July 2022

Cascading Augmented Sound Interactions Among Virtual and Physical Objects

Corville Allen

Andy Lavery

Cascading Augmented Sound Interactions Among Virtual and Physical Objects

ABSTRACT

Current augmented reality (AR) technology does not enable the transitioning of sounds across scenes on multiple AR devices, or even between AR objects within a single scene. This disclosure describes techniques for audio generated by virtual or physical objects in one or more AR scenes viewed via one or more viewports to create interactions between virtual objects in or across scenes. The interactions can cascade from one or more virtual or physical objects in an AR scene to one or more other virtual or physical objects in or across AR scenes across multiple AR devices.

KEYWORDS

- Augmented reality (AR)
- Virtual sound
- Multi-viewport AR
- Virtual reality (VR)
- Immersive experience
- Transitivity
- Cascaded audio
- AR platform

BACKGROUND

Augmented reality (AR) incorporates virtual objects into the visual display of objects in the real world, typically through a single viewport attached or associated with a physical viewing device such as glasses, headset, etc. Current AR technology does not enable the transitioning of sounds across scenes on multiple AR devices, or even between AR objects within a single scene.

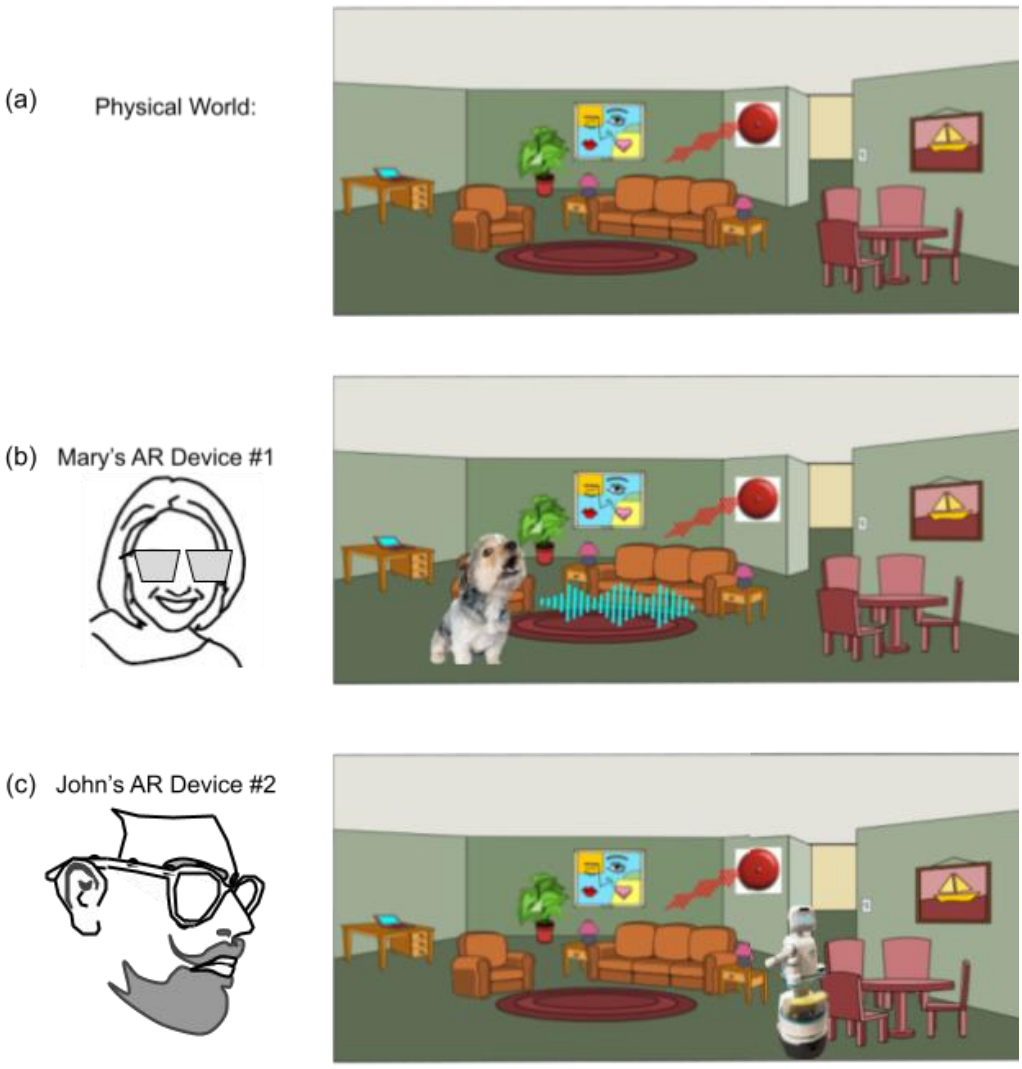

Fig. 1: Virtual sounds generally do not travel across devices

For example, as illustrated in Fig. 1, a physical scene (Fig. 1a) is viewed by two users, Mary and John, using their respective AR devices. Mary's viewport (Fig. 1b) includes a virtual dog, while John's viewport (Fig. 1c) includes a virtual robot. The ring of a physical doorbell (red sound waves) is heard by both Mary and John. The virtual dog in Mary's AR environment reacts to the ring by barking (blue sound waves). The virtual robot in John's AR environment is not programmed to react to a doorbell, but it is programmed to react to a dog's bark. However, the

virtual dog's bark does not travel across devices. The virtual robot doesn't hear the virtual dog bark and hence doesn't react.

AR techniques today can play virtual sounds or modify environmental sounds, e.g., they can cancel out certain types of sounds or noises. However, current techniques do not enable virtual objects to respond to physical or virtual sounds, nor can they cascade the effect of a sound across multiple viewports, e.g., have the virtual sound that originates in one viewport cause an effect in another viewport.

DESCRIPTION

This disclosure describes techniques for audio generated by virtual or physical objects in one or more AR scenes (e.g., one or more viewports) to create interactions between virtual objects in or across scenes. The interactions can cascade from one or more virtual or physical objects in an AR scene to one or more other virtual or physical objects in or across AR scenes (e.g., across AR devices).

The real and virtual audio produced or received by an AR object is designated to be transitional (transitive) - capable of transitioning between virtual objects in a cause-and-effect manner. Based on the transitive state of the audio, the AR platform tests scenes for AR objects (which can be across multiple viewports or devices) that can respond to the audio. An AR object in any viewport found to be responsive to transitive audio receives the spatial audio properties for the scene and relational objects therein. A device receiving transitive audio transfers the properties of the transitive audio to the AR objects within its scene, such that AR objects therein can respond.

