Multimedia Tools and Applications (2024) 83:63847–63870
https://doi.org/10.1007/s11042-023-17943-w

A new listener-centered directional attenuation sound model for augmented reality environments

Marina Martínez-Cabrejas[1] · Cristina Portalés[1] · Jesús Gimeno[1] · Manolo Pérez[1] · Sergio Casas-Yrurzum[1]

Received: 3 July 2023 / Revised: 14 December 2023 / Accepted: 15 December 2023 / Published online: 10 January 2024

Abstract

Augmented Reality (AR) involves the combination of synthetic and real stimuli, not being restricted to visual cues. For the inclusion of computer-generated sound in AR environments, it is often assumed that the distance attenuation model is the most intuitive and useful system for all users, regardless of the characteristics of the environment. This model reduces the gain of the sound sources as a function of the distance between the source and the listener. In this paper, we propose a different attenuation model not only based on distance, but also considering the listener orientation, so the user could listen more clearly the objects that they are looking at, instead of other near objects that could be out of their field of view and interest. We call this a *directional attenuation model*. To test the model, we developed an AR application that involves visual and sound stimuli to compare the traditional model versus the new one, by considering two different tasks in two AR scenarios in which sound plays an important role. A total of 38 persons participated in the experiments. The results show that the proposed model provides better workload for the two tasks, requiring less time and effort, allowing users to explore the AR environment more easily and intuitively. This demonstrates that this alternative model has the potential to be more efficient for certain applications.

Keywords Augmented reality · Spatial sound · Directional attenuation · Evaluation

1 Introduction

Augmented Reality (AR) can be seen as a transversal technology since it is applicable in multiple scenarios and in multiple areas of knowledge, including education/learning [13, 38, 44], entertainment [24, 34, 41], cultural heritage [19, 28, 40], surgery [53],Edwards et al. 2021; [3], engineering [14, 36, 39], etc. The term AR was first coined by Caudell and Mizell [9] to describe a display used by aircraft electrical technicians that mixed virtual

✉ Sergio Casas-Yrurzum
Sergio.Casas@uv.es

[1] Institute of Robotics and Information and Communication Technologies, Universitat de València, València, Spain

graphics with physical reality. At that time, the definition of AR was linked to a specific type of displays (Head-Mounted Displays, HMD), and was focused only on visual information. It was a little later when AR applications were developed for other types of displays, so the definition of AR was extended and separated from display technology [42]. Currently, the definition set out in the work of Azuma [1] is followed, which defines AR as systems that have the following three characteristics: 1 It combines real and virtual objects,2 It is interactive in real time; 3 It is registered in 3D. These properties can be applied to different stimuli, not only visual cues.

Although the current definition of AR is not restricted to visual stimuli, in the last decades the research in this technology has been focused mainly on pattern recognition and on tracking techniques as well as on overlaying 2D and 3D models onto real environments [47]. This may be due to the technical difficulties of proper visual integration of the virtual and real worlds in real time, which implies, for example, an accurate registration of the position and orientation of the virtual objects with respect to the camera and a solution for occlusions. For that reason, research on other perceptual cues in AR has been limited. Still, the consideration of other stimuli, such as sound, might bring important benefits.

Indeed, researchers have reported that the integration of sound in AR environments significantly improves the accuracy of depth judgment and improves task performance, also suggesting that it contributes significantly to the feeling of human presence and collaboration [58, 59]. Thus, it can lead to strong AR illusions, contributing to the sense of immersion [20]. The consideration of sound stimuli can also shorten the task completion time and improve the efficiency [6, 30, 47, 58], while it can be a significant factor in searching, identifying and navigating for hidden objects within AR scenes [47, 50, 59]. It has also been reported that, on the contrary, when sound component is absent from a system, then participants can easily become isolated from the environment [33], and that the visual display bears some weaknesses that the audio modality can overcome, such as limited screen space, overload of information, vulnerability to sunlight, and the necessity for constant attention [45]. Another study states that it is relevant to 'see' the sound source in order to give the participants the psychological impression that the sound source exists [57].

Despite the many benefits of integrating sound in AR environments, most AR applications do not incorporate a sound component and/or the sound design is given a marginal role, and thus more research is needed in this field [4, 10, 33, 46, 47]. When sound is implemented, it is generally assumed that the distance attenuation model is the most intuitive and useful system for all users, regardless of the characteristics of the environment. This model reduces the gain of the sound sources as a function of the distance between the source and the listener.

In this article, we challenge this assumption and propose a different attenuation model in which instead of applying gains to the sound sources based solely on the distance, we considered also an attenuation based on the listener orientation. In this listener-centered attenuation model, we assume that the listener's and the viewer's frame of reference will coincide, so we can use the position of the virtual camera as the listener's position and the viewing direction as the listener's orientation. Our model calculates sound attenuation based on a combination of the distance from the listener to the sound source and a focus angle, in such a way that only those sound sources that are close to the camera viewing direction will be perceived by the user. We call this a *directional attenuation model*. It is relevant to comment on the aura-based interactional teleconferencing model developed by [21], where audio interaction is sensitive to both the distance and the relative orientations of the objects involved. In this sense, our model is similar, albeit with a different spatial approach. Both systems are flexible in terms of parametrization, but Greenhalgh and Benford's model is intended on providing a natural environment for the spatial mediation of conversation, while ours is not focused on providing natural sound cues. Our system defines a sound model intended to enhance performance, aid navigation and interaction within the virtual environment.

This approach could be very important in handheld AR systems where the user's view is very narrow and is restricted by the field of view (FOV) of the device's camera. In these cases, the user is likely to be more interested in the objects that he/she is looking at than those that are near but invisible because they are outside the narrow FOV. Moreover, due to the narrow FOV, users must maintain a certain distance from the virtual object to visualize it in its entirety, and in this case, users would like to continue listening to the object even if they are a little farther away, so a sound model based only on distance seems not sufficient to achieve a satisfactory experience. It must be noted that our system creates an inherently unnatural perception of sound, intended to aid navigation and interaction within the virtual environment. For instance, it could help avoid possible errors in perception in case of multiple sound sources –e.g., in a museum, where the pieces are placed next to each other–, as users can focus on either one or the other. Thus, it can be understood in terms of 'gaze' and 'focus' models, referring to where the user is looking at and what is the angle of audition, respectively.

In order to test the proposed model, we have developed an AR application that involves visual and sound stimuli with these two attenuation models. We investigate the differences in usability, workload and task performance between the two attenuation models. Our contribution is two-fold. In the first place, we propose a new sound model for AR applications, which is also easy to implement and computationally cheap. To the best of our knowledge, our system is the first one to mimic real-time spot lighting to attenuate sound in an AR environment. Second, we experimentally compare the new model against the classical one, demonstrating that this alternative model has the potential to be more effective for certain applications.

With these two models in mind, we are interested in knowing if the new model is suitable for AR applications in which sound is an important factor and there could be multiple sound sources playing sound simultaneously. In such AR applications, we hypothesize that:

- Hypothesis 1 (H1). Users will need less time to identify a virtual sound when it is attenuated using the new directional model than with the classical distance-based attenuation model.
- Hypothesis 2 (H2). Users will make more errors when trying to identify a virtual sound with the classical distance-based attenuation model than using the new directional model.
- Hypothesis 3 (H3). Users will be better able to focus their attention on items of interest using the new directional model than using the distance-based attenuation model.
- Hypothesis 4 (H4). Usability and workload will be improved with the new directional model.

The reminder of this article is organized as follows. The next section presents related work, focusing in previous works on AR applications that involve sound stimuli and commenting about the sound attenuation models. In Sect. 3, we describe the directional attenuation model and how it can be integrated in an AR application. Section 4 describes a comparative experimental assessment between the new model and the classical one, whereas Sect. 5 discusses the results of such evaluation. Finally, Sect. 6 presents conclusions, describes the limitations and proposes further work.

2 Related work

The auditory perception of people is omnidirectional since pressure waves propagate in all directions. In addition, people hear with more intensity sound sources as they get closer to them, while also being able to identify their direction. In many multimedia-based applications, sound models try to emulate the reality, where sound is output to the user using either headphones or loudspeakers, which can play sound according to different characteristics. For instance, considering a pair of sound players, stereo panning reliably positions sound to the left or right of a listener, while variations in intensity can simulate changes in the distance between a source and a listener [54, 55]. On the other hand, binaural 3D audio algorithms that make use of specific filters, or HRTF's (Head-Related Transfer Functions) [5], allow more accurate localization of a sound source around the user. Some works also deal with room acoustics to simulate how audio waves propagate through space by bouncing from different surfaces [18, 31, 56]. For instance, [31] propose an approach to answer the question 'can one hear the size of a room?'. Such approach only requires a single speaker and an arbitrarily placed microphone, opening new directions in the field of multimodal scene perception and visualization for AR. And [56] propose and evaluate an integrated method for 3D binaural audio with synthetic reverberation. The approach brings interactive auralization using vector base amplitude panning and a scattering delay network. The rendering model also allows direct parameterization of room geometry and absorption characteristics. In this sense, it is worth mentioning audio spatializers. They use "physical" characteristics of a scene, such as the distance and angle between a listener and an audio source, to modify the properties of sound transmitted to the user. For instance, the Unity audio engine supports spatialization through plugins from, e.g., Google, Oculus or Microsoft (Unity [52].

