 March 2015 / Accepted: 27 August 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract This paper presents an experimental study of spatial sound usefulness in searching and navigating through augmented reality environments. Participants were asked to find three objects hidden within no-sound and spatial sound AR environments. The experiment showed that the participants of the spatialized sound group performed faster and more efficiently than working in no-sound configuration. What is more, 3D sound was a valuable cue for navigation in AR environment. The collected data suggest that the use of spatial sound in AR environments can be a significant factor in searching and navigating for hidden objects within indoor AR scenes. To conduct the experiment, the CARE approach was applied, while its CARL language was extended with new elements responsible for controlling audio in 3D space.

Keywords Spatial sound · Augmented reality · 3D sound in AR · Audio augmented reality · AR · HCI · CARE · CARL

1 Introduction

Augmented reality (AR) technology enables superimposing computer-generated content, such as interactive 2D and 3D multimedia objects, in real time, on a view of real objects. Widespread use of AR technology has been enabled in the recent years by the remarkable progress in consumer-level hardware performance—in particular—in the

computational and graphical performance of mobile devices and the quickly growing bandwidth of mobile networks. Education, entertainment, e-commerce, and tourism are examples of application domains in which AR-based systems are increasingly being used. Since, in AR systems, synthetic content can be superimposed directly on a view of real objects, AR offers a powerful tool for presentation of different kinds of contextual information.

While scientists are focused on engineering techniques of recognizing and tracking patterns as well as overlaying 2D and 3D models onto real environments, methods of modeling spatial sound have not been sufficiently addressed in the context of creating AR applications, especially in the mobile domain. What is more, spatial sound can play a significant role in human–computer interaction (HCI) systems including augmented reality applications. Researches reported that presence of spatial sound in AR applications improves efficiency in performing tasks (Billinghurst et al. 1998; Zhou et al. 2004), enhances depth perception of AR scenes, contributes to the AR experience (Zhou et al. 2004), and helps to identify various 3D objects placed in AR scenes (Sodnik et al. 2006; Zhou et al. 2004).

Cohen and Wenzel (1995) extensively explain multidimensional sound techniques including VR and to some extent AR domains. On the other hand, issues related to design, development, evaluation, and application of audio AR systems including perceptual studies are well summarized by Mariette (2013). Audio technologies that can be used to build immersive virtual environments and gaming are discussed by Murphy and Neff (2011). The presented review takes into consideration the widely varying levels of audio sophistication, which is being implemented in games. Authors present various audio capabilities that are important while implementing 3D sound in virtual environments. Playing multiple sounds at the same time,

✉ Dariusz Rumiński
ruminski@kti.ue.poznan.pl

¹ Department of Information Technology, Poznań University of Economics, Al. Niepodległości 10, 61-875 Poznań, Poland

looping, volume modulation, or audio balancing are examples of 3D sound techniques applied in gaming.

Not only 3D sound is important in gaming, but also in a music industry. Stampfl (2003) combined AR with 3D sound to create immersive 3D musical presentations. The augmented reality disk jockey (AR/DJ) system allows two music DJs to play various sound effects and place them anywhere in 3D space in a music club. Sound sources are visualized in a 3D model of a dance floor and can be controlled using a pen with visual tracking markers placed on it. While this solution looks interesting and promising to use, no scientific results have been published yet.

Many researches reported that an application of 3D sound in AR applications enhances user experience. For instance, Hatala et al. (2005) presented *ec(h)o*—an AR interface utilizing spatialized soundscapes and employing a semantic web approach to retrieve artefacts' data and spatialized audio. The study was designed to create an interactive audio museum guide. The AR interface tracks the visitor's location and dynamically plays audio information related to soundscapes the visitor is seeing. Authors conducted usability tests of the *ec(h)o* system. Participants were asked to use this system in real conditions and complete a questionnaire. Most of the participants gave navigation and engagement of the audio information a high rank (a mean was 4.0 in a 5-point scale).

Findings, reported by Regenbrecht et al. (2004), showed also that 3D sound was an important element of a video conferencing system (*cAR/PE!*). Authors presented a prototype which comprises live video streams of the participants arranged around a virtual table with spatial sound support. Spatial sound driven by headphones audio hardware was used to indicate different user positions. The task was to decide on the most esthetic out of five car models placed on one side of the virtual meeting room. Following, the participants were asked to complete a questionnaire. The *cAR/PE!* system was rated as easy to use and overall user satisfaction was good. Moreover, a method of exchanging information verbally was also rated as satisfactory. In turn, Poeschl et al. (2013) reported that integrating 3D sound in a synthetic scene leads to higher levels of presence experienced compared to no-sound display. Participants of the experiment were asked to complete a questionnaire including six questions related to audio perception. The process was repeated subsequently with a scene where the audio display swapped from no-sound to spatial sound or the opposite.

While 3D sound can enhance an immersive user experience, performed studies show that an application of spatial sound can improve efficiency in performing tasks. For instance, Billinghurst et al. (1998) presented a wearable augmented reality interface that provides spatialized 3D graphics and audio cues to aid performance in finding objects. In the presented experiment, authors examined how spatial cues affected performance using various configurations

combining spatialized visual and audio information. Authors measured average task completion time of searching information from eight pages of data displayed in various conditions. Authors found that adding spatial audio cues to the body-stabilized display configuration dramatically improved performance in finding target icons.

