Technical Disclosure Commons

Defensive Publications Series

May 2020

Synesthetic Soundtrack

Elly Nattinger

Meltem Oktem

Isabel Lee

Synesthetic Soundtrack

ABSTRACT

This disclosure describes techniques to generate an audio experience or soundscape corresponding to the visual field of a user. With user permission, objects within the feed of a head-mounted camera are semantically identified using computer vision techniques. Based on the detected objects, a unique audio experience, shaped by the world around a user and by the physical items they engage with, is generated.

KEYWORDS

- Soundscape
- Virtual reality
- Augmented reality
- Semantic identification
- Object identification
- Computer vision

BACKGROUND

Virtual or augmented reality (VR or AR) technologies enable a user to modify their visual field, e.g., add virtual objects to their visual fields. The VR/AR experience thus far does not include the addition of virtual sounds that are appropriate to the current environment of the user or generable by physical manipulation of everyday objects.

Although disc jockeys and other music professionals have at their disposal a large range of software and hardware to control, synthesize, or shape music, these do not incorporate computer vision or semantic analysis of objects in a wearable form factor.

Some specialized hardware is available that enable users to control interfaces (including generating sounds) based on everyday objects, but these work in a tactile mode, e.g., capacitive sensing that controls a keyboard. For example, if a fruit is attached to this hardware, the user can touch the fruit to trigger a sound. There is no use made of semantic knowledge of the identity of the object as a fruit in the triggering of sounds.

A musical instrument in the form of a table has been demonstrated that enables users to manipulate pre-determined, RFID-tagged objects on the tabular surface to shape a generated sound. The instrument is custom hardware designed for permanent installations or performances, does not incorporate computer vision, and only works with specific pre-made objects. Other demonstrations that capture audio from the user's environment to activate a generative soundtrack are similar, but do not semantically map the user's visual environment to an audio soundscape.

DESCRIPTION

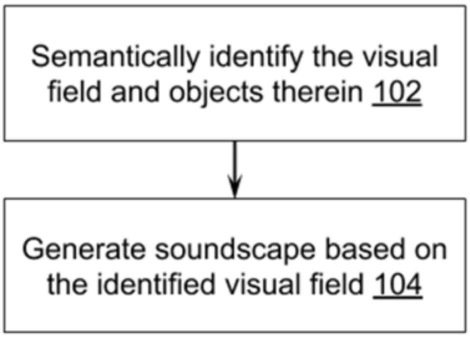
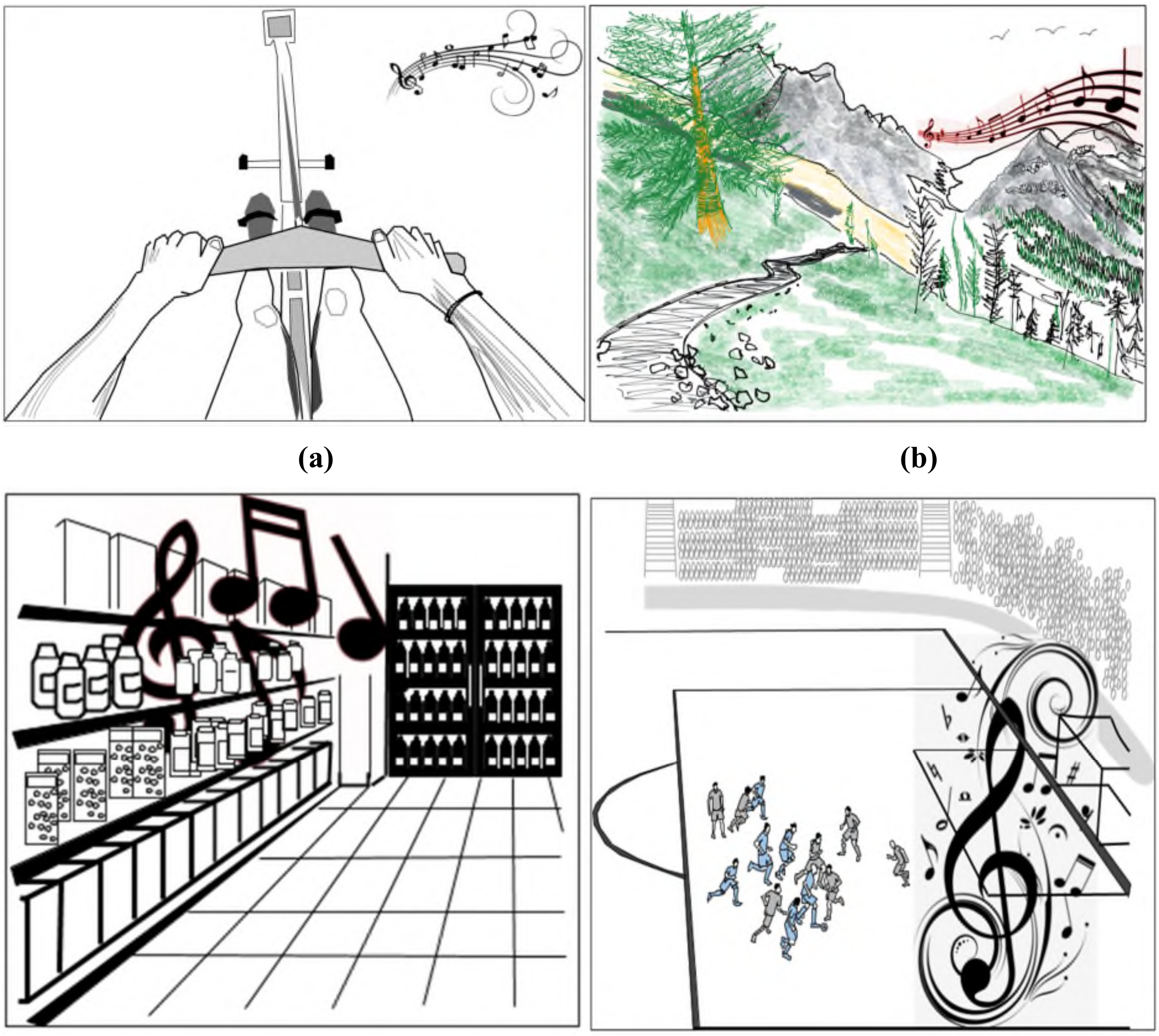