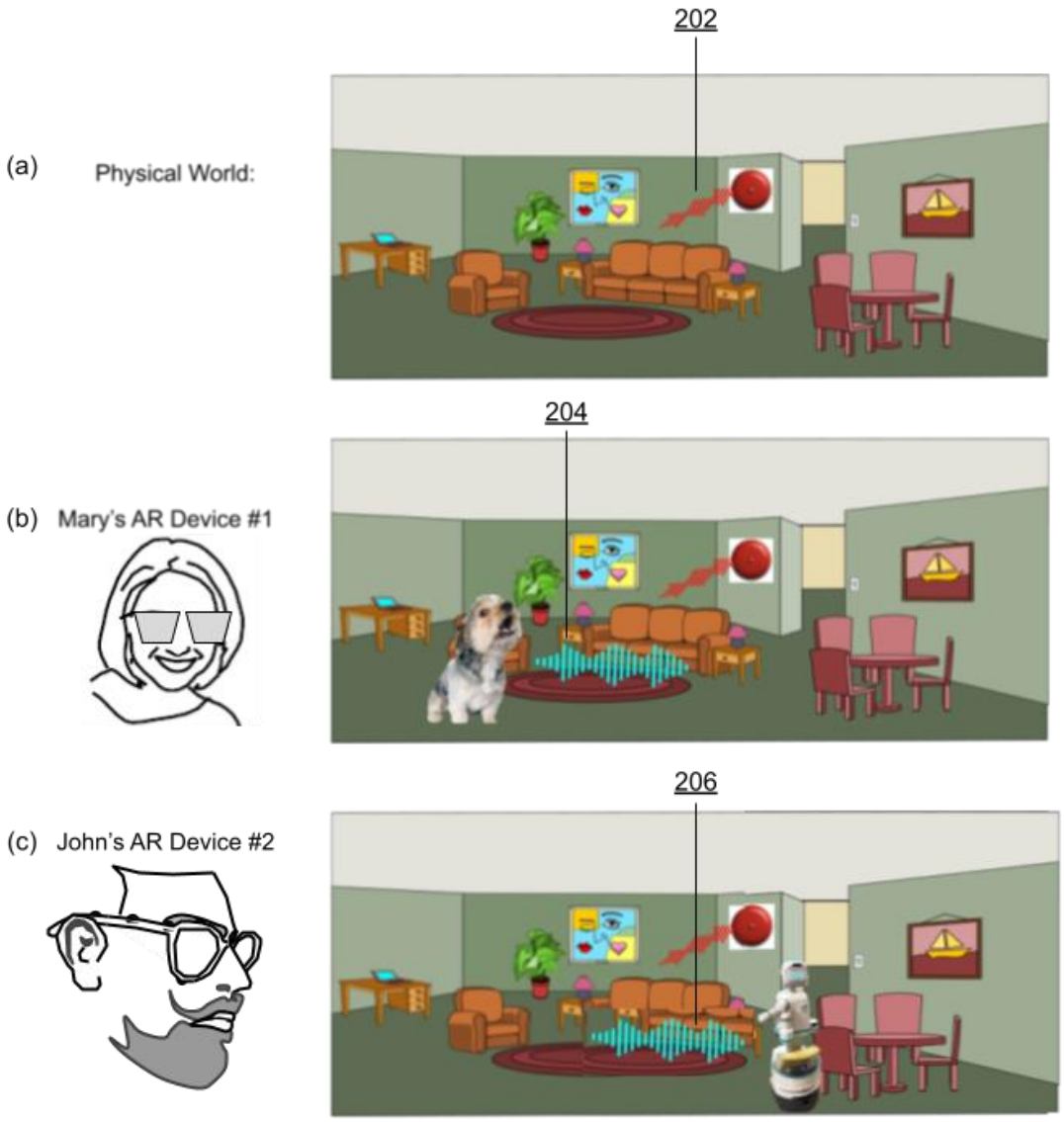

Fig. 2: Cascading augmented sound interactions among virtual and physical objects

Fig. 2 illustrates cascading augmented sound interactions among virtual and physical objects. A physical scene (Fig. 2a) is viewed by two users, Mary and John, using distinct AR devices. Mary's viewport (Fig. 2b) includes a virtual dog, while John's viewport (Fig. 2c) includes a virtual robot. The ring of a physical doorbell (202) is designated as transitive and is heard by both Mary and John. The virtual dog in Mary's AR environment reacts to the ring by barking (204). The virtual robot in John's AR environment is not programmed to react to a

doorbell, but it is programmed to react to a dog's bark. The bark of a dog is marked as a transitive sound, e.g., one that travels across devices. Therefore, the virtual bark generated in Mary's AR device is heard in John's AR device (206). In particular, the virtual robot in John's AR device hears the virtual dog bark and reacts, e.g., with a waving gesture to calm down the dog.

As illustrated by the example of Fig. 2, there are two types of sound, e.g., a virtual sound produced by a virtual object (e.g., a virtual dog that barks); and a real-world sound, e.g., the ring of a physical doorbell or a physical door being shut. Either type of sound can be marked as transitive - capable of propagating across objects and devices, or non-transitive - not propagated across devices.

Examples

- The real audio ring generated by the physical doorbell can be designated as transitive, such that it is propagated to objects in both AR scenes.
- Mary's speech, "John, can you please check the door" can be designated as non-transitive by the AR platform system. It does not get transferred further to the AR devices and their virtual objects.
- The bark of the virtual dog can be designated as transitive, such that it propagates to objects in both AR scenes.

Transitive audio from the doorbell is received at Mary's device, causing the virtual dog in Mary's AR scene to turn towards the sound source (the doorbell) and bark. Although the doorbell ring is transitive and received by John's device, the virtual robot in John's AR scene isn't programmed to respond to it. However, since the bark of the virtual dog is transitive, it is propagated from Mary's AR scene to John's AR scene, where it is received by the virtual robot.

The virtual robot, which is pre-programmed to respond to a dog's bark, performs a waving gesture to the virtual dog to calm it down. The barking sound is associated with a spatial location and sentiment, for which the virtual robot has pre-configured reactions.

Effectively, multiple virtual objects in one or more AR scenes, viewed via distinct viewports, each make virtual sounds in response to physical or virtual sounds in the one or more scenes. The virtual sounds are coordinated to effect a realistic interaction between virtual objects across scenes. The multiple viewports act as one AR platform that coherently handles cross-viewport sounds.

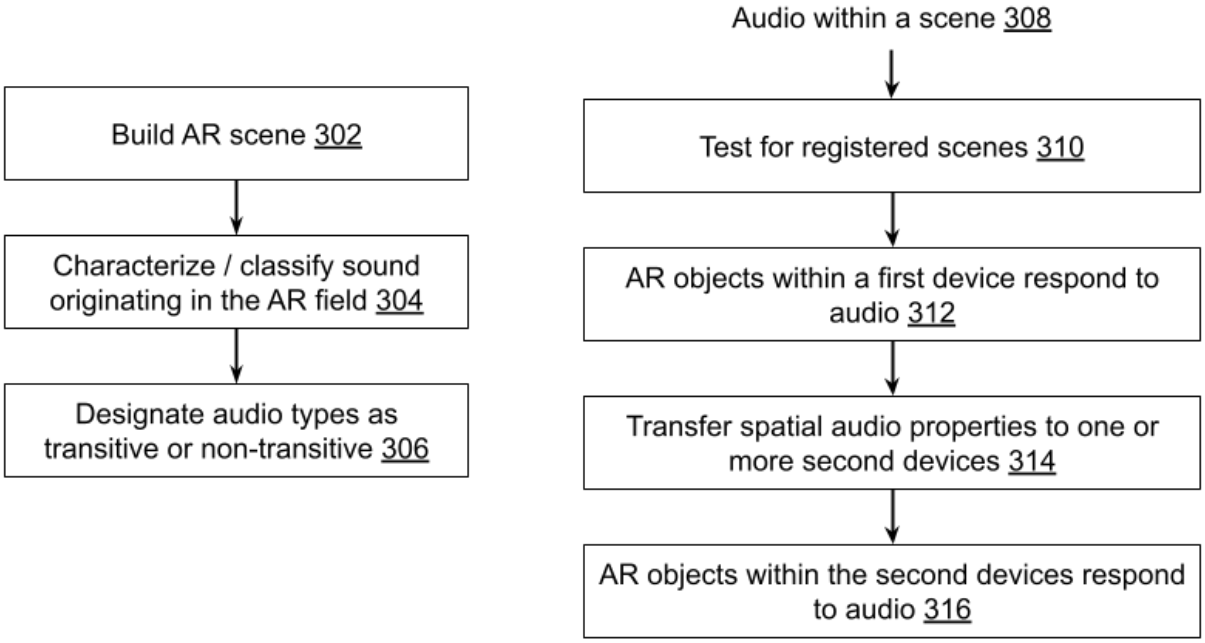

Fig. 3: Flowcharts for cascading sound interactions in augmented reality

Fig. 3 illustrates example flowcharts for cascading sound interactions in augmented reality. An augmented reality scene is built that includes one or more physical objects and one or more virtual objects (302). Audio signals from the scene are observed, spatially located, and classified across various dimensions (304), such as amplitude, wavelength, frequency, period, velocity, text of spoken language, sentiment of spoken language, sentiment of musical sounds