An improved efficiency in performing tasks was also reported by Zhou et al. (2004). Authors investigated the use of 3D sound in an experimental AR game. Participants of the experiment searched for a virtual princess using visual interface and the same interface enhanced with 3D sound. Participants judged also the relative depth of augmented virtual objects. Experimental data suggest that the use of 3D sound significantly improves task performance and accuracy of depth judgement. The results of the study also indicate that 3D sound contributes to the feeling of human presence and collaboration and helps to identify spatial objects. It is also worth emphasizing that Sodnik et al. (2006) affirmed findings presented by Zhou et al. (2004). Sodnik et al. (2006) presented an evaluation of the perception and localization of 3D sound in a tabletop AR environment. The goal of the research was to explore sound–visual search in a number of conditions, whereby different spatial cues were compared and evaluated. Findings of the experiment indicate that humans localize the azimuth of a sound source much better than elevation or distance. Presented findings show that the distance perception of near sources is poor and in this case an elevation can play an important role to distinguish sound sources.

Localization study on applying spatial sound to virtual reality environments was presented by Jones et al. (2008). Authors reported that users could find routes while using spatial audio cues. This study investigated navigation of three routes in VR environments. Participants of the study performed three evaluation tasks. The goal of each of the tasks was to walk from a start location to a target location along a route. The only directional information provided was the continuously adapted music which was played according to the user's location. The study measured completion rate and mean time to complete each route. Results of the study show that an approach of audio panning to indicate heading direction and volume adjustment for distance to target were sufficient factors to finish tasks.

An overview concerning challenges of augmented audio reality for position-aware services can be found in a technical paper presented by Cohen and Villegas (2011). As the authors state, methods of distance reproduction and tracking the observer's orientation in such AR environments are required to enhance value of AR applications. For instance, Villegas and Cohen (2010) presented a software audio filter—head-related impulse response (HRIR)—which can be used to create distance effects within AR environments. This filter allows dynamic modification of a sound source's

apparent location by modulating its virtual azimuth, elevation, and range in realtime.

This paper presents a study that was designed to test the hypothesis that an application of spatial sound significantly improves performance in searching and navigating through indoor augmented reality environments. To conduct the experiment, a novel approach to model spatial sound in AR environments was used. The approach allows dynamic composition of visual objects with aural objects within AR environments. As a result, a user can detect and navigate to individual sound sources which lead to hidden virtual objects.

The remainder of this paper is structured as follows. Section 2 introduces the concept of contextual augmented reality environments (CARE) and presents an extension of CARL language with audio elements and its implementation. Section 3 describes how spatial sound is integrated with CARE environments. The experimental design of the study is described in Sect. 4. This is followed by an analysis of the gathered data and an interpretation (Sect. 5). Next, Sect. 6 contains a discussion of the experiment. Finally, Sect. 7 concludes the paper and indicates the possible directions of future research.

2 Dynamic contextual AR environments

This section provides an overview of the semantic CARE concept that was used to design an experimental study concerning usefulness of spatial sound in indoor AR environments (Rumiński and Walczak 2014b). The CARE approach avoids fragmentation of AR functionality between multiple independent applications and to simplify integration of various information services, providing visual and aural objects, into seamless, contextual and personalized AR interfaces. CARE handles with updating of AR data without requiring actualization of the AR browser application based on the user context. In the CARE environment, AR presentations are described by three semantically described elements:

- *Trackables*—a set of visual markers, indicating the elements (views) of reality that can be augmented (e.g., paintings, posters);
- *Content objects*—a set of virtual content objects including visual and auditory data provided by the available information/content sources (e.g., 3D models representing objects drawn on paintings, audio, and video);
- *Interface*—a description of the user interface, indicating the forms of interaction available to user (e.g., gestures affecting AR scenes).

These elements are encoded in contextual augmented reality language (CARL) and are sent to a generic *CARL Browser* to compose an interactive AR scene (Rumiński

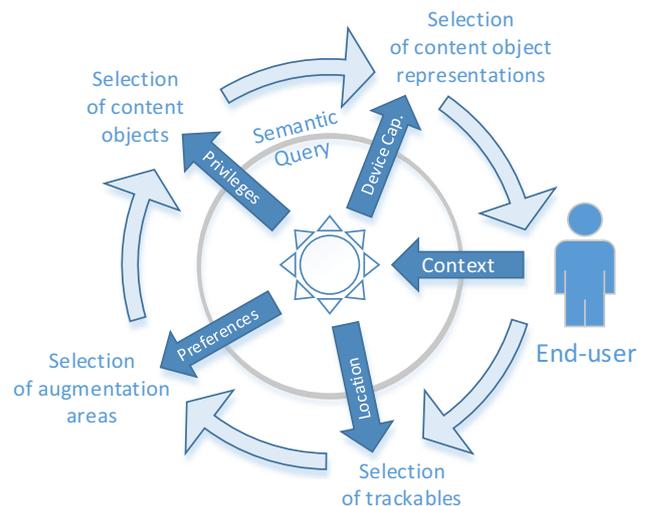


Fig. 1 Dynamic composition of contextual AR scenes in a CARE environment

and Walczak 2014). Examples of CARL descriptions are provided in Sect. 3. In general, the particular elements of a CARE environment are partially independent and may be offered by different service providers in a distributed architecture. Discovery and matching between these elements are possible based on their semantic descriptions with the use of shared ontologies and vocabularies. To enable creation of contextualized AR presentations, the specific elements of AR scenes sent to the client browser can be selected based on the current *context*—user location, preferences, privileges, device capabilities, etc.

Figure 1 depicts a process of dynamic configuration of AR elements to be presented to a user. In each iteration cycle, the AR browser updates context information. Based on the user location and preferences, the visual markers (trackables) and content objects (visual and audio) are selected. Privileges and device capabilities influence the selection of content objects and their representations to be presented to a user. The mapping between context properties and AR elements can differ in different applications: In particular, complex semantic rules are used for selection and modification of these elements (Flotyński and Walczak 2014; Rumiński and Walczak 2014b; Walczak et al. 2014).