Fig. 1: Generating a synesthetic soundscape

This disclosure describes techniques to generate an audio experience, or soundscape, corresponding to the visual field of a user. As illustrated in Fig. 1, with user permission, objects within the feed of a head-mounted camera are semantically identified (102) using computer vision techniques. Based on the detected objects, a unique audio experience shaped by the world around a user and by the physical items they engage with is generated (104). The techniques of audio generation can be incorporated into any device, e.g., a wearable device such as glasses or head-mounted device, that can capture the user's point of view.



(a) (b) (c) (d)

Fig. 2: Generating a soundscape appropriate to the user's environment (a) User rides a rowing machine in the gym; (b) User is on a nature walk; (c) User is at a store; (d) User is at a game.

Per the techniques, the scenery or the objects detected in a user's field of view inform the soundscape generated for the user. As illustrated in Fig. 2(a): when the user is working out in the gym, one set of sounds or music is played; Fig. 2(b): when the user is outside in nature, another set of sounds is played; Fig. 2(c): when the user is at the store, a corresponding set of sounds is played; Fig. 2(d): when the user is at a game, yet another set of sounds is played.

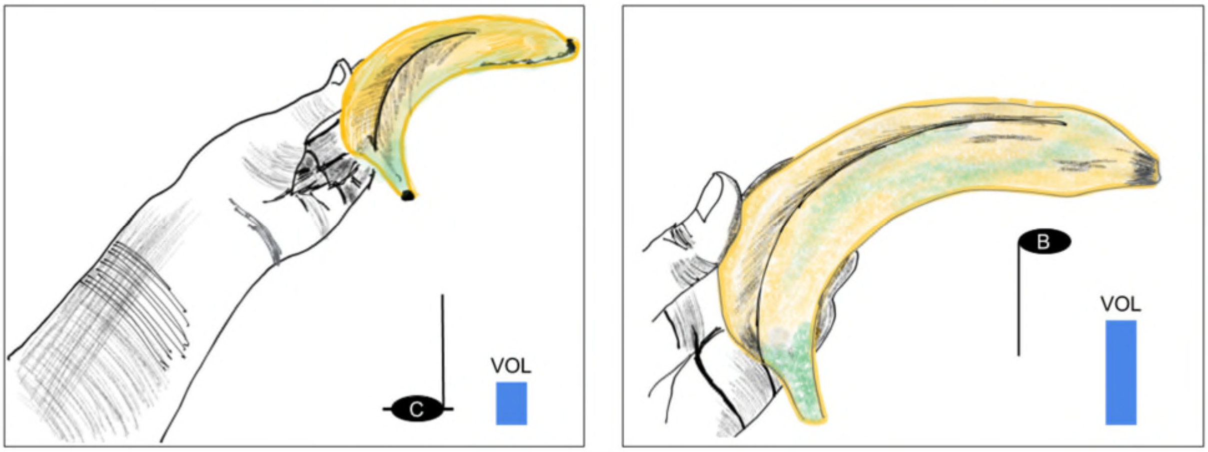


Fig. 3: Soundscape based on the manipulation of everyday objects

The user can also manipulate objects and their positions in their three-dimensional visual field to control the blend of a set of audio tracks, the timbre, or the frequency of specific sonic components. This is illustrated in Fig. 3, where the pitch and the volume of the audio vary with the distance of the banana to the user. Movement of a specific object within the user's viewpoint can also affect sonic behaviors, as can collecting specific numbers or groups of objects.

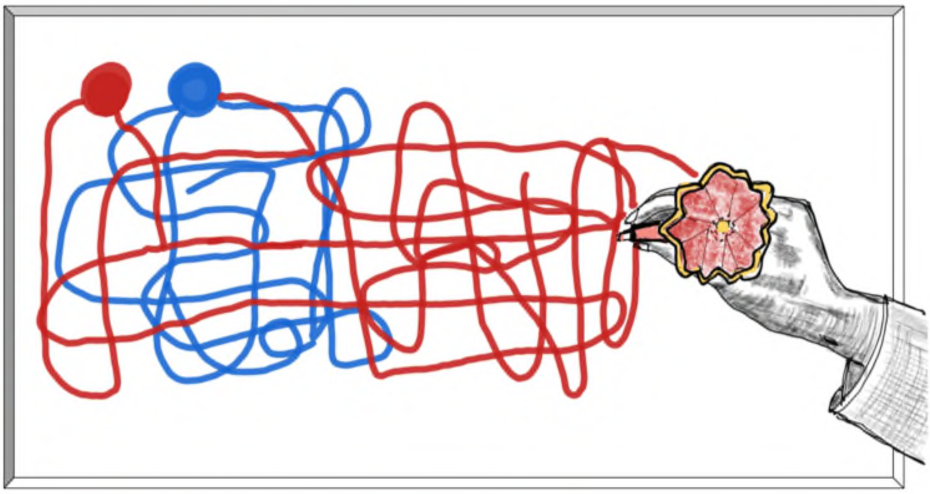


Fig. 4: Sketching sound