(cheerful, calming, energetic, dangerous, etc.). The real and virtual audio produced or received by a (physical or virtual) object is designated as transitional or not (306). For example, virtual audio can be designated as transitive. Audio generated by real objects in the scene can be designated as transitive (or not) based on the type of audio (e.g., speech, non-speech, non-human sounds, etc.).

Given audio produced or received within a scene (308), the AR platform tests registered scenes (which may be across multiple devices) for AR objects that respond to the type of the audio (310). For example, a first AR device can check for AR objects within its own scene that are responsive to the audio. Such AR objects respond to the audio (312). The first AR device transfers the spatial audio properties for the scene and relational objects to one or more second devices (314). The one or more second devices transfer the audio to the AR objects within their own scenes, where AR objects respond depending on the properties of the audio (316).

Response to audio can vary across objects in the viewport space and can depend on the type of object, the relationship to other objects in the viewport space, the classification attributes of the audio, spatial characteristics of the audio, and spatial characteristics of the objects in the viewport space. For example, AR objects can respond based on one or more key-action types. Some AR objects may create virtual audio responses, such as talking; laughing; communicating through non-language sounds (purring, barking, grunting, etc.) that communicate emotions; or other sounds. These sounds, which are responses to other sounds, are in turn introduced into the scenes to cascade across the scenes again.

Virtual sound generated within the AR scene can be analyzed for structured information included within. For example, spoken language or text, whether human or computer-generated,

can be semantically analyzed and understood, enabling high-fidelity representation of the semantic and spatial characteristics of virtual sounds in the AR scene.

Additionally, virtual objects in an AR scene can also create physical sounds, e.g., real sound waves that propagate through the air as opposed to virtual sounds that are played through the earphones of AR devices. Physical sounds created by virtual devices can in turn participate in propagating through the one or more AR scenes. For example, a physical drum-set can take AR input to make physical sounds, e.g., when a virtual object hits the drum. The physical sounds produced by the virtual object hitting the physical drum cascade through the AR scenes in the manner described above.

In this manner, by handling virtual sound interactions accurately, the techniques of this disclosure enable multiple AR objects spread across multiple viewports to interact realistically. The techniques apply to a wide variety of AR settings, including events, meetings, logistics warehouses, virtual companions, advertisements, maps, navigation, AR walking directions, etc. The techniques generally enable the user to participate more fully in augmented reality with audio controls.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs, or features described herein may enable the collection of user information (e.g., information about a user's social network, social actions or activities, profession, a user's preferences, or a user's current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location

information is obtained (such as to a city, ZIP code, or state level) so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

Current augmented reality (AR) technology does not enable the transitioning of sounds across scenes on multiple AR devices, or even between AR objects within a single scene. This disclosure describes techniques for audio generated by virtual or physical objects in one or more AR scenes viewed via one or more viewports to create interactions between virtual objects in or across scenes. The interactions can cascade from one or more virtual or physical objects in an AR scene to one or more other virtual or physical objects in or across AR scenes across multiple AR devices.

REFERENCES

[1] Yang, Jing, and Friedemann Mattern. “Audio augmented reality for human-object interactions.” In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, pp. 408-412. 2019.

[2] Wong, Adrian, and Xiaoyu Miao. “Methods and devices for rendering interactions between virtual and physical objects on a substantially transparent display.” U.S. Patent 8,990,682, issued March 24, 2015.