2.1 The CARL language

The CARL language has been designed to compose dynamic contextual augmented reality environments consisting of visual content. To provide spatialized audio content, the CARL language has been extended with new audio elements responsible for modeling audio within CARE environment.

To enable conditional steering of various auditory as well as visual content depending on the relative position of

the camera and a real-world object, the *Sector* element was used (Rumiński and Walczak 2014a). The *Sector* element declares an active area in 3D space, which can trigger actions when the camera enters or leaves the sector boundaries. It contains information about what actions should be triggered when the camera enters or leaves the sector, e.g., pan-specific sound. The actions that should be triggered when the camera enters the sector are specified within the *In* element. In turn, the 'leaving' actions are specified within the *Out* element.

The sector boundaries are described by four ranges. The first range (\min_{θ} , \max_{θ}) indicates pitch angle between the camera's position and the reference plane of a trackable object. The second range (\min_{ϕ} , \max_{ϕ}) indicates yaw angle between the camera position and tracked real object. The third range (\min_{Dist} , \max_{Dist}) specifies the minimal and the maximal distance from the camera to the center of the trackable object. This distance is calculated on the basis of the length of the translation vector. The last range (\min_{Height} , \max_{Height}) describes the minimal and the maximal height from the camera to the trackable object.

Listing 1: Shortened description of content object representing the lion's roaring sound.

```

1 <ContentObjects>
2 <ContentObject id="lionFX">
3 <Resources>
4 <Component id="http://.../res/lionFX">
5 <Location details="high" uri=".../lionFX"/>
6 </Component>
7 </Resources>
8
9 <Actions>
10 <Action name="startRoaring">
11 <SetLooping>true</SetLooping>
12 <SetVolume left="5" right="5"/>
13 <Play/>
14 </Action>
15
16 <Action name="stopRoaring">
17 <Stop/>
18 </Action>
19
20 <Action name="setVolume10">
21 <SetVolume left="10" right="10"/>
22 </Action>
23
24 <Action name="setVolume20">
25 <SetVolume left="20" right="20"/>
26 </Action>
27 ...
28 <Action name="moreOnRight">
29 <SetVolume left="30" right="60"/>
30 </Action>
31 ...
32 <Action name="roar">
33 <SetVolume left="99" right="99"/>
34 </Action>
35 </Actions>
36 </ContentObject>
37 ...

```

2.2 Implementation

The *CARL browser* application has been implemented using Qualcomm's Vuforia computer vision library to recognize and track planar images and 3D objects in the real time (Qualcomm 2015). The application is implemented in Java and runs on the *Android* platform (Android 2015). *CARL Browser* processes *CARL* descriptions provided by the distributed REST-based AR services. Each AR service can supply various information including models of 3D objects, audio sources, or specification of possible interaction between a user and synthetic content.

The CARE implementation is based on the REST architectural paradigm where *CARL Browser* can communicate with multiple distributed servers providing *CARL* specifications of particular elements of the CARE environment (trackables, content objects, and interfaces).

Listing 2: Sectors responsible for starting and controlling sounds.

```

1 <Trackable id="landscape">
2
3 <Begin>
4 <ObjectBegin id="lionFX"/>
5 <ObjectBegin id="lion3D"/>
6 ...
7 </Begin>
8
9 <Active>
10 <Sector name="startToPlay"
11 <minTheta="0" maxTheta="180"
12 <minPhi="0" maxPhi="360"
13 <minDistance="0" maxDistance="1500"
14 <minHeight="0" maxHeight="1500">
15 <In>
16 <ObjectAction id="lionFX" action="startRoaring"/>
17 <ObjectAction id="pigFX" action="startSquealng"/>
18 <ObjectAction id="horseFX" action="startNei..."/>
19 </In>
20
21 <Out>
22 <ObjectAction id="lionFX" action="stopRoaring"/>
23 ...
24 </Out>
25 </Sector>
26
27 <Sector name="lionVolume10"
28 <minTheta="0" maxTheta="60"
29 <minPhi="35" maxPhi="50"
30 <minDistance="1400" maxDistance="1500"
31 <minHeight="1212" maxHeight="1300">
32 <In>
33 <ObjectAction id="lionFX" action="setVolume10"/>
34 </In>
35 </Sector>
36
37 <Sector name="lionVolume20"
38 <minTheta="0" maxTheta="60"
39 <minPhi="35" maxPhi="50"
40 <minDistance="1300" maxDistance="1400"
41 <minHeight="1126" maxHeight="1212">
42 <In>
43 <ObjectAction id="lionFX" action="setVolume20"/>
44 </In>
45 </Sector>
46 ...

```

The architecture of CARE's server is based on Spring and Apache CXF frameworks (Spring 2015; Apache 2015).

3 Modeling spatial sound in CARE

Each auditory object is described as a *ContentObject* element consisting of *Resources* as well as *Actions* elements. A *Resources* element points to locations of the same audio element with different levels of quality. The *Actions* element describes actions that can be called on a content object. In this case, actions are responsible for controlling audio (e.g., starting/stopping audio, setting right/left channel's volume, and setting a sound's looping). These actions are triggered by the *CARL Browser*, depending on where the camera is situated in 3D space.

Listing 1 presents the *ContentObject* with an *id=lionFX* representing a lion's roaring. The sound's resource is identified by *uri=http://.../res/lionFX*. The six actions are declared within the *lionFX* content object that can be called by *CARL Browser*. The first action—*startRoaring*—is responsible for setting up and starting the sound effect (lines 10–14). At first, it sets the sound's looping to true to repeat it without any gap between its end and start (line 11). Next, it sets the volume level to five units on *left* and *right* channels (line 12). In the end, it starts to play the sound effect with the *play* command (line 13).