Regarding spatial audio for AR applications, the academic literature refers to the concept of Augmented Audio Reality (AAR) [12], which is to present an overlay of synthetic sound sources upon real world objects, also displayed in real time and registered in 3D. In this way, the 3D sound (or spatial sound) is the sound that seems to come from various directions, creating the effect of a 3D space [58]. Examples of such applications are described in [4, 10, 11, 25, 55]. For instance, [11] presents the results of participants' interactions with an experimental AAR installation in a museum, also following the aura-based model described in [21]. In this work there is a focus on engagement and awareness, whereas our work is not explicitly focused on awareness but on facilitating the understanding and usefulness of the sounds. There are also other works where the AR system is complemented with stereo sound or stereo effects, like in [17, 48, 51], while others do not describe with enough detail the sound model they are using, such as in [38]. In our work, we will test the proposed model with stereo audio, as horizontal navigation is the main use case of AR applications.

Additionally, some AR works that are based on the principle of attaching virtual sounds to locations for exploration and interaction are often composed of two concentric levels of audio feedback, one in wide proximity providing cues that guide users towards it and the other in a narrower activation zone allowing further interaction –although other configurations are possible–. Examples of AR applications that involve sound interaction zones can be found in [26, 46, 54, 60]. For instance, in [26] an AR-based application is presented to explore the use of sound to enhance museum visits. The system allows visitors to express their interests by means of simple and intuitive gestures,to that end, they introduce the notion of the Audibility Zone (AZ which delimits the space in which a particular sound is audible. Also, the work in [60] describes the LISTEN project, a personalized and interactive location-based audio experience based on an adaptive system model. It does this by tracking aspects of the visitors' behavior to assign them a behavioral model and adjust the delivery of audio content accordingly. LISTEN also introduces the concept of the attractor sound, which, based on the visitor's personalized profile model, suggests other nearby artworks to the visitor that may be of interest to them via spatially located audio prompts.

On the other hand, in AR applications that involve vision-based tracking methods where sounds are applied onto augmented objects adhered to fiducial markers, it is quite common that the audio cues are triggered when the objects are located inside the camera FOV. Many examples can be found, as the works described in [27, 32, 38, 48]. For instance, in [48] an AR-based e-magazine consisting of all the planets in different pages is presented. The application works on a smartphone and performs a visual tracking,once the camera is focused on a marker for detection, the related augmented object of the specific planet to the marker is projected and the corresponding audio (stereo effects are played on headphones. In these cases, it has been reported that the requirement of always keeping a marker in the scene all the time limits the user's movement as well as viewpoint and thus restricts the user to a relatively small space and a narrow angle of view [58]. However, this can also be perceived as positive if the user needs to focus on a specific sound stimulus, and thus it might aid not to hear other sound sources that can distract him/her. This is further supported by the fact that hearing attracts attention more easily than vision [43]. The problem with triggering is that sound changes abruptly from inaudible (listener outside the triggering zone) to audible state (listener inside the triggering zone). Differently, in our proposed model, the transition can be smooth and, in fact, it is not necessary that the sound source lies within the FOV.

Additionally, it is quite common that in AR indoor-based environments the space is relatively small. In such condition, [59] pointed out that the attenuation of sound with distance in this small area can be too vague to be perceived, and thus they proposed to exaggerate the intensity difference of 3D sound according to distance (depth) changes of the virtual sound sources. Another option are directional sound sources that can be configured to focus the sound on a specific direction (through the configuration of a dispersion cone) [15]. With this kind of source, the sound could be projected to specific parts of the room, but the user would not hear the sound from other places even if he/she is looking directly to the source. In conclusion, these two attenuation models depend on the user position, but do not take into account his/her point of view. Differently, we propose to apply the attenuation based both on distance and the user orientation (referred as focus angle). This angle could be smaller, equal or larger than the camera FOV, making it very flexible. The mathematical formulation of the model will be explained in the next section.

Finally, it is worth referring to the sound reproduction technology. As commented above, sound is output to the user using either headphones or loudspeakers. However, as pointed by [29], there are some advantages associated with headphone-based spatial sound delivery including the fact that headphones provide a high level of channel separation, thereby minimizing any crosstalk that arises when the signal intended for the left (or right) ear is also heard by the right (or left) ear. In our case, we make use of headphones because we will the test the model with a mobile AR application, and headphones can be directly connected to the smartphone/tablet where the AR application is running, they are relatively cheap and, most important, the sound is played for a single person, not interfering with other prospective users,this allows that different users could simultaneously interact with the augmented environment if running the app in their own smartphones. Other recent works making use of headphones are reported in [10, 17, 47, 48, 54, 55]. For instance, in [47] an experimental study is presented, where the spatial sound usefulness in searching and navigating through AR environments is explored,the participants were equipped with Sennheiser HD 202 headphones. Also, in [17] it is explored how stereo sound affects assembly guidance in an AR environment. For the visual and sound cues, they make use of a HoloLens 2 device, which includes a pair of small speakers close to the ears.

3 Materials and methods

3.1 Model description

As aforementioned, when introducing sound in computer applications, it is quite common to attenuate sound based on the distance from the sound source to the user. The attenuation function based on distance (att_d) could be logarithmic, linear, could follow the inverse law or even the inverse square law of sound. Real-time audio software libraries such as OpenAL, CoreAudio, DirectSound or Unity 3D implement some of these models, usually adding roll-off factors and minimum and maximum distances to clamp the attenuation levels at certain distances. Although sound sources could be directional, this model is omnidirectional; i.e. the listener receives the same amount of sound intensity regardless of his/her orientation. Figure 1 depicts this omnidirectional model, hereafter referred to as the *classical attenuation model.*

Instead, we propose a listener-centered attenuation model in which the listener is not omnidirectional. To build our directional attenuation model, we first define a cutoff angle θ_c between 0 and 180º. This angle is measured from the camera viewing direction

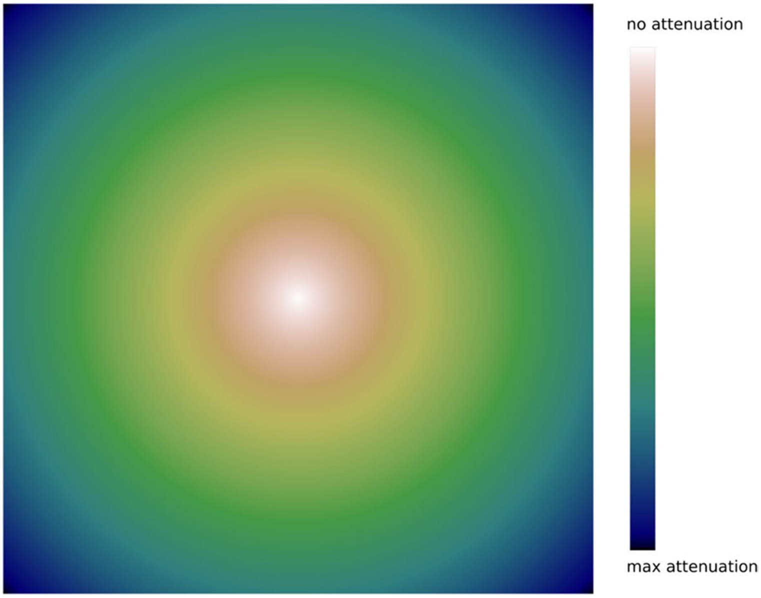

Fig. 1 Heat map of the sound model, where a hypothetic user is located at the center of the map, assuming that the sound is attenuated only by distance (calculated by the inverse law)

and defines a cone outside which attenuation is absolute and the listener is not able to hear sounds. The equations for calculating angular attenuation (att_a) according to our directional model are the following:

$$c = \cos(\theta_c) \tag{1}$$

$$\overrightarrow{v_{ls}} = \frac{\overrightarrow{s_s} - \overrightarrow{s_l}}{\left|\overrightarrow{s_s} - \overrightarrow{s_l}\right|} \tag{2}$$

$$p = \overrightarrow{v_l} \cdot \overrightarrow{v_{ls}} \tag{3}$$

$$\begin{gathered} if(p < c) \\ atta_a = 0 \\ else \\ atta_a = p^{exp} \end{gathered} \tag{4}$$

where exp is a factor that makes the attenuation narrower or broader, $\overrightarrow{v_l}$ is a unit vector representing the direction in which the listener is facing (the viewing direction), $\overrightarrow{s_s}$ represents the position of the sound source and $\overrightarrow{s_l}$ represents the position of the listener. The final result is the value of att_a, shown in in (4), which represents the angular attenuation based on the orientation of the listener, where 1 means no attenuation and 0 means full attenuation. Figure 2 and 3 depict this calculation. This model is inspired by real-time spot lighting (implemented in OpenGL, for instance) in which light attenuation follows a similar procedure.

