Enhancing Embodied Telepresence in First-Person View Drone Pilots through Real-Time Data-Driven Sound Design

Samuel Adam SWOPE
The Hang Seng University of Hong Kong
Hong Kong
samuelswope@hsu.edu.hk; swwwope@gmail.com

Abstract

This academic paper investigates the integration of real-time, data-driven sound design and construction within the context of first-person view (FPV) drone piloting. The objective of this study is to enhance the sense of embodied telepresence experienced by FPV drone pilots. The paper addresses several research gaps and questions, including the potential for shared experiences between the human body and the drone machine, the impact of different sound designs on the overall piloting experience, and the influence of new ocular and aural feedback on the pilot's understanding of aerodynamic position and movement. Additionally, the paper examines the limitations of current human-drone sensory overlay and emphasizes the necessity for a widely applicable methodology for aural sensory stimulation in drone operation. The research methodology encompasses the development of a physical prototype and the exploration of flight controller data sonification. The paper concludes with an evaluation of the prototype and the identification of areas that require further research and improvement, laying the foundation for integrating sound design into FPV drone control systems and opening up new possibilities for creating immersive and dynamic airborne art.

Keywords

Art and technology, Embodied telepresence, First-person view (FPV) drone piloting, Real-time data-driven sound design, Algorithmic composition, Physical computing, Human-machine relationship, Spatial awareness, Multisensorial interface, Sonification of flight controller data.

Introduction

As human-drone interaction becomes increasingly prevalent, there is an essential need for continual alignment between artistic creation and technological development to facilitate the evolution of airborne art into a more captivating and dynamic medium. This alignment requires the integration of innovative hardware and software formats, such as aerial vehicles, sensors, and machine learning.

The purpose of this research project is to establish a new methodology and theoretical framework that enhances the embodied telepresence experienced by first-person view (FPV) drone pilots through the integration of real-time, data-driven sound design and construction. By incorporating electroacoustic transducers, such as headphones, and employing innovative methods derived from physical computing, FPV drone pilots experience real-time audio synthesis based on algorithmic composition. This synthesis is mapped from in-flight data sourced from drone sensors, aerobatics, and operator inputs. The resulting auditory output enables FPV drone pilots to transcend the limitations of ocularcentrism in drone control and access an aural landscape that induces a heightened sense of spatial awareness.

In FPV drone operation, there are three main components: a remote controller, often a radio transmitter with 2 joysticks, for controlling the drone while in flight; goggles that cover the pilot's eyes and display immersive live video; and the drone, which is often a quadcopter and can vary in size and power. The drone is equipped with a forward-facing camera capable of wirelessly streaming real-time video from the drone's perspective to the pilot's goggles. The pilot, who can now only see from the drone's point of view, must keep their mind and transmitter input synchronized with the drone to keep it airborne and fly with precision. Professional FPV piloting considers a range of flying styles, including racing, freestyle, long range, and filming. Racing competitions involve pilots navigating through challenging courses at high-speed. Freestyle focuses on acrobatic maneuvers, allowing pilots to showcase their creativity by performing impressive tricks. Long range pushes the limits of the drone’s range capabilities while exploring vast distances. Filming concentrates on capturing cinematic shots and footage for various purposes, such as movies, commercials, or other projects that prioritize high speed live action and a wide range of perspectives.

Research Gaps and Objectives The significance of this research lies in reimagining flight and air as aesthetic systems and creating a more immersive and dynamic airborne art by aligning artistic creation with innovative drone technologies. By expanding the sensory dimensions of aerial art beyond the visual realm, the research development aims to enrich spatial experience and contribute to the discourse on art and technology which considers airborne systems.