In addition, user gestures can be used to trigger specific components of the soundscape. As an example, illustrated in Fig. 4, a soundscape can accompany a pattern that the user sketches out. As another example, a particular sound can be triggered when a user points at a specific semantically-identified object.

With user permission, additional information about the user’s context can feed into any of these modes of control. For example, the location or time-of-day can be used to affect the underlying material of the soundscape that the user is controlling. The specific semantic categories of objects that are used for sonic control and shaping can be assigned by an application developer, or be made flexible and assignable by the user.

In this manner, by mapping the user’s visual field to an ensemble of sounds, the described techniques extend the interaction of the user with their world beyond semantic classification of objects. A user can walk through the world and hear a custom soundscape based on what they are seeing, or recast everyday objects as music controllers that shape a sonic experience. The techniques can be applied as a solo experience of a generated soundtrack (or musical exploration), or, can be utilized in performance contexts or group experiences. Example applications include music technology, education, play-experiences (e.g., at museums) for children and adults, interactive technology, etc.

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 surroundings, camera feed, 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

This disclosure describes techniques to generate an audio experience or soundscape corresponding to the visual field of a user. With user permission, objects within the feed of a head-mounted camera are semantically identified using computer vision techniques. Based on the detected objects, a unique audio experience, shaped by the world around a user and by the physical items they engage with, is generated.

REFERENCES

1. https://makeymakey.com/
2. https://reactable.com/
3. https://dspace.mit.edu/handle/1721.1/36161