The next action—*stopRoaring*—is responsible for stopping audio content representing lion's roaring sound (lines 16–18). The next two actions, *setVolume10* and *setVolume20*, presented in lines 20–26, set the volume to ten and twenty units, respectively (where one hundred is the maximum volume). Within line 27, there could be written more *setVolumeN* actions where *N* means a sound volume, but the description of listing 1 was specially shortened to only show the idea of constructing volume actions. Next, the *moreOnRight* action increases volume level to 60 on the right channel and 30 on the left channel, giving a sense that the sound's source comes from the right side of a user's camera (lines 28–30). The last action—*roar*—sets the volume for both channels to the maximum value indicating that the user just found the source of sound (lines 32–34). These actions will be used when a user's camera approaches the source of sound to give the effect of depth perception.

Listing 3: Combining spatial sound with visual content.

```

1 <Sector name="lionOnRight"
2   minTheta="0" maxTheta="60"
3   minPhi="15" maxPhi="35"
4   minDistance="0" maxDistance="600"
5   minHeight="0" maxHeight="600">
6 <In>
7   <ObjectAction id="horseFX" action="setVolume70"/>
8   <ObjectAction id="lionFX" action="moreOnRight"/>
9   <ObjectAction id="pigFX" action="little0Left"/>
10  <ObjectAction id="horse3D" action="showHorse"/>
11  ...
12 </In>
13  ...
14 </Sector>
15
16 <Sector name="oppositeToLion"
17   minTheta="0" maxTheta="60"
18   minPhi="35" maxPhi="50"
19   minDistance="0" maxDistance="500"
20   minHeight="0" maxHeight="500">
21 <In>
22   <ObjectAction id="lionFX" action="roar"/>
23   <ObjectAction id="lion3D" action="showLion"/>
24   ...
25 </In>
26   ...
27 </Sector>
28   ...

```

Listing 2 presents a shortened description of the trackable object with declared sectors responsible for starting and controlling sounds. Lines 3–7 present audio as well as visual objects that are associated with the *landscape* object.

The first sector *startToPlay* is responsible for starting sound effects of animals (lines 10–25). When a user's camera enters the sector boundaries, then the application triggers *startRoaring*, *startSquealing*, and *startNeighing* actions using *ObjectAction* elements declared within the *IN* element (lines 15–19). In turn, when the camera leaves the *startToPlay* sector, then the application calls actions declared within the *Out* element—in this case sounds of animals will be stopped (lines 21–24). The next two sectors—*lionVolume10* and *lionVolume20*—are used when a user approaches the lion's sound source (lines 27–45). In this case, decreasing the distance causes the volume of lion's sound to be turned up.

Listing 3 describes sectors that trigger actions responsible for combining spatial sound with visual content. The *lionOnRight* sector declares actions that will be called when a camera points to the center of a trackable object from the right side and the distance between the camera and the trackable object is small (lines 1–12). A visual-

ization of being in the *lionOnRight* sector is depicted in Fig. 6b. The first three actions are responsible for giving an effect that the neighing (line 7) is the most audible sound. What is more, the roaring sound is also audible, but in this case only on the right channel (line 8). The squealing sound is little audible on a left channel (line 9), but it does not play an important role in this sector. This construction gives a clue that the lion object can be located on the right side of the horse. The last action (*showHorse*—line 10), called on content object with the *id=horse3D*, shows a 3D model representing the horse object.

The last sector—*oppositeToLion*—contains actions that will be called when the camera points to the right side of the trackable object (lines 16–27). The first action—*roar* (line 22)—is responsible for turning the roaring sound up. This action is called on content object with the *id=lionFX*. The next action—*showLion* (line 23)—triggered on content object with *id=lion3D*, is responsible for showing a 3D model of the lion object.

4 The experiment

In this section, a study investigating whether 3D sound can be valuable for localizing hidden objects and navigating between them through indoor augmented reality environments, even if visual objects are not visible in an AR interface, is presented. Following sections describe the design of the study, the characteristics of participants that took part in the study, as well as the experiment’s procedure.

4.1 Design of the study

To conduct the experiment, two CARE environments—*Env1* and *Env2*—were designed. In both environments, three content objects representing 3D models of animals augment a real-world object.

The first environment—*Env1*—does not use sound-based contents. The *Env2* environment uses a spatial auditory information to give a user audio cues, which can help to search and navigate to hidden content objects. The context of a user was static—it did not change in a runtime and the user was not able to change preferences and privileges while performing a task. The reader can refer to the video¹ presenting how audio was controlled while conducting the experiment.

3D models of animals are hidden in 3D space and only appear when the user’s camera is located in one of the following sectors: *oppositeToPig*, *horseOnRight*, *pigOnLeft*, *oppositeToHorse*, *horseOnLeft*, and *oppositeToLion*.

Figure 2 visualizes active sectors of the *Env2* environment, in which various audio contents are turned on with

varying sound volume switched between different channels. The first space, the green hemisphere, presents boundaries of the *startToPlay* sector (described in Sect. 3, list 2). When the user’s camera enters this sector, *CARL Browser* triggers actions declared within the *In* element of the *startToPlay* sector. In this case, *CARL Browser* starts to play audio giving a sense that various animals are making noise, but ‘being’ in this sector it is hardly possible to locate positions of hidden objects. When the camera leaves, this sector sound effects will be stopped: *CARL Browser* will trigger actions declared within the *Out* element.

The colored triangular pyramids represent spaces in which a user can hear animals’ sounds, such as a pig’s squealing (the pink color), a horse’s neighing (the gray color), and a lion’s roaring (the yellow color). Every single gradient visualizes the sound intensity level which dynamically increases when the user’s camera moves closer and decreases, when camera moves away. For instance, when the camera enters the *pigVolume10* sector, the volume of the squealing sound is set to ten. Moving forward, *CARL browser* sets volume to twenty, thirty, and forty until the camera enters the loudest sector—*oppositeToPig*—where squealing sound is the most audible.

The next interesting sector is *horseOnRight*, which is responsible for playing the squealing of pig as well as the neighing of horse. The sound volume of neighing is set quieter relative to the sound volume of squealing sound. Additionally, the neighing sound is more audible on the right channel giving a sense that on the right side of the pig object is situated the horse object, even if the camera does not point to the center of the scene, in which the horse object is hidden. Conversely to the *horseOnRight* sector, entering the *pigOnLeft* sector causes that *CARL Browser* automatically pans the neighing sound to be more audible than squealing sound. However, in this case, the squealing effect is more audible on the left channel. The same thing happens when the camera enters the *lionOnRight* or *horseOnLeft* sectors, but in these cases, appropriate sounds are dynamically panned between the left/right channels depending on where the camera is situated in the 3D space.

4.2 Task

Participants were randomly divided into two separate groups—the control group and the spatialized sound group. The users from each group were asked to find hidden objects within AR environment. The participants of the control group used the *CARL Browser* application with configured no-sound environment. The users of the spatialized sound group were also using *CARL Browser*, but in this case, with configured audio environment. The participants were also equipped with Sennheiser HD 202 headphones.

¹ <https://youtu.be/XAzRyHin9io>.

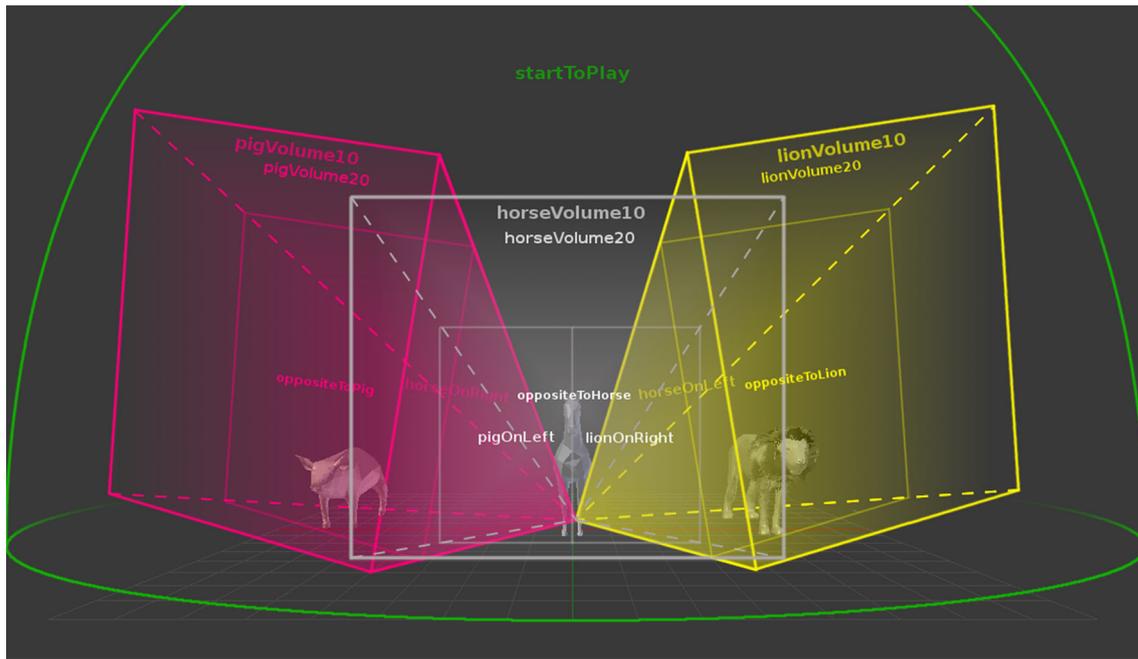


Fig. 2 Visualization of modeled audio sectors

Table 1 Participants' characteristics

Groups	Control	Spatialized sound
<i>N</i>	7	9
Age	21.9 ± 2.67	22.4 ± 1.51
Female	3 (43 %)	2 (22 %)

4.3 Participants

A total of 16 test subjects (5 female and 11 male) participated in the study. The subjects ranged in age from 20 to 27 years ($M = 22.19$ years, $SD = 2.04$ years). Table 1 presents participants' characteristics. All the participants reported normal or corrected-to-normal eyesight and hearing.

Figure 3 presents the user equipped with headphones performing the task. The user's camera is situated in the *oppositeToPig* sector, in which the user interacts with visual and auditory contents. In the presented example, perceiving various sound effects, dynamically panned between channels with different sound intensity dependent on the camera's position, helps to search for hidden objects and navigate through the AR environment.

4.4 Procedure

After a brief introduction to the experiment, the participants were asked to provide information including their age, gender, hearing, and eyesight disabilities. Next, the



Fig. 3 Performing the task using spatial sound environment

participants were given instructions regarding what is a sector in context of CARE environment, how an exemplary sector works, and what exemplary conditions must be met to recognize that user's camera has just entered a sector's boundaries by *CARL Browser*. After that, a mobile device with running *CARL Browser* and configured test environment was given to the users to get used to the task and the application. Lastly, the participants from the control group ran the AR application with configured no-sound environment. The participants from the spatialized sound group started the AR application with configured audio environment.

During the experiment, the task completion time was measured. The measurements began when a participant started performing the task and were stopped when all hidden objects were discovered. After that, the users were asked to complete a short questionnaire.

5 Results and analysis

5.1 Task completion time

While conducting the experiment, the time taken to complete the task was measured for all the participants. The mean time taken to complete the task, as presented in Fig. 4, was analyzed using an independent t test. The two-sample unequal variance (heteroscedastic) test was performed.

The test confirmed that the significant differences in the task completion times existed between control and spatialized sound groups: $T = 2.89, p < 0.05$. The critical value is 2.145 for a 95 % confidence interval and 14 degrees of freedom ($df = 14$).

The task completion time was significantly shorter for the participants of the spatialized sound group (Fig. 4). The perception of auditory events such as hearing various sounds of animals with dynamically changing sound volume and panning audio, depending on where the camera is situated, might have caused the participants to be more focused on localizing hidden objects, which as a result shortened task completion times. For instance, the participants carefully decreased distance to the trackable object, having the right angle, so that at least one sound object was increasing its volume. As a result, the participants did not want to lose auditory cues by using random camera's movements. The participants could assume that in some specific location, a hidden object could be placed, even if it was not visible in the application.

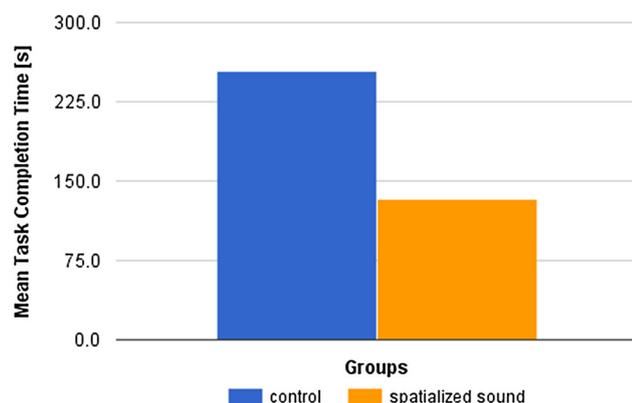


Fig. 4 Mean time taken to complete the task for the control and spatialized sound groups

Conversely, the task completion time was significantly greater for participants of the control group. Participants used no-sound environment. No aural cues might have caused the participants to test various randomly chosen distances, angles, and heights to find hidden object, which as a result might have given greater task completion times for the participants of the control group.

5.2 Evaluation of the spatial sound usefulness

After completing the task, each participant completed a questionnaire with a 1–10 rating scale that was used as a measurement of the spatial sound usefulness in searching and navigating through the augmented reality environment. Table 2 was used in rating the use of spatial sound within augmented reality environment.

The results of an independent t test with the 95 % confidence interval confirmed the existence of a significant difference between the ratings of usefulness spatial sound in two groups: $T = 3.28, p < 0.05$. Fig. 5 presents ratings of spatial sound usefulness given by the participants of both groups.

The participants of the control group were asked to rate whether the use of spatial sound would help in searching for hidden objects and navigating through the AR environment, if they could use it. Most of the participants

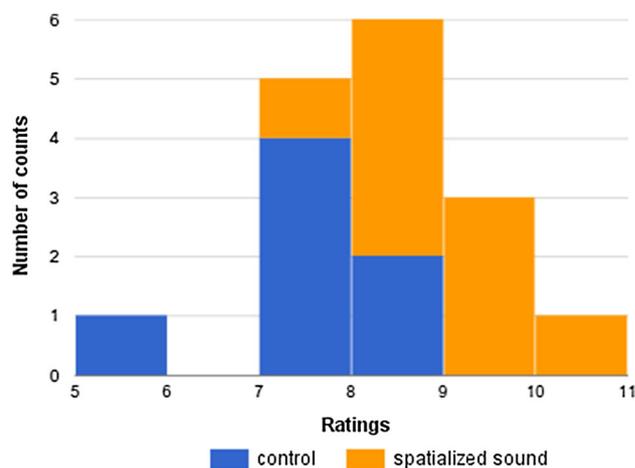


Fig. 5 The ratings of the spatial sound usefulness

Table 2 A scale of 1–10 used in rating the usefulness of spatial sound

Meaning	Rating
Useless	1
Slightly useful	2
Useful	3
Very useful	4
Critical	5
	6
	7
	8
	9
	10

answered that the use of spatial sound would be very useful in locating and navigating hidden objects in such AR environment. Only one person rated the use of spatial sound on five. In total, seven participants ($M = 7.0$) indicated that the use of spatial sound could improve the task performance for the participants in the control group.

The participants of the spatialized sound group were asked to rate whether the use of spatial sound helped them in seeking out hidden objects and navigating between them. The use of spatial sound was very useful for five of the participants. What is more, for four of them, spatial sound was a critical factor to find hidden objects. In total, nine participants ($M = 8.44$) indicated that the use of spatial sound was a significant factor to solve the task for the participants of the spatialized sound group.

5.3 Evaluation of searching and navigating methods

This evaluation was only applied to the participants of the spatialized sound group. Before starting the experiment, the participants were asked to identify what is being heard. Next, the participants started searching for hidden objects. When the camera entered the *startToPlay* sector, the participants answered that they heard animals.

The participants limited the area of search by slightly shortening the distance to the part of trackable object telling which object was being sought. Depending on where the camera was situated, all the participants gave the correct answer, e.g., when the camera was in one of the horse's sectors, the participant answered that the horse was being sought. Shortening distance to the particular part of trackable object caused the volume of the aural content object to increase. These audio hints, as a result, were limiting boundaries of the searching space.

The questionnaire included a question asking about the usefulness of audio panning technique while recognizing the direction of possible location of hidden objects. Six of participants (67 %) answered that the panning technique

gave a cue to navigate to possible location of hidden object. For the rest (33 %), this technique was distracting while searching for hidden objects.

6 Discussion

The collected data suggest that an application of spatial sound within indoor AR environment can significantly improve search as well as navigation to find hidden virtual objects. These results affirm that an application of 3D sound in AR applications improves efficiency in searching tasks (Billinghurst et al. 1998; Zhou et al. 2004). What is more, the use of spatial auditory information helped the participants to identify hidden objects and navigate between them.

The experiment showed that the participants of the spatialized sound group performed faster and more efficiently than the participants of the control group. A clear strategy was used by the participants relying on perceiving sounds in 3D space. When the participants of the spatialized sound group were shortening the distance, at the right angle, to the particular part of the trackable object, they heard that the volume of the audio object was changing. These audio cues resulted in limiting boundaries of the searching space. The participants could assume that the hidden object could be located in particular parts of 3D space.

Moreover, when a single object was found and the camera was within the *pigOnLeft* sector, the participants could assume that on the left side of the horse the pig object could be located, though it was not visible in the AR interface (as presented in Fig. 6a when only the horse object is visible). As a result, most participants navigated to the left side of the trackable object to seek the pig. The same reaction was when, e.g., the camera entered boundaries of the *lionOnRight*'s sectors (Fig. 6b). In this case, the participants navigated to the right side of the trackable object to see the lion object. These results confirm that an

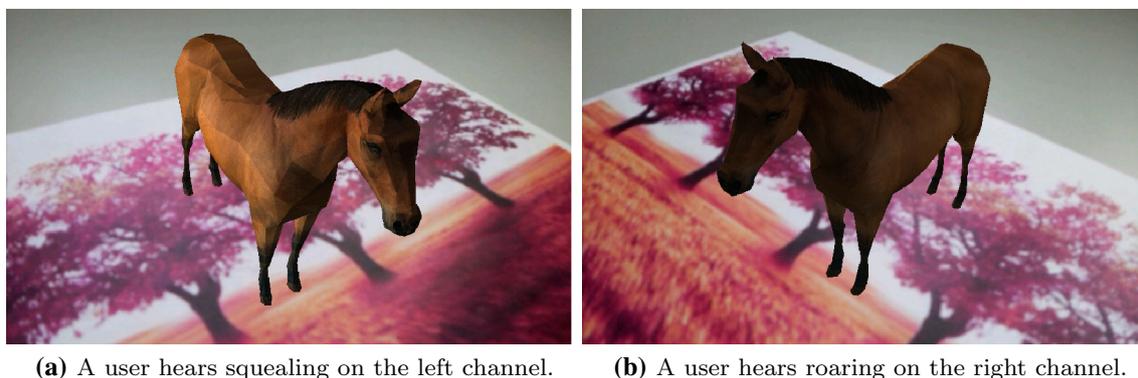


Fig. 6 Depending on various camera positions, a user hears various sounds. For both pictures, the most audible sound is neighing. Additionally, in **a** a user can hear squealing on the left channel, while in **b**, roaring sound is audible on the right channel

approach of audio panning to indicate heading direction and volume adjustment for distance to target can be a valuable navigation cue—not only applied to VR environments (Jones et al. 2008), but also to indoor AR environments.

To carry out the presented experiment, CARE approach was used to model auditory AR environment. *CARE Browser* allowed playing various sounds together, coming from independent sources, and panning audio in an automatic way. What is more, these sounds were automatically panned, depending on where the camera was situated in 3D space. The AR application permitted also to easily manage sound volume. However, *CARL Browser* and *CARL* do not fully support sophisticated spatial sound techniques such as interaural differences of a sound source intensity, time, or direct to reverberant energy ratio for distance estimation. Currently, it is possible to place a sound in a 3D space giving an impression that the sound can be heard in one ear or the other accompanied by overall level adjustments for distance effects taking into account pitch and yaw angles. Although current research findings are promising, additional research is needed. In the presented experiment, three models were equally distributed in the horizontal direction. The study should be repeated with other AR environment configurations, including randomly distributed models in the horizontal as well as the vertical directions. Another limitation was the number of participants. This study should be repeated with more participants to confirm presented results.

7 Conclusions and future work

In this paper, an experimental study of spatial sound usefulness in augmented reality environments was presented. The study explores the possibilities of localizing and navigating in augmented reality environments.

Spatial sound is a great tool which can help in locating and navigating for virtual objects within augmented reality environments. Moreover, the results suggest that the spatial sound may increase a user's perception in three dimensions while searching for hidden objects. Shortening the distance to a particular part of a trackable object, having the right angle, caused the volume of the aural content object to increase, which helped to limit boundaries of the search space. Moreover, 3D sound was a valuable factor to perform the task for the most participants. The participants of the control group, who were using no-sound environment, were at least aware that the usage of spatial sound could be an important factor to perform the task.

To perform the experiment, a novel approach to model spatial auditory information in the CARE environment was used. This approach allows playing various sounds together

coming from independent sources, dynamically panning audio dependent on the camera's position. Apart from that, CARE enables combining sound contents with visual 3D models coming from distributed AR services. The presented approach goes beyond the current state of the art in the field of modeling spatial sound in AR environments. The CARE environment and the CARL language provide a basis for building a generic augmented reality audio browser that takes into account data representing real-world objects, multimedia content (visual as well as audio), various interfaces, and user context coming from independent and distributed sources. The proposed approach avoids fragmentation of AR functionality between multiple independent applications and simplifying integration of various information services into unified, ubiquitous, contextual AR interfaces. The presented design of the study is one example of AR spatialized audio application. The approach to model 3D sound in the CARE environment can be also applied to building AR applications including domains such as gaming, education, entertainment, cultural heritage, and advertising.

Possible directions of future research incorporate several facets. First, a study including randomly distributed models in the horizontal as well as vertical direction should be conducted to confirm the results of the presented research. Second, an evaluation of the spatial sound in a learning process will be examined. Finally, an extension of the CARL language and *CARL Browser* to more sophisticated spatial sound techniques within augmented reality environments should be implemented including interaural differences in intensity, time, and shadowing.

Acknowledgments This research work has been supported by the Polish National Science Centre (NCN) Grants No. DEC-2012/07/B/ST6/01523.

Conflict of interest The author declares that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Android (2015) Android Platform Home Page. <http://www.android.com/>
- Apache (2015) Apache CXF Home Page. <http://cxf.apache.org/>
- Billinghurst M, Bowskill J, Dyer N, Morphet J (1998) An evaluation of wearable information spaces. In: Virtual Reality Annual International Symposium, Proceedings, IEEE 1998, pp 20–27. doi:10.1109/VRAIS.1998.658418

- Cohen M, Villegas J (2011) From whereware to whence- and whitherware: augmented audio reality for position-aware services. In: VR Innovation (ISVRI), 2011 IEEE International Symposium on, pp 273–280. doi:[10.1109/ISVRI.2011.5759650](https://doi.org/10.1109/ISVRI.2011.5759650)
- Cohen M, Wenzel EM (1995) Virtual environments and advanced interface design. In: The Design of Multidimensional Sound Interfaces. Oxford University Press, Inc., New York, NY, USA, pp 291–346. <http://dl.acm.org/citation.cfm?id=216164.216180>
- Flotyński J, Walczak K (2014) Conceptual knowledge-based modeling of interactive 3d content. Vis Comput. doi:[10.1007/s00371-014-1011-9](https://doi.org/10.1007/s00371-014-1011-9)
- Hatala M, Wakkary R, Kalantari L (2005) Rules and ontologies in support of real-time ubiquitous application. Web Semant 3(1):5–22. doi:[10.1016/j.websem.2005.05.004](https://doi.org/10.1016/j.websem.2005.05.004)
- Jones M, Jones S, Bradley G, Warren N, Bainbridge D, Holmes G (2008) Ontrack: dynamically adapting music playback to support navigation. Pers Ubiquitous Comput 12(7):513–525. doi:[10.1007/s00779-007-0155-2](https://doi.org/10.1007/s00779-007-0155-2)
- Mariette N (2013) Human factors research in audio augmented reality. Springer, New York. doi:[10.1007/978-1-4614-4205-9_2](https://doi.org/10.1007/978-1-4614-4205-9_2)
- Murphy D, Neff F (2011) Spatial sound for computer games and virtual reality. IGI Global, Hershey
- Poeschl S, Wall K, Doering N (2013) Integration of spatial sound in immersive virtual environments an experimental study on effects of spatial sound on presence. In: Virtual Reality (VR), 2013 IEEE, pp 129–130. doi:[10.1109/VR.2013.6549396](https://doi.org/10.1109/VR.2013.6549396)
- Qualcomm (2015) Vuforia. <https://www.qualcomm.com/products/vuforia>
- Regenbrecht H, Lum T, Kohler P, Ott C, Wagner M, Wilke W, Mueller E (2004) Using augmented virtuality for remote collaboration. Presence 13(3):338–354. doi:[10.1162/1054746041422334](https://doi.org/10.1162/1054746041422334)
- Rumiński D, Walczak K (2014) Carl: a language for modelling contextual augmented reality environments. In: Camarinha-Matos LM, Barrento NS, Mendonça R (eds) Technological innovation for collective awareness systems, vol 432., IFIP advances in information and communication technologySpringer, Berlin, pp 183–190. doi:[10.1007/978-3-642-54734-8](https://doi.org/10.1007/978-3-642-54734-8)
- Rumiński D, Walczak K (2014a) Dynamic composition of interactive ar scenes with the carl language. In: The 5th International Conference on Information, Intelligence, Systems and Applications, IEEE, pp 329–334. doi:[10.1109/IISA.2014.6878808](https://doi.org/10.1109/IISA.2014.6878808)
- Rumiński D, Walczak K (2014b) Semantic contextual augmented reality environments. In: The 13th IEEE International Symp. on Mixed and Augmented Reality (ISMAR 2014), IEEE, pp 401–404. doi:[10.1109/ISMAR.2014.6948506](https://doi.org/10.1109/ISMAR.2014.6948506)
- Sodnik J, Tomazic S, Grasset R, Duenser A, Billingham M (2006) Spatial sound localization in an augmented reality environment. In: Proceedings of the 18th Australia Conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments, ACM, New York, NY, USA, OZCHI '06, pp 111–118. doi:[10.1145/1228175.1228197](https://doi.org/10.1145/1228175.1228197)
- Spring (2015) Spring home page. <http://www.spring.io/>
- Stampfl P (2003) Augmented reality disk jockey: Ar/dj. In: ACM SIGGRAPH 2003 Sketches and Amp; Applications, ACM, New York, NY, USA, SIGGRAPH '03, pp 1. doi:[10.1145/965400.965556](https://doi.org/10.1145/965400.965556)
- Villegas J, Cohen M (2010) Hrir : modulating range in headphone-reproduced spatial audio. In: Proceedings of the 9th ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry, ACM, New York, NY, USA, VRCAI '10, pp 89–94. doi:[10.1145/1900179.1900198](https://doi.org/10.1145/1900179.1900198)
- Walczak K, Rumiński D, Flotyński J (2014) Building contextual augmented reality environments with semantics. In: Virtual Systems Multimedia (VSMM), pp 353–361. doi:[10.1109/VSMM.2014.7136656](https://doi.org/10.1109/VSMM.2014.7136656)
- Zhou Z, Cheok AD, Yang X, Qiu Y (2004) An experimental study on the role of 3d sound in augmented reality environment. Interact Comput 16(6):1043–1068. doi:[10.1016/j.intcom.2004.06.016](https://doi.org/10.1016/j.intcom.2004.06.016)