There is no broadly used system that enables an FPV pilot to experience drone control and operation with an acoustic element. This project aims to explore the effects of real-time, data-driven sound design on the embodied telepresence of FPV drone pilots. It seeks to address key research questions related to the potential for an out-of-body experience in FPV drone pilots, the possibility of shared experiences between the human body and the drone machine, and the correlations between the human-machine relationship and spatial awareness. The project aims to determine whether the addition of data-driven sonics to FPV drone control systems can foster shared experiences between the pilot's body and the machine. Additionally, it aims to identify which types of drone sensors, algorithms, techniques, and sounds can enhance the aerial experience. The project will also examine how different sound designs can impact the overall experience. Furthermore, it will investigate how the pilot's understanding of aerodynamic position and movement may change with the introduction of new ocular and aural feedback.

Human-drone Sensory Overlay and its Current Limitations FPV drone operation has garnered adoption as a hobbyist and professional sport as well as in commercial settings. Drones are frequently deployed in a variety of roles: racers, remote inspection units, tools for landscape surveys, and more. However, since the proliferation of affordable drones in the 2010s, the methods of control have been exclusively dedicated to sight, with FPV drone pilots primarily relying on an onboard camera to determine the trajectory of their controlled unit. Other forms of sensory output such as audio feeds have been largely absent from the control process. Consequently, drone operation remains an unintuitive affair that requires specialized training. Additionally, it remains an exclusionary medium in that visually impaired individuals are unable to effectively participate in drone operation.

As FPV drones are deployed in even more contexts, including infrastructure inspection and surveys in the energy, agriculture, mining, security, and other sectors, it is crucial to ensure that the control environment for FPV drone operators advances to facilitate an effective and meaningful piloting experience so that the data recorded via drone flight is maximally useful. Once developed, these advancements can lead to a paradigm change in human drone operation, including a more intuitive operational setting that can lower the barrier of entry to drone operation. Additionally, this can lead to the formation of new applications for drone deployment, giving researchers, engineers, creatives, and commercial stakeholders such as entities in the entertainment and filmmaking industries an extra dimension to the medium to build upon while they formulate innovative technologies and methods to use drones that employ audio sensors and sonic composition that give drone operators a meaningful audio feedback during drone flight. This formulation can also provide visually impaired individuals a means to become drone operators, extending their reach to experience the world through kinetic aerial devices.

The presentation of drones in cultural contexts has so far been limited to the physical and visual aspect, with other sensory information inaccessible to viewers. This has limited the ways in which drones can be used in many contexts, whether in the development of culturally significant settings, commercial applications, or in a significant portion of engineering and academic research related to the operation and control of FPV drones. Research related to the incorporation of sound for the pilot's benefit during drone operation has thus far been unable to attain a quality and consistency that would make the addition worthwhile and meaningful. However, it remains crucial to pursue this line of research and development, as the benefits that aural signals can facilitate for drone pilots in commercial and creative contexts are expected to lead to significant paradigm changes in terms of drone application as well as our understanding of the sensory convergence of humans and machines.

In the quest to develop a method that can be widely distributed, a tangible model was constructed to capture signals and data during drone flights. A qualitative approach, employing iterative design methodologies, has been selected to effectively identify the various types of drone sensors and sounds that enhance an immersive and out-of-body aerial experience. Additionally, this approach aims to analyze and present connections between the human-machine relationship and spatial awareness, which are influenced by aerodynamics and sonic composition. Ultimately, this research endeavor aims to contribute to the existing knowledge surrounding human-drone interaction.

Review

Rendering Flight and Air as Aesthetic Systems The concept of rendering flight and air as mediums for art has emerged as a way to distance them from their traditional scientific understanding and reimagine them as aesthetic systems. One of the earliest instances of using air as a medium in art can be traced back to Marcel Duchamp's creation of *50cc of Paris Air* in 1919, a Dadaism sculpture consisting of a glass ampoule filled with air purchased by the artist from a pharmacist in Paris. While this artwork raises questions about the preservation of its gaseous contents over time, it signifies a shift in perceiving air as an active and perceptible medium within the realm of art.

In the 21st century, the integration of innovative technologies, such as drones, sensors, and machine learning, has opened up new possibilities for creating dynamic and spectacular airborne art. By constructing and controlling aesthetic systems that engage with air, artists can explore the artistic, dynamic, and technical aspects of flight and air as aesthetic dimensions. This approach goes beyond visual representation and aims to create a more immersive and multi-sensorial experience.

Multisensorial Interface in Drone Perception To fully enhance the viewing experience of aerial art, it is essential to develop technologies that allow for the perception of art beyond the visual dimension. While aerial sculptures and installations can be visually striking, they often lack sensory stimulation beyond sight. To address this limitation, a multidisciplinary approach is required, integrating various disciplines into an aesthetic system that enables the perception of multiple dimensions in art. This concept draws inspiration from research on "systems esthetics" in the 1960s, which highlighted the shift in art towards responsive environments and the reflection of technological dynamism. [1]

The aerial view, now coupled with innovative machine vision and artificial intelligence, transforms the perception of drones as objects and the images which they produce. While aerial imagery is commonly associated with warfare and surveillance, autonomous airborne art relies on complex aerial views supported by sensors, cameras, and code. The behaviors exhibited by these flying artworks, guided by their creators, evoke the notion of behavior and intelligence. This raises questions about the behavior and intelligence of aerial drone art. Computer scientist Valentino Braitenberg explored the idea of machines or vehicles exhibiting behaviors in a natural environment, challenging the notion of what it means to be alive. [2] However, it is important to note that their intelligence does not mirror human cognition, but rather operates according to predetermined rules established by the creators, or pilots, themselves.

When a human operator controls an FPV drone, the perception of the drone merges with that of the pilot through a first-person view headset. Understanding the limitations of this sensory convergence is crucial for further advancements in the deployment of FPV drones in various contexts, whether commercial or creative, or for research or recreation.

Beyond Visual Sensory Treatment of Drones Recent studies have examined aerial drones as part of a wider sensorial assemblage that generates non-human sensing for human operators. [3] This perspective acknowledges that drones facilitate a multisensorial interface, expanding human capabilities to interpret physical surroundings. By adding significant verticality and increasing the speed at which distances can be traversed, drones provide operators with a new level of spatial awareness. This is made possible through the interpretation of machinic data signals, which are then visually presented to the pilot. As a result, the aesthetics of drone movement become imprinted upon the operator, shaping their perception of the environment. [4] While these aesthetics are predominantly visual, they have gained significant attention in mainstream media due to the military and entertainment applications of drones. [5]

Although much of the existing research focuses on the visual aspects of drone movement, some studies have explored the potential for drones to offer sensory experiences beyond vision alone. Viewing drones as a form of media technology, researchers have investigated whether these devices can be considered "more-than-optical" objects that can provide additional sensory stimulation. [6]

Developing a Methodology

Current Framework for Aural Sensory Stimulation in Drone Operation Under the current framework of FPV drone piloting, there is no broadly used system that enables a pilot to experience drone control and operation with an acoustic element. In fact, pilots are typically bound by visual output, depending on a single sense to control their device remotely. There have been some experiments performed to explore the possibility of creating an aural dimension that augments the ocular feed that FPV drone pilots have become accustomed to. Some pilots go to the lengths of wearing earbuds to focus on the audio aspect of drone control. [7] This involves onboard microphones transmitting real-time audio to pilot goggles, but this sound is limited to noises generated by the drone's motor and propellers, as well as wind that the drone encounters while traveling through the air.

Due to the distracting nature of these sounds, FPV drone pilots have opted to suppress and eliminate all noises while operating their devices, describing a preference for an entire absence of microphones and audio feeds in their hardware setup. While pilots have become accustomed to this ocularcentric mode of drone operation, there have been attempts to add additional sensory signals that create an immersive control environment, leading to the rejection of the previous proposed solution to provide signals to pilots in audio form. Other research into drone noise has taken place with the exploration of active or passive noise filters, while research has been published to suggest that the application of 3D binaural audio capture from a drone may enhance immersive consumption of news media content. [8]

To date, research on ex-visual sensory treatment of drones has made progress in determining the technical implementation required to create a meaningful and intuitive aural environment for pilots and viewers. This includes hardware tests for microphones to determine the most suitable and effective way of gathering aural data input. However, further research is needed to develop a widely distributable methodology for implementing aural sensory stimulation as perceived by drones. This methodology would enable the creation of a consistent and immersive aural experience that complements the visual aspects of drone flight. By integrating aural stimulation, operators and viewers can enhance their spatial awareness and deepen their engagement with the drone's multisensorial capabilities.

Sonification of Flight Controller Data An innovative work that merges sound design and a sculptural drone was created for the seventh edition of sound art series, *Haptic Collision* (2022). The resulting performance was staged for 100 invited guests, with its ranks including academics, artists, musicians, writers, and curators from varying Hong Kong sectors and communities. The artwork and performance involves a custom, silver foil, helium assisted drone driven by an autonomous flight controller and human interrupt. A three-axis accelerometer, three-axis gyroscope, and five time-of-flight sensors return in-flight data such as the drone's roll, pitch, yaw, acceleration, altitude, and spatial position as well as the drone's autonomous decisions. The full data package is wirelessly transmitted for algorithmic processing by a computer that is set up remotely. This leads to real-time audio granular synthesis and algorithmic composition as output for an audience and cultural context (figure 1). The *Haptic Collision* performance yielded significant outcomes, encompassing comprehensive data collection, meticulous analysis, and a solid foundation for ongoing research and development endeavors. The show *Nervous Thrasher* (2024) is a good case in point (figure 2). The methodology as described with the following prototype departs from the *Haptic Collision* and *Nervous Thrasher* model and expands to introduce FPV systems and covers a comprehensive range of outcomes (figures 3 and 4).

Figure 1. *Haptic Collision #7.* Live performance documentation. ©Samuel Swope. 2022

Figure 2. *Nervous Thrasher*. Live performance documentation. ©Samuel Swope. 2024

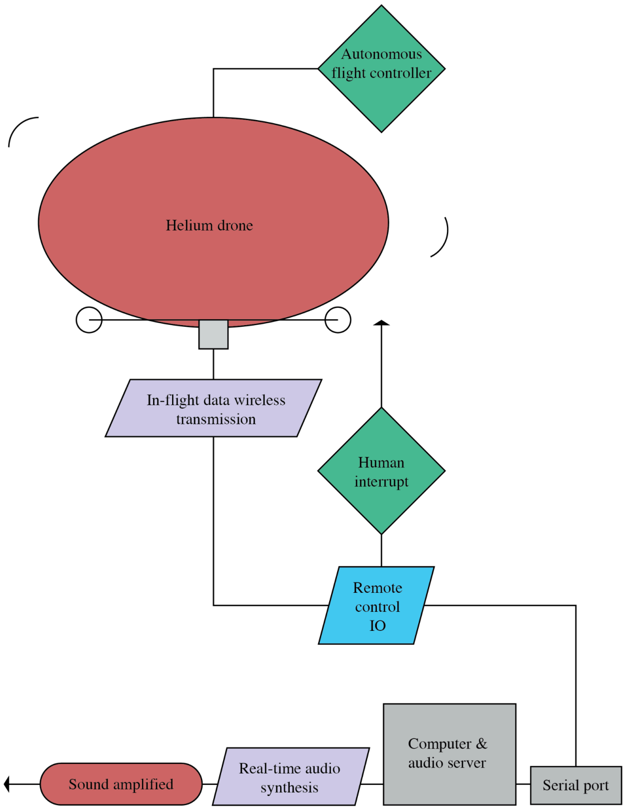

Figure 3. *Haptic Collision* #7 and *Nervous Thrasher* flow diagram chart. ