Synesthesia AR: Creating Sound-to-Color Synesthesia in Augmented Reality

Shashaank N* Steven Feiner†
Department of Computer Science
Columbia University
New York, USA

Figure 1: Paintings created through an interactive sound-to-color synesthesia experience in Synesthesia AR.

Abstract

Sound-to-color synesthesia is a neurological condition in which people experience different colors and shapes when listening to music. We present an augmented reality application that aims to create an interactive synesthesia experience for non-synesthetes. In this application, users can visualize colors corresponding to each unique note in the 12-tone equal-temperament tuning system, and the auditory input can be selected from audio files or real-time microphone. A gestural hand-tracking interface allows users to paint the world space in visualized synesthetic colors.

Keywords: Augmented Reality, Synesthesia, Chromesthesia

Index Terms: Computing methodologies—Computer graphics—Graphics systems and interfaces—Mixed / augmented reality; Human-centered computing—Human–computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality; Human-centered computing—Human–computer interaction (HCI)—Interaction techniques—Gestural input

1 Introduction

Synesthesia (from the Greek meaning "to perceive together") is a neurological condition affecting ~4% of the world's population [1], in which the stimulus of one sense involuntarily triggers other senses. People with this condition (referred to as "synesthetes") often mentally superimpose visual images [2], such as associating words/numbers with colors. One common form of this condition is sound-to-color synesthesia (also known as chromesthesia), in which listening to music triggers associations with different colors and sometimes shapes. Previous demonstrations of sound-to-color synesthesia involve synchronization of music with colorful animations presented interactively on a 2D monitor [3] or in virtual reality (VR) [4]–[6]. Other work has explored color-to-sound synesthesia in augmented reality (AR) by processing camera imagery to generate audio [7].

We present Synesthesia AR, an AR application that aims to treat sound-to-color synesthesia as an interactive art medium. Users are presented with a multisensory experience akin to that of a synesthete, with auditory input triggering visual stimuli. They can make the world their canvas with a simple but expressive hand-tracking interface to paint the world space with colors generated from the auditory input (Figure 1). While there are different variations of sound-to-color synesthesia, Synesthesia AR implements note-to-color synesthesia, where each unique note (called a pitch class) in the 12-tone equal-temperament (12-TET) tuning system is associated with a different color.

2 System

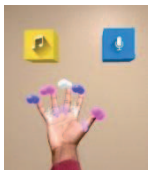
Synesthesia AR was designed for Snap Spectacles 2021 [8], lightweight eyewear that can display stereoscopic optical–see-through AR within a 26.3° diagonal field of view (FOV). It was developed in JavaScript using the Lens Studio AR engine from Snap [9].

The sound-to-color synesthesia pipeline in the application has two different input types: MP3 audio files and real-time microphone input. Audio files are manually transcribed to determine the pitch class and duration of notes in each polyphonic voice; this information is used to synchronize audio playback with synesthesia animations. Real-time microphone input is processed from the Spectacles microphone at a sampling rate of 44.1 kHz in buffers of 2048 samples. Each buffer is processed using an autocorrelation algorithm [10] to detect monophonic pitch and determine the currently playing note. Once the note is identified, it is mapped to a unique color in the program.

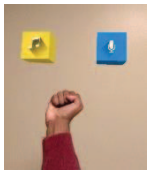
Hand tracking is incorporated using the Snap Spatial Engine built into Lens Studio to return finger positions relative to the world space. Tracking of individual finger joints is enabled to map the synesthesia colors to the bottoms of the fingers for reference (Figure 2) and the fingertips for painting "brushstrokes" (Figure 3). Continuous hand movements are interpreted as paint actions to spawn brushstrokes of different paint materials in the trails of the fingertips.

* shashaank.n@columbia.edu
† feiner@cs.columbia.edu

2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) | 978-1-6654-8402-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/VRW55335.2022.00288

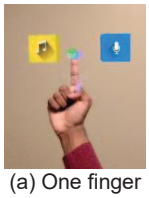


(a) Open Palette

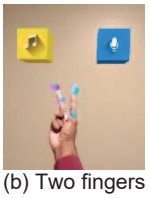


(b) Closed Palette

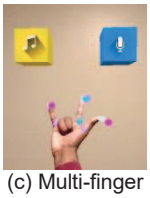
Figure 2: Color palette gestures.



(a) One finger



(b) Two fingers



(c) Multi-finger

Figure 3: Painting gesture examples.

In addition to the hand-tracked painting interface, the Spectacles touchpad is used for simple actions, such as double tapping to select the auditory input for the synesthesia experience (Music Mode or Microphone Mode) or clear the painted brushstrokes in the world space. Music Mode has additional touchpad interactions to control the music playlist (left swipe/right swipe) and play/pause (single tap) the currently playing song.

3 Experience

When the user wearing Spectacles first opens the application, two floating cubes appear at the left and right. These two cubes correspond to the two modes in the application: Music Mode and Microphone Mode. The user can orient their head toward one of the cubes and double-tap the Spectacles touchpad to start the synesthesia experience.

To begin painting, the user must bring a hand up within the Spectacles tracking space. The palette of available painting colors is then displayed at the bottoms of the fingers (Figure 2a), and the palette is synchronized to the notes of the currently playing MP3 or microphone audio. When the user actively moves one or more of their fingers, the application paints "brushstrokes" near the fingertips that match the colors presented on the palette. The user can stop painting brushstrokes by not moving their fingers and can hide the palette by closing their fingers (Figure 2b). To clear the painted brushstrokes from the world space, the user can orient their head toward a red cube anchored at their feet and double-tap the Spectacles touchpad. As each finger can paint brushstrokes independently, the user is not limited to keeping their entire hand open and can use a wide variety of expressive hand gestures that selectively open and close different fingers (Figure 3).

In Music Mode, the colors are synchronized to predefined song excerpts stored within the application. The user can play/pause the current song using a single tap on the Spectacles touchpad and can switch between songs using "left swipe" or "right swipe" actions on the Spectacles touchpad. Depending on the number of polyphonic voices in the selected song, the palette displays multiple colors corresponding to each voice. In Microphone Mode, the colors are synchronized to real-time microphone input from the user, which can include singing, speaking, or playing an instrument. As this mode supports only monophonic input, the palette displays the same color on all five fingers.

4 Conclusions

We presented Synesthesia AR, an augmented reality application for Snap Spectacles 2021 that enables users to experience sound-to-color synesthesia. The application uses songs stored as audio files and real-time microphone input from the Spectacles to map each unique note in the 12-TET tuning system to a color. While there has been previous work in this field to create interactive visualizations of sound-to-color synesthesia, Synesthesia AR extends sound-to-color synesthesia into the realm of interactive art by including a simple and intuitive hand-tracked interface to translate the synchronized colors from music into paintings.

The scope of this application is currently limited by the Spectacles' hardware constraints, as an 8 MB maximum application size allows the Spectacles to store only short 15–20 second excerpts of songs. Accessing an online music database using a REST API is currently not possible using the Snap software. Furthermore, the Spectacles' hand-tracking technology can track only one hand at a time; this results in songs supporting a maximum of five polyphonic voices using the current 1:1 voice-to-finger mapping, which narrows the option of painting songs with multiple, nuanced layers.

Future work could include incorporating a faster real-time pitch-detection algorithm (monophonic and polyphonic) using machine learning, as the current JavaScript autocorrelation algorithm has some lag due to the limited processing power of the Spectacles. Additional improvements to enhance the user experience could include the ability to "touch" brushstrokes in the projected environment with each hand to modify their position, as well as a user interface to customize the painting, for example by changing the size and shape of brushstrokes and modifying the synesthesia color palette.

Acknowledgments

We thank Snap Inc. for providing Snap Spectacles 2021 for this project. We thank Ted Brown and Ilter Canberk at Snap Inc. for their technical assistance in using the Spectacles with Lens Studio.

References

[1] M. J. Banissy, C. Jonas, and R. Cohen Kadosh. "Synesthesia: An introduction." *Front. Psych.* 5:1414. Frontiers, Dec. 2014.

[2] J. Simner. "Why are there different types of synesthete?" *Front. Psych.* 4:558. Frontiers, Sep. 2013.

[3] M. J. Adiletta and O. Thomas. "An artistic visualization of music modeling a synesthetic experience." *arXiv*. Dec. 2020. https://arxiv.org/abs/2012.08034 (accessed Jan. 10, 2022).

[4] Denkwerk. "Synesthesia VR." Synesthesia.world. https://www.synesthesia.world/en/vr-experience (accessed Jan. 10, 2022)

[5] A. Kose, A. Tepljakov, S. Astapov, D. Draheim, E. Petlenkov, and K. Vassiljeva. "Towards a synesthesia laboratory: Real-time localization and visualization of a sound source for virtual reality applications." *J. of Comm. Soft. And Sys.*, 14, 112–120, Mar. 2018.

[6] Y. Konishi, N. Hanamitsu, K. Minamizawa, B. Outram, A. Sato, and T. Mizguchi. "Synesthesia Suit: The full body immersive experience." *Proc SIGGRAPH '16,* Jul. 2016, p. 1, doi: 10.1145/2929490.2932629.

[7] M. Benitez and M. Vogl. "SARA: Synesthestic augmented reality application." *Proc ISWC '14 Adjunct,* Sep. 2014, pp. 21–25, doi: 10.1145/2641248.2641277.

[8] Snap Inc. "New Spectacles." Spectacles.com. https://www.spectacles.com/new-spectacles/ (accessed Jan. 10, 2022)

[9] Snap Inc. "Lens Studio." Snapchat.com. https://lensstudio.snapchat.com/ (accessed Jan. 10, 2022).

[10] D. Antonios. "Fundamental frequency estimation for music signals." B.S. Thesis, Dept. Info. Eng. and Comm. Systems, Univ. Aegean, Mytilene, Greece, 2015. [Online]. Available: https://hellanicus.lib.aegean.gr/handle/11610/8650