It is worth noting that the two models can be applied independently (only distance attenuation or only angular attenuation) or simultaneously. In this latter case the final attenuation (att_f) will be:

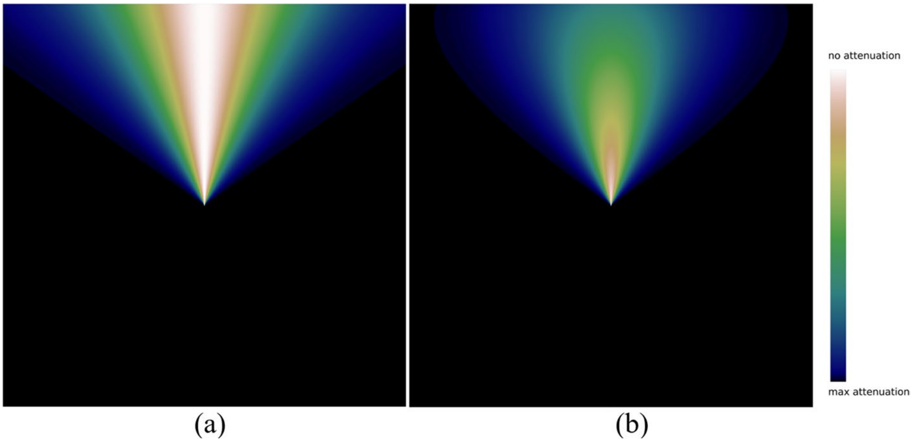

Fig. 2 Heat map of the sound model, where a hypothetic user is located at the center of the maps, assuming that the sound is: (**a**) attenuated only by angle, with an exponent of 10; (**b**) attenuating by distance and by angle, with an exponent of 10. The cutoff angle was set to 180° in both cases

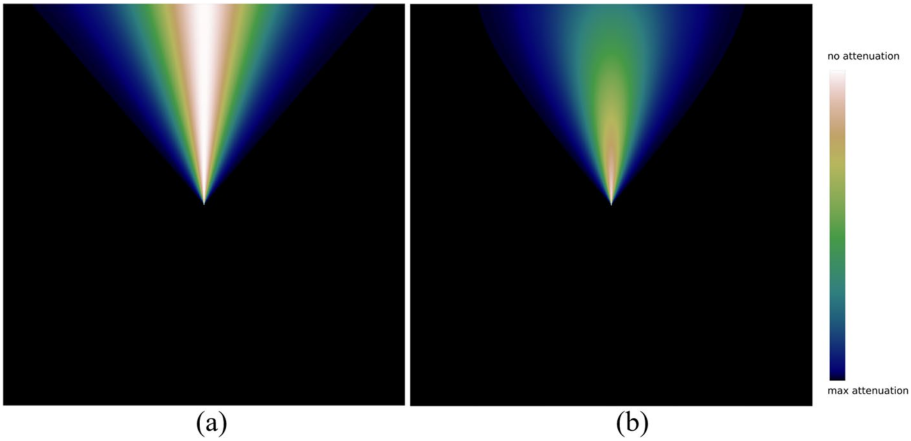

Fig. 3 Heat map of the sound model, where a hypothetic user is located at the center of the maps, assuming that the sound is: (**a**) attenuated only by angle, with an exponent of 20; (**b**) attenuated by distance and by angle, with an exponent of 20. The cutoff angle was set to 180° in both cases

$$att_f = att_d \cdot att_a \tag{5}$$

We choose this option to build the proposed listener-centered directional model since we want sounds to be attenuated when they are far away from the listener, and also when they are not within the listener's focus of attention. Figure 2 b and Fig. 3 b depict this new listener-centered directional model.

3.2 System implementation

In order to verify the validity of the hypotheses raised in this research, we built a handheld mobile AR application in which we implemented the new sound model. The system was implemented using Unity 2020.3.18 together with the Vuforia 10.5.5 library. The application was run on a Samsung Active Tab 3 Android-based tablet with a 13 Mp ARCore-compatible camera and a Razer Barracuda stereo headset. For the sake of simplicity, the AR application was built with Vuforia image targets. Therefore, an image target needs to be printed in order for the virtual objects to be properly placed with respect to the real world. A sheet music target image was created for that purpose. Figure 4 shows a picture of the AR application.

The AR application included two different sound-based tasks. Both tasks were designed to have multiple simultaneous sound sources, on which the user must extract information and make decisions. In Task 1, the user is told to look at a set of virtual movie posters and focus the attention on the poster that he/she believes is reproducing a horror genre movie. The posters were placed vertically on a virtual wall and did not provide visual information about the movie. The information was provided by the music of the movie. In this task, the cutoff angle was setup to 90° and the exponent to 20.

In Task 2, a well-known song (*Bohemian Rhapsody* by English band *Queen*) was played. The song is played on three separate tracks (vocals, piano and violins). Each track is played by a separate virtual loudspeaker and the user must navigate around the augmented scene in order to recognize which of these 3 fixed virtual objects is playing the violins track. The

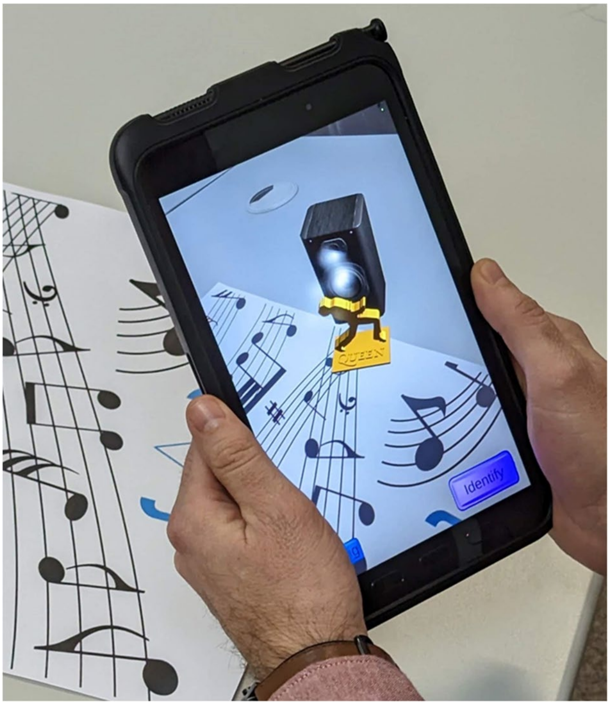

Fig. 4 The AR application

loudspeakers were arranged horizontally in an equilateral triangle shape. In this task, the cutoff angle was also setup to 90º but the exponent was set to 10. Therefore, the effective audible cone in Task 1 is much narrower than in Task 2 because of the exponent.

In both tasks, users provided feedback (answers to complete the task) by pressing a button on the mobile application to communicate that the sound was identified (see Fig. 4). Therefore, we could easily record the amount of time needed to complete the tasks and the number of errors before a successful completion.

In order to implement the new model and be able to compare it with the classical one, we created a component with a C# script that we called *AudioListenerAttenuator*. The script allows to switch from the distance-based attenuation model –implemented by default in Unity– to the new directional attenuation model. This component is applied to the *GameObject* of the AR camera, since we assume that the listener and the viewer will be the same. The script allows also to parametrize the cone used in the proposed attenuation model.

Two Unity scenes were created, one for each task. The first scene is built based on a Vuforia image target with a 3D object in each of its corners that simulates a movie poster. In this scene, each of the four posters has an associated *AudioSource* with the sound of a movie soundtrack. The volume of each *AudioSource* is modified by the *AudioListenerAttenuator* previously described, so that the amount of sound received by the listener could change depending on the distance and the orientation of the listener, provided that the new sound model is chosen. This scene is intended to place the image target on a wall so that the user can explore the movie posters. Figure 5 shows the Unity scenario for Task 1.

The second Unity scene consists of a Vuforia image target containing three 3D objects (speakers) that also emit sound (*AudioSource*). The volume of each *AudioSource*

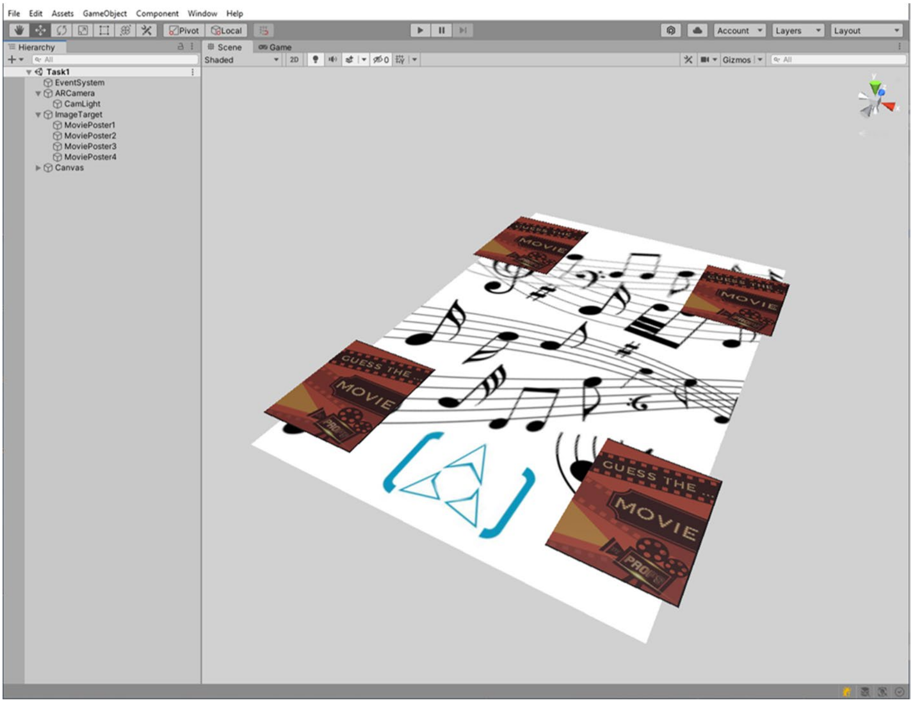

Fig. 5 Unity scenario and object hierarchy for Task 1 with the image target and the virtual posters on it

is also modified by the *AudioListenerAttenuator*. The image target will be placed in the real environment on a table for the user to explore the scene. Figure 6 shows the Unity scenario for Task 2.

Unity's distance-based attenuation model is parameterized based on a minimum distance (d_{min}) and a maximum distance (d_{max}) that are assigned in the *AudioSource* component. The former indicates up to what distance the sound source has a gain of 1. The latter indicates from what distance the sound source is no longer audible, and the gain is 0. In our AR application, the default Unity model was used for all sources, d_{min} was set to 0.1 m and d_{max} was set to 2.0 m.

The *AudioListenerAttenuator* script allows the cone-based directional model to be enabled or disabled, simultaneously allowing the default distance-based gain calculation to continue working. If the directional attenuation model is chosen, the script calculates an angular attenuation value (att_a) based on Eqs. (1), (2), (3) and (4). This attenuation is applied to the *AudioSource* volume to simulate the new *Listener* behavior since, in Unity, the distance attenuation calculation is always active. The final effect is that the total attenuation follows Eq. (5). If, on the other hand, the classical attenuation model is chosen, the volume of the *AudioSource* is not modified and only the default Unity model is applied.

The parameters of the *AudioListenerAttenuator* component that defines the angular attenuation are the *cone angle (cutoff angle)* (θ_c) and the *cone exponent* (*exp*). A third parameter, the *cone direction*, can be set, but the most common option is to copy it from the viewing direction of the camera. Thus, it corresponds to $\vec{v_l}$ in the model's equations.

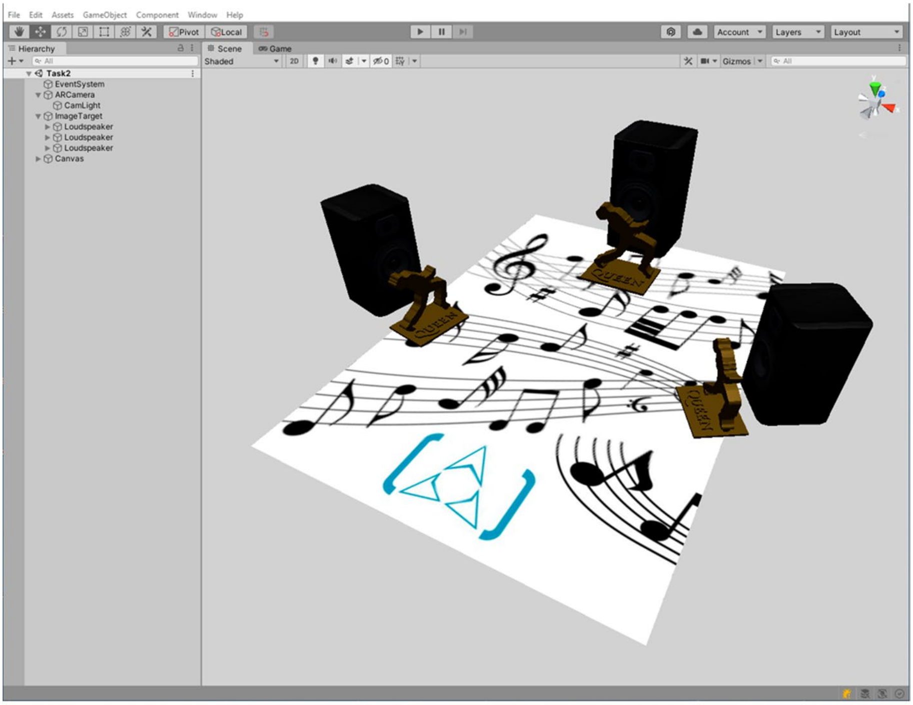

Fig. 6 Unity scenario and object hierarchy for Task 2 with the image target and the virtual loudspeakers

This component stores an array with all the *AudioSources* of the scene to be able to modify the volume of each one.

It is of utmost importance to highlight that although our AR application uses a Vuforia's image target, the sounds are not linked to the marker and or not triggered when the marker is visible. The image target is only used for the 3D registration of the real world with respect to the device camera. In fact, our implementation is ready to be switched to markerless AR systems. Therefore, it could work with no effort in optical see-through AR devices or in SLAM-based mobile AR paradigms.

4 Experimental study

In order to verify the validity of the hypotheses raised in this research, it is necessary to assess the models within the AR application, and measure user performance, usability, workload and subjective preference for either of the models.

4.1 Participants

To that end, 38 people over 18 years of age were recruited for the experimental comparative assessment. They were recruited by social media and were not rewarded in any way for their participation. In Table 1 we can see the statistical distribution of the participants including age, gender and previous experience with AR. As can be seen, there is a majority

Table 1 Statistical distribution of the participants of the experiment

Age	*<20*	*20–29*	*30–39*	*40–49*	*>50*
	2 (5.26%)	29 (76.31%)	3 (7.89%)	2 (5.26%)	2 (5.26%)
Gender	*Female*	*Male*			
	24 (63.16%)	14 (36.84%)			
Previous Experience	*Not at all familiar*	*Slightly familiar*	*Occasionally familiar*	*Moderately familiar*	*Extremely familiar*
	9 (23.68%)	16 (42.11%)	6 (15.79%)	3 (7.89%)	4 (10.53%)
Number of samples = 38					

of young people but most users consider themselves unfamiliar or just slightly familiar with AR technology. None of the participants were professionally engaged in AR-related jobs.

4.2 Data collection

In order to perform the comparative assessment, we tested the two previously defined tasks. We recorded the actions of the users and measured the amount of time and the number of errors done while completing the tasks. Each user had to perform both tasks with both sound models. They were also prompted to fill a usability questionnaire (SUS) [7] and a workload questionnaire (NASA-TLX) [22] for each task completed with each of the two models. They also filled a final comparative questionnaire with five two-choice questions and two open-ended questions in order to provide their subjective opinions on which model they found more appropriate according to different dimensions: preference, recommendation, usefulness, ease and attention. Table 2 shows the final comparative questionnaire. The ten questions of the SUS and the six questions of the NASA-TLX questionnaire can be found in the corresponding references.

It should also be noted that each of the tasks to be analyzed was tested with the two sound models, although not at the same time. That is, the tasks were repeated twice: once with the classical attenuation model and once with the directional attenuation model. The test order was randomized: 19 participants tested the classical model first, and 19 participants tested the proposed model first. Also, the order of the sound models was counterbalanced, since participants testing Task 1 with the classical model first, tested the classical model in second place for Task 2 and vice versa, although the users were not aware about

Table 2 Final two-choice comparative questionnaire

Question	
Q1	Which model did you like the most, in general?
Q2	What model would you recommend in the use of AR applications?
Q3	Which model do you think is more useful in the development of AR applications?
Q4	In which model do you think you had the easiest time recognizing the sound you needed to find?
Q5	Which of the two models has allowed you to better focus your attention?
Q6	Explain/justify your answers
Q7	What differences did you experienced in recognizing the sound you were supposed to find?

the order of the tests. Under no circumstances users knew which sound model was being used in their AR application. In addition, the arrangement of the virtual objects was randomly changed in every test, making impossible for users to find patterns in the correct answers.

4.3 Procedure

The experimental protocol consisted of 5 steps:

1) Users were first briefed into the tasks that they were going to do. They were told that they had to complete each task twice with two different sound models, but they were not told about how the sound models work. They were also warned that each execution of the task would have a different solution because the order of the right answers would be randomized.
2) Then, they were provided with a free-practice application in which virtual sounds were played for 5 min and users had to identify them. None of these sounds were later used in the two tasks of the experiment.
3) Then, the real experiment begun. First, each user had to complete Task 1 with one of the two sounds models (with a randomized counterbalanced order). After completing the task, each user had to complete the SUS and the NASA-TLX questionnaires about Task 1 using that sound model. Then, the task was launched again using the other sound model and upon completion users had to complete again the SUS and NASA-TLX questionnaires. Figure 7 shows a user testing Task 1.
4) After they finished Task 1 with both sound models, the user was prompted to complete the final comparative questionnaire, in order to choose between the first AR experience and the second AR experience, without knowing the internal differences between the two.
5) A similar procedure was carried out for Task 2. Figure 8 shows a user testing Task 2.

As aforementioned, the AR application was also able to gather objective measures about the performance of the users. Two numeric datasets were produced for each execution of the AR application. The first one records the total time the user needed to complete the task. The second one records the number of errors. As a result, we have 8 objective numeric datasets of length 38. Using letter A to refer to the classical sound model and letter B to refer to the new proposed model, we name the objective datasets as T1A (Task 1, time for the classical model), E1A (task 1, number of errors for the

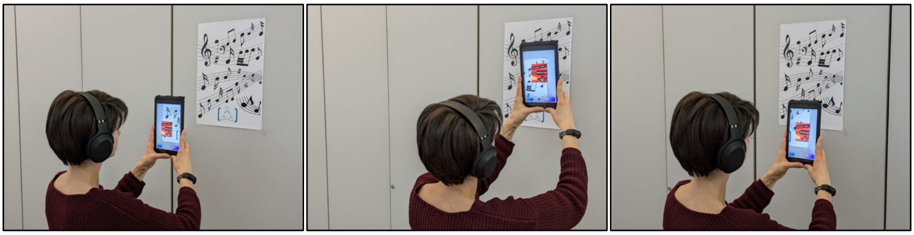

Fig. 7 A user testing Task 1

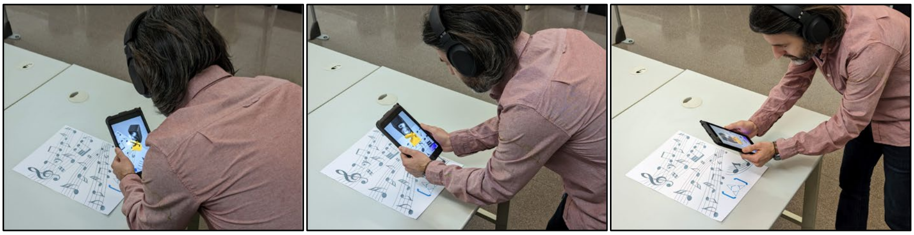

Fig. 8 A user testing Task 2

classical model), T1B (Task 1, time for the directional model), E1B (Task 1, number of errors for the directional model), T2A, E2A, T2B and E2B. Similarly, there are 4 SUS datasets (with 10 questions each) of length 38 (named SUS1A, SUS1B, SUS2A and SUS2B) and 4 NASA-TLX datasets (with 6 questions each) of length 38 (named TLX1A, TLX1B, TLX2A, TLX2B), plus 10 Boolean datasets (Q1, Q2, Q3, Q4, Q5 – for each task) from the two-choice questions of the final questionnaire. Table 3 summarizes the numeric datasets. Since all participants tested two different conditions and they did it in all possible orders, the experiment can be considered to follow a two-treatment within-subjects design with complete counterbalancing.

These datasets were analyzed using IBM SPSS 28.0. First, we checked the normality hypothesis using the Kolmogórov-Smirnov test (Massey Jr 1951) and the Shapiro–Wilk test [49]. Then, we applied parametric and non-parametric paired tests in order to compare the results obtained using model A versus the ones obtained using model B, for both tasks. We also applied a binomial test to the two-choice questions of the final questionnaire. All the analyses were two-tailed and were conducted at the 0.05 significance level, unless otherwise indicated.

Table 3 Numeric dataset summary

Dataset	Description
T1A	Task 1 – Completion time – Model A (classical)
E1A	Task 1 – Number of errors – Model A (classical)
T1B	Task 1 – Completion time – Model B (new)
E1B	Task 1 – Number of errors – Model B (new)
T2A	Task 2 – Completion time – Model A (classical)
E2A	Task 2 – Number of errors – Model A (classical)
T2B	Task 2 – Completion time – Model B (new)
E2B	Task 2 – Number of errors – Model B (new)
SUS1A	Task 1 – SUS score – Model A (classical)
SUS1B	Task 1 – SUS score – Model B (new)
SUS2A	Task 2 – SUS score – Model A (classical)
SUS2B	Task 2 – SUS score – Model B (new)
TLX1A	Task 1 – NASA-TLX score – Model A (classical)
TLX1B	Task 1 – NASA-TLX score – Model B (new)
TLX2A	Task 2 – NASA-TLX score – Model A (classical)
TLX2B	Task 2 – NASA-TLX score – Model B (new)

4.4 Results

In this section we describe the results of applying statistical analysis to both the objective and the subjective datasets collected as a result of the experiments.

First, we passed normality tests to the numeric datasets. Table 4 shows the results of applying both Shapiro–Wilk and Kolmogorov–Smirnov tests. The former is more suitable for this experiment since the number of samples is smaller than 50 [37]. Thus, this test will be used to classify the datasets as normal or non-normal. As can be seen, almost all datasets can be considered non-normal, with the exception of T1B and E1B. E1B is not analyzed because all values are zero. Therefore, it could be considered normal.

First, we analyze the objective datasets (times and errors). As most objective variables cannot be considered normally-distributed, we apply a paired Wilcoxon signed-rank test in order to compare the results from group A (classical sound model) to the results of group B (new proposed model). As two of these datasets could be considered normal, we also apply a parametric paired t-test, in order to complete the analysis. The results are shown in Table 5 and in Table 6. In both tests, it can be clearly seen that there are significant differences in favor of model B in all but one of the objective datasets. The results are consistent across the two tasks, regarding the time needed to complete the tasks. Figure 9 shows a box plot of the time variable for both tasks. With respect to the number of errors, the difference is only significant for Task 2. The results are very similar for both the parametric and the non-parametric tests.

In addition to the objective data previously analyzed, we also collected subjective data in the form of usability and workload tests. Table 7 shows the results of the usability tests for both tasks and both models. According to the mean SUS scores, the new model is perceived as more usable than the classical one. A paired Wilcoxon

Table 4 Normality tests

Dataset	Kolmogorov–Smirnov		Shapiro–Wilk		Decision
	Statistic (D)	Significance	Statistic (W)	Significance	
T1A	0.227	$<10^{-3}$	0.796	$<10^{-3}$	Non-normal
E1A	0.535	$<10^{-3}$	0.302	$<10^{-3}$	Non-normal
T1B	0.150	0.031	0.943	0.051	Normal
E1B	-	-	-	-	Normal
T2A	0.171	0.007	0.723	$<10^{-3}$	Non-normal
E2A	0.463	$<10^{-3}$	0.560	$<10^{-3}$	Non-normal
T2B	0.142	0.050	0.900	0.003	Non-normal
E2B	0.539	$<10^{-3}$	0.237	$<10^{-3}$	Non-normal
SUS1A	0.206	$<10^{-3}$	0.810	$<10^{-3}$	Non-normal
SUS1B	0.188	0.002	0.831	$<10^{-3}$	Non-normal
SUS2A	0.255	$<10^{-3}$	0.752	$<10^{-3}$	Non-normal
SUS2B	0.221	$<10^{-3}$	0.738	$<10^{-3}$	Non-normal
TLX1A	0.134	0.081	0.940	0.042	Non-normal
TLX1B	0.153	0.025	0.886	0.001	Non-normal
TLX2A	0.130	0.108	0.909	0.005	Non-normal
TLX2B	0.149	0.033	0.873	$<10^{-3}$	Non-normal

Number of samples = 38

Table 5 Non-parametric paired Wilcoxon signed-rank tests. Time and errors. Group A vs group B

Dataset	Median ± IQR	Mean rank [neg., pos.]	Rank sum [neg., pos.]	Statistic (Z)	Significance
T1A	27.000 ± 30.750	[21.21, 7.29]	[615.00, 51.00]	4.431	$<10^{-3}$
T1B	16.500 ± 8.250				
E1A	0.000 ± 0.000	[2.00, 0.00]	[6.00, 0.00]	1.732	0.083
E1B	0.000 ± 0.000				
T2A	21.500 ± 17.500	[8.30, 20.67]	[41.50, 661.50]	4.680	$<10^{-3}$
T2B	12.000 ± 8.000				
E2A	0.000 ± 0.250	[0.00, 4.50]	[0.00, 36.00]	2.828	0.005
E2B	0.000 ± 0.000				
Number of samples = 38					

Table 6 Parametric paired t-tests. Time and errors. Group A vs group B

Dataset	Mean ± SD	Statistic (T)	Significance
T1A	42.263 ± 34.913	4.397	$<10^{-3}$
T1B	18.158 ± 7.951		
E1A	0.079 ± 0.273	1.781	0.083
E1B	0.000 ± 0.000		
T2A	26.632 ± 19.838	4.629	$<10^{-3}$
T2B	14.132 ± 7.215		
E2A	0.263 ± 0.503	3.141	0.003
E2B	0.053 ± 0.226		
Number of samples = 38. Degrees of freedom = 37			

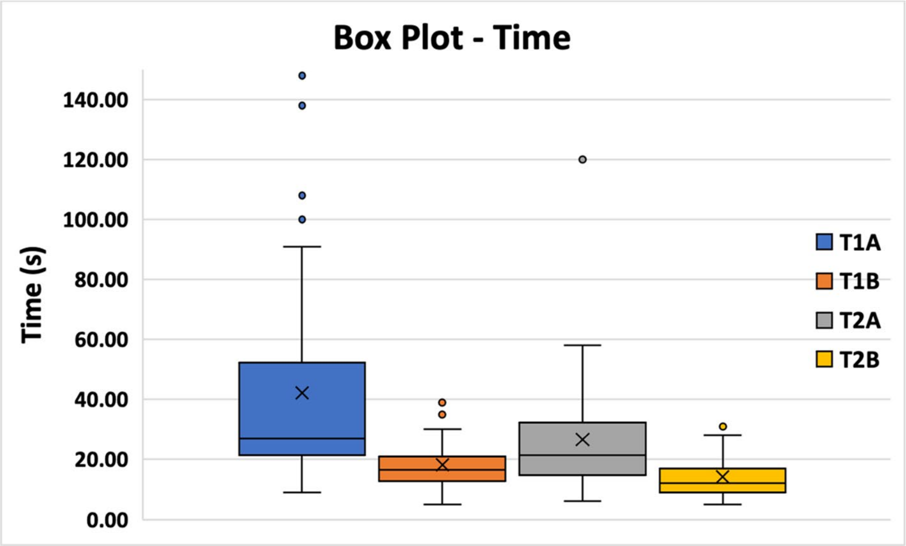

Fig. 9 Box plot for the time variable

Table 7 Results of the SUS test for both tasks and groups

Dataset	Task 1		Task 2	
	Task 1-A (distance only)	Task 1-B (distance + angle)	Task 2-A (distance only)	Task 2-B (distance + angle)
SUS1	3.605	4.316	4.079	4.289
SUS2	2.158	1.342	1.421	1.342
SUS3	4.316	4.711	4.395	4.684
SUS4	1.868	1.605	1.632	1.421
SUS5	4.263	4.684	4.579	4.632
SUS6	1.842	1.263	1.474	1.342
SUS7	4.447	4.763	4.579	4.789
SUS8	1.553	1.289	1.579	1.211
SUS9	4.395	4.895	4.474	4.816
SUS10	1.684	1.316	1.684	1.342
SUS-Score	79.803	91.382	85.789	91.382
Number of samples = 38				

Table 8 Non-parametric paired Wilcoxon signed-rank tests. SUS-Score. Group A vs group B

Dataset	Median ± IQR	Mean rank [neg., pos.]	Rank sum [neg., pos.]	Statistic (Z)	Significance
SUS-Score A Task 1	86.25 ± 21.875	[14.14, 16.54]	[99.00, 397.00]	-2.930	0.003
SUS-Score B Task 1	95.00 ± 13.125				
SUS-Score A Task 2	90.00 ± 13.750	[13.71, 17.45]	[164.50, 331.50]	-1.640	0.101
SUS-Score B Task 2	96.25 ± 13.125				
Number of samples = 38					

signed-rank test (Table 8) shows that the difference between the models is significant for Task 1. In any case, both models exceed 79 points. This means that the whole AR application lies within the *good* qualification [2, 8]. In addition, model B exceeds 90 points and model A for Task 2 exceeds 85. These lie within the *excellent* qualification.

A similar analysis can be done for the workload test. Table 9 shows the results of the NASA-TLX questionnaire in a 0–100 scale, where lower values mean lower workload. For Task 1, both mental and temporal demand, effort and frustration mean levels reported by users are clearly higher in the distance attenuation model, whereas performance and physical demand are similar. For Task 2, the situation is similar, although the differences are smaller. It is worth noting that Task 1 was considered more mentally demanding, because overall mean levels of mental demand decrease in Task 2 with respect to Task 1. The mean value of all six NASA-TLX dimensions – where the performance measure is subtracted from 100 before the mean is calculated, in order to harmonize all six dimensions to the 0–100 scale – provides a clearer picture, since this value is

Table 9 Mean results of the NASA-TLX test for both tasks and groups

Dataset	Task 1		Task 2	
	Task 1-A (distance only)	Task 1-B (distance + angle)	Task 2-A (distance only)	Task 2-B (distance + angle)
Mental demand	40.582	22.161	27.285	19.252
Physical demand	9.141	7.895	10.942	9.003
Temporal demand	26.870	11.773	22.992	13.296
Effort	31.579	16.482	24.792	18.421
Performance	79.501	88.504	83.102	89.474
Frustration	19.945	5.817	15.235	7.756
NASA-TLX (mean)	24.769	12.604	19.691	13.042
Number of samples = 38				

much smaller for model B in both tasks. A paired Wilcoxon signed-rank test (Table 10) shows that the difference between the models is significant for both tasks.

Finally, we analyze the results of the final questionnaire by means of a one-tailed binomial test. This test allows us to understand if the proportion of choices for either of the models is significantly different than the expected a priori value (50%). As seen in Table 11, the new model is the chosen one for both tasks for questions Q2 (recommendation for AR), Q3 (usefulness in AR), Q4 (sound identification) and Q5 (attention focus). Q1 (overall preference) remains inconclusive, albeit there is a majority of people who prefer the new model for both tasks.

5 Discussion

From the results of the objective and subjective data analyses, we can clearly see that the proposed new model is positively assessed and allows some tasks to be performed faster. Indeed, for both tasks, completion times are shorter for the directional model and the perceived temporal demand is also lower than when the classical model is used. Therefore, we can confidently say that H1 is true, even for two simple tasks like the ones proposed in the experiment.

Table 10 Non-parametric paired Wilcoxon signed-rank tests. NASA-TLX. Group A vs group B

Dataset	Median ± IQR	Mean rank [neg., pos.]	Rank sum [neg., pos.]	Statistic (Z)	Significance
NASA-TLX A Task 1	22.807 ± 23.903	[18.48, 15.67]	[536.00, 94.00]	-3.620	$< 10^{-3}$
NASA-TLX B Task 1	9.211 ± 15.351				
NASA-TLX A Task 2	14.912 ± 23.904	[20.39, 12.21]	[448.50, 146.50]	-2.583	0.010
NASA-TLX B Task 2	9.211 ± 17.543				
Number of samples = 38					

Table 11 Binomial tests for the two-choice questions

Task	Question	Model A (distance only)	Model B (distance + angle)	Binomial p-value
Task 1	*Q1*	16 (42.1%)	22 (57.9%)	0.209
	Q2	8 (21.1%)	30 (78.9%)	$<10^{-3}$
	Q3	8 (21.1%)	30 (78.9%)	$<10^{-3}$
	Q4	4 (10.5%)	34 (89.5%)	$<10^{-3}$
	Q5	8 (21.1%)	30 (78.9%)	$<10^{-3}$
Task 2	*Q1*	14 (36.8%)	24 (63.2%)	0.072
	Q2	12 (31.6%)	26 (68.4%)	0.017
	Q3	13 (34.2%)	25 (65.8%)	0.036
	Q4	6 (15.8%)	32 (84.2%)	$<10^{-3}$
	Q5	11 (28.9%)	27 (71.1%)	0.007

Number of samples = 38

Regarding H2, users make fewer mistakes with the new model in Task 2 and feel more frustrated with the classical model. Thus, H2 could be considered partially true, although the amount of room for making mistakes in this AR application is really small, because the use cases are simple and there are very few choices to make. In fact, none of the participants made a single error in Task 2 with the new model. Time is a more reliable measure in this experiment.

Regarding H3, the binomial tests for Q4 and Q5 leave no room for doubt. Users clearly feel that better sound identification and attention focus is achieved with the new model. The open-ended questions Q6 and Q7 also explain why the new model is preferred for these tasks. Here we list some of the answers to Q6: "*because you can hear it more clearly with model B", "because the sounds are further apart and easier to be distinguished with model B", "in model A sounds overlapped too much"*. With respect to Q7, some people did not know exactly what was going on and were unable to tell the difference between the models: "*not many*", "*I did not see much of a difference, but I found model B easier"*, whereas other participants did notice a difference: "*when I pointed at an object with model A, I could hear two sounds if I got too close, and I had the feeling that the device was not picking it up properly. This did not happen with model B", "using model A you had to be really focused on the task because the sounds were mixed", "with model A the sounds were rather mixed throughout the task, while with model B each sound had its own space".*

Regarding H4, the usability tests and the NASA-TLX also point in the same direction: the new model is, generally speaking, more usable and needs less workload on the part of the users. Thus, H4 can be considered true, according to the data. Usability is significantly increased for Task 1 and the new model obtains a very high score (91.382), whereas for Task 2 the difference between the models is not statistically significant but only the new model obtains a SUS score that is higher than 90 points. The fact that the usability tests obtain such high scores for both models reinforces the validity of the results because they are obtained from a tool that is perceived positively.

Workload is significantly reduced for both tasks when the new model is used. In fact, for Task 1 the NASA-TLX index is halved with respect to the classical one. Regarding the six workload dimensions, mental and temporal demand, effort and frustration levels are lower with the new model, but the differences are smaller in Task 2. This is probably because

Task 2 was easier for the participants. The amount of time they needed to complete Task 1 was higher than for Task 2, which reaffirms this idea. In fact, both tasks are not physically challenging and are relatively easy to accomplish. This is why physical demand and performance are not particularly affected by the choice of model. The differences between Task 1 and Task 2 can also be analyzed in terms of the width of the attenuation zone. In Task 1, the attenuation zone is larger because the exponent is twice the value as in Task 2. Therefore, the differences between the classical attenuation model and the new one are more evident in Task 1. We argue that this is the reason why there are greater differences (in usability and workload) between the two models for Task 1 than for Task 2.

The fact that the results of questions Q2 (recommendation for AR) and Q3 (usefulness in AR) show a significant difference in favor of the new model is consistent with the rest of the results and points out that the new model is the recommended one and is perceived as more useful than the classical one.

Overall, the new model is positively assessed and can be seen as a generalization of the classical distance-based attenuation. In fact, setting up the cutoff angle to 180º degrees –which makes the cone omnidirectional– and the exponent to a very small value makes the two models equivalent. An interesting question, which is difficult to answer, is if the suitability of using the new method could generalize to most AR environments, especially to environments that are very different from the one tested. The answer probably depends on two factors: (i) the need of achieving sound isolation or sound identification and (ii) the number of simultaneous sound sources. In tasks where sound identification/isolation is not necessary, limiting the hearing range of the listener could be counterproductive. However, not doing so could also lead to an unintelligible combination of sounds that is no different from noise, turning sound into a nuisance rather than a perceptual cue. Most AR applications combine several virtual objects and it is often important to identify which virtual object is producing a particular sound. Therefore, it is not difficult to imagine that this model could be useful in a wide range of situations, let alone knowing that it is a generalization of the classical one. Of course, this would depend on the number of sources and their spatial density. For instance, there could be situations in which a single source is present, the sound could come from behind and it would be important to be able to hear it. Thus, the proposed model may not be necessary for AR environments with very few sound sources or with sound sources that are not likely to overlap, but it can be of great help in complex environments in which filtering the amount of perceptual cues perceived by the user could be the difference between success and failure. In any case, it is beneficial to have an alternative model that demonstrates a potential for greater efficiency than the usual distance-based attenuation model typically found in multimedia creation environments.

6 Conclusions and further work

Sound is often given a marginal role in AR systems and when is implemented, it is generally assumed that the traditional distance attenuation model is the most intuitive and useful system for all users and cases. In this paper, we challenge this assumption and propose an attenuation model in which the listener orientation is also considered. We call this a *directional attenuation model*. In order to test the new proposed model, we have developed an AR application that involves visual and sound stimuli and we have created two tasks in which a user needs to identify a specific sound while multiple sound sources placed at different locations coexist. We hypothesize that users will need less time and will make fewer errors when trying to recognize the position of a sound with the new model. We also hypothesize that users will be better able to focus their attention with the new model, and that usability and workload will be improved.

A total of 38 persons were asked to perform these two tasks with the two models to test the hypotheses. Objective and subjective measures were collected and analyzed, including usability and workload data. Our experiments show that the proposed model provides better workload for the tested tasks, requiring less time and effort, allowing users to explore the AR environment more easily and intuitively. Therefore, we conclude that the proposed model could be useful in AR applications in which several sounds coexist and need to be identified. In addition, given that the proposed model can be seen as a generalization of the classical distance-based attenuation model, and the computational load of the new proposed model is low, there are few reasons not to include this model in future AR systems, even if multiple sound sources are not present. The model could also be useful in Virtual Reality (VR) environments, since the underlying working assumptions and the spatial constraints can be similar. In this regard, our model aims at AR/VR applications in which an effective functional virtual acoustic reality (i.e., one that is functional for the deployed purpose of navigating a virtual environment, relaying audio information to users, etc.) is needed, and not at applications where the simulated sound must be physically realistic.

The study has also some limitations. First, during the experimental assessment we did not track the user's head but the camera of the mobile device. While we can assume that the user is looking at the display and keeps the same orientation as the device, it is true that the viewing direction refers to the camera and not to the user, and thus the attenuation of sound is applied according to the camera rather than the user. We did this because we did not want the setup to be too complex, so that the participants would be quickly familiar with the use of the tool. Indeed, this interaction metaphor –where users move the device to detect virtual sound–, has been described before in [23]. They proposed a system called AudioTorch, also demonstrating that it allows quick orientation and easy discrimination between proximate sources.

A similar experiment could be setup with an optical see-through AR device, such as Microsoft HoloLens 2 or Magic Leap One. With these immersive devices, the user will keep the same orientation as the device and the user experience could be more consistent. Therefore, it is not unlikely that the benefits of the new model will increase if this limitation is not present. The second limitation comes from the fact that we tested the system with a stereo setup, using headphones and the user was expected to move on the floor plane. Nevertheless, it is important to emphasize that the proposed model is three-dimensional as it calculates the attenuation in 3D, both in terms of distance and angle; i.e. the effects of sound attenuation with respect to the listener position and direction are calculated in 3D, not assuming that the listener moves on a plane.

Our future work will revolve around four main areas. First, we will incorporate HRTF into our system. We also hope to develop new tasks in order to test the proposed in a wide range of AR-related practical situations. It would also be interesting to perform a comprehensive analysis of the effects of the parameters (cutoff, exponent) in the usability of the proposed model. Finally, we plan to test the new model in different AR paradigms, such as see-through AR or spatial AR, and also in VR applications.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. Cristina Portalés is supported by the Spanish government postdoctoral grant Ramón y Cajal, under grant No. RYC2018-025009-I. Jesús Gimeno and Sergio Casas-Yrurzum are supported by Spain's Agencia Estatal de Investigación and the Spanish Ministry of Science and Innovation, through the ViPRAS project (Virtual

Planning for Robotic-Assisted Surgery)—PID2020-114562RA-I00 (funded by MCIN/AEI/https://doi.org/10.13039/501100011033).

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests The authors have no competing interests to declare that are relevant to the content of this article.

Informed consent All the participants in this research volunteered for the experiment and signed an informed consent regarding their participation in the study, which was completely anonymous.

References

1. Azuma RT (1997) A Survey of Augmented Reality. Presence: Teleoperators and Virtual Environments 6:355–385
2. Bangor A, Kortum P, Miller J (2009) Determining what individual SUS scores mean: Adding an adjective rating scale. J Usability Stud 4:114–123
3. Barcali E, Iadanza E, Manetti L et al (2022) Augmented Reality in Surgery: A Scoping Review. Appl Sci 12:6890. https://doi.org/10.3390/app12146890
4. Bauer V, Nagele A, Baume C et al (2019) Designing an Interactive and Collaborative Experience in Audio Augmented Reality. In: Bourdot P, Interrante V, Nedel L et al (eds) Virtual Reality and Augmented Reality. Springer International Publishing, Cham, pp 305–311
5. Begault DR (1994) 3D Sound for Virtual Reality and Multimedia, 1st edn. Academic Press, Boston
6. Billinghurst M, Bowskill J, Dyer N, Morphett J (1998) An evaluation of wearable information spaces. In: Proceedings. IEEE 1998 Virtual Reality Annual International Symposium (Cat. No.98CB36180). pp 20–27
7. Brooke J (1996) SUS-A quick and dirty usability scale. Usability Evaluation in Industry 189:4–7
8. Brooke J (2013) SUS: a retrospective. J Usability Stud 8:29–40
9. Caudell TP, Mizell DW (1992) Augmented reality: An application of heads-up display technology to manual manufacturing processes. IEEE, pp 659–669
10. Chatzidimitris T, Gavalas D, Michael D (2016) SoundPacman: Audio augmented reality in location-based games. In: 2016 18th Mediterranean Electrotechnical Conference (MELECON). pp 1–6
11. Cliffe L, Mansell J, Greenhalgh C, Hazzard A (2021) Materialising contexts: virtual soundscapes for real-world exploration. Pers Ubiquit Comput 25:623–636. https://doi.org/10.1007/s00779-020-01405-3
12. Cohen M, Aoki S, Koizumi N (1993) Augmented audio reality: telepresence/VR hybrid acoustic environments. In: Proceedings of 1993 2nd IEEE International Workshop on Robot and Human Communication. pp 361–364
13. Costa MC, Santos P, Patrício JM, Manso A (2021) An Interactive Information System That Supports an Augmented Reality Game in the Context of Game-Based Learning. Multimodal Technol Inter 5:82. https://doi.org/10.3390/mti5120082
14. Dini G, Mura MD (2015) Application of Augmented Reality Techniques in Through-life Engineering Services. Procedia CIRP 38:14–23. https://doi.org/10.1016/j.procir.2015.07.044
15. Doerr K, Rademacher H, Huesgen S, Kubbat W (2007) Evaluation of a Low-Cost 3D Sound System for Immersive Virtual Reality Training Systems. IEEE Trans Visual Comput Graphics 13:204–212. https://doi.org/10.1109/TVCG.2007.37

16. Edwards PJE, Chand M, Birlo M, Stoyanov D (2021) The Challenge of Augmented Reality in Surgery. In: Atallah S (ed) Digital Surgery. Springer International Publishing, Cham, pp 121–135
17. Feng S, He X, He W, Billinghurst M (2022) Can you hear it? Stereo sound-assisted guidance in augmented reality assembly. Virtual Reality. https://doi.org/10.1007/s10055-022-00680-0
18. Frank M, Rudrich D, Brandner M (2020) Augmented practice-room—Augmented acoustics in music education. Fortschritte der Akustik-DAGA
19. Gimeno J, Portalés C, Coma I, et al (2017) Combining Traditional and Indirect Augmented Reality for Indoor Crowded Environments. A Case Study on the Casa Batlló Museum. Computers & Graphics. https://doi.org/10.1016/j.cag.2017.09.001
20. Goudeseune C, Kaczmarski H (2001) Composing Outdoor Augmented-Reality Sound Environments
21. Greenhalgh C, Benford S (1995) MASSIVE: a collaborative virtual environment for teleconferencing. ACM Trans Comput-Hum Interact 2:239–261. https://doi.org/10.1145/210079.210088
22. Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Advances in psychology. Elsevier, pp 139–183
23. Heller F, Borchers J (2014) AudioTorch: using a smartphone as directional microphone in virtual audio spaces. In: Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services. Association for Computing Machinery, New York, NY, USA, pp 483–488
24. Hung S-W, Chang C-W, Ma Y-C (2021) A new reality: Exploring continuance intention to use mobile augmented reality for entertainment purposes. Technol Soc 67:101757. https://doi.org/10.1016/j.techsoc.2021.101757
25. Inoue A, Ikeda Y, Yatabe K, Oikawa Y (2019) Visualization system for sound field using see-through head-mounted display. Acoust Sci Technol 40:1–11. https://doi.org/10.1250/ast.40.1
26. Kaghat FZ, Azough A, Fakhour M (2018) SARIM: A gesture-based sound augmented reality interface for visiting museums. In: 2018 International Conference on Intelligent Systems and Computer Vision (ISCV). pp 1–9
27. Kataoka Y, Teraoka W, Oikawa Y, Ikeda Y (2018) Real-time measurement and display system of 3D sound intensity map using optical see-through head mounted display. In: SIGGRAPH Asia 2018 Posters. Association for Computing Machinery, New York, NY, USA, pp 1–2
28. Kim H, Matuszka T, Kim J-I et al (2017) Ontology-based mobile augmented reality in cultural heritage sites: information modeling and user study. Multimed Tools Appl 76:26001–26029. https://doi.org/10.1007/s11042-017-4868-6
29. Lam J, Kapralos B, Kanev K et al (2015) Sound localization on a horizontal surface: virtual and real sound source localization. Virtual Reality 19:213–222. https://doi.org/10.1007/s10055-015-0268-2
30. Larsson P, Västfjäll D, Kleiner M (2001) Ecological acoustics and the multi-modal perception of rooms: real and unreal experiences of auditory-visual virtual environments
31. Liang BS, Liang AS, Roman I, et al (2023) Reconstructing room scales with a single sound for augmented reality displays. Journal of Information Display
32. Liarokapis F, Petridis P, Lister PF, White M (2002) Multimedia Augmented Reality Interface for E-learning (MARIE). World Trans Eng Technol Educ 1:173–176
33. Liarokapis F, White M, Lister P (2004) Augmented reality interface toolkit. In: Proceedings. Eighth International Conference on Information Visualisation, 2004. IV 2004. pp 761–767
34. Mahmood Z, Ali T, Muhammad N et al (2017) EAR: Enhanced Augmented Reality System for Sports Entertainment Applications. KSII Trans Internet Inform Syst (TIIS) 11:6069–6091. https://doi.org/10.3837/tiis.2017.12.021
35. Massey FJ Jr (1951) The Kolmogorov-Smirnov test for goodness of fit. J Am Stat Assoc 46:68–78
36. Meža S, Turk Ž, Dolenc M (2015) Measuring the potential of augmented reality in civil engineering. Adv Eng Softw 90:1–10. https://doi.org/10.1016/j.advengsoft.2015.06.005
37. Mishra P, Pandey CM, Singh U et al (2019) Descriptive statistics and normality tests for statistical data. Ann Card Anaesth 22:67
38. Muliyati D, Bakri F, Ambarwulan D (2019) The design of sound wave and optic marker for physics learning based-on augmented reality technology. J Phys: Conf Ser 1318:012012. https://doi.org/10.1088/1742-6596/1318/1/012012
39. Noghabaei M, Heydarian A, Balali V, Han K (2020) Trend Analysis on Adoption of Virtual and Augmented Reality in the Architecture, Engineering, and Construction Industry. Data 5:26. https://doi.org/10.3390/data5010026
40. Panou C, Ragia L, Dimelli D, Mania K (2018) An Architecture for Mobile Outdoors Augmented Reality for Cultural Heritage. ISPRS Int J Geo Inf 7:463. https://doi.org/10.3390/ijgi7120463
41. Portalés C, Viñals MJ, Alonso-Monasterio P, Morant M (2010) AR-Immersive Cinema at the Aula Natura Visitors Center. IEEE Multimedia 17:8–15

42. Portalés Ricart C (2009) Entornos multimedia de realidad aumentada en el campo del arte. Universidad Politécnica de Valencia
43. Posner MI, Nissen MJ, Klein RM (1976) Visual dominance: An information-processing account of its origins and significance. Psychol Rev 83:157–171. https://doi.org/10.1037/0033-295X.83.2.157
44. Prokhorov A, Klymenko I, Yashina E, et al (2017) SCADA Systems and Augmented Reality as Technologies for Interactive and Distance Learning. pp 245–256
45. Ren G, Wei S, O'Neill E, Chen F (2018) Towards the Design of Effective Haptic and Audio Displays for Augmented Reality and Mixed Reality Applications. Adv Multimed 2018:1–11. https://doi.org/10.1155/2018/4517150
46. Rovithis E, Moustakas N, Floros A, Vogklis K (2019) Audio Legends: Investigating Sonic Interaction in an Augmented Reality Audio Game. Multimodal Technol Inter 3:73. https://doi.org/10.3390/mti3040073
47. Rumiński D (2015) An experimental study of spatial sound usefulness in searching and navigating through AR environments. Virtual Reality 19:223–233. https://doi.org/10.1007/s10055-015-0274-4
48. Sagayam M, Henesey L, Ho CC et al (2020) Augmented reality-based solar system for e-magazine with 3-D audio effect. Int J Simul Process Model 15:524. https://doi.org/10.1504/IJSPM.2020.10034721
49. Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). Biometrika 52:591–611
50. Sodnik J, Tomazic S, Grasset R, et al (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. Association for Computing Machinery, New York, NY, USA, pp 111–118
51. Stahl C (2007) The roaring navigator: a group guide for the zoo with shared auditory landmark display. In: Proceedings of the 9th international conference on Human computer interaction with mobile devices and services. Association for Computing Machinery, New York, NY, USA, pp 383–386
52. Unity Technologies (2023) Unity - Manual: Audio Spatializers. https://docs.unity3d.com/Manual/VRAudioSpatializer.html. Accessed 2 Nov 2023
53. Vávra P, Roman J, Zonča P et al (2017) Recent Development of Augmented Reality in Surgery: A Review. J Healthcare Eng 2017:e4574172. https://doi.org/10.1155/2017/4574172
54. Vazquez-Alvarez Y, Oakley I, Brewster SA (2012) Auditory display design for exploration in mobile audio-augmented reality. Pers Ubiquit Comput 16:987–999. https://doi.org/10.1007/s00779-011-0459-0
55. Villegas J (2015) Locating virtual sound sources at arbitrary distances in real-time binaural reproduction. Virtual Reality 19:201–212. https://doi.org/10.1007/s10055-015-0278-0
56. Yeoward C, Shukla R, Stewart R et al (2021) Real-Time Binaural Room Modelling for Augmented Reality Applications. JAES 69:818–833
57. Yewdall DL (2011) Practical Art of Motion Picture Sound. Waltham, MA
58. Zhou Z, Cheok AD, Qiu Y, Yang X (2007) The Role of 3-D Sound in Human Reaction and Performance in Augmented Reality Environments. IEEE Trans Syst Man Cybern Part A: Syst Humans 37:262–272. https://doi.org/10.1109/TSMCA.2006.886376
59. Zhou Z, Cheok AD, Yang X, Qiu Y (2004) An experimental study on the role of software synthesized 3D sound in augmented reality environments. Interact Comput 16:989–1016. https://doi.org/10.1016/j.intcom.2004.06.014
60. Zimmermann A, Lorenz A (2008) LISTEN: a user-adaptive audio-augmented museum guide. User Model User-Adap Inter 18:389–416. https://doi.org/10.1007/s11257-008-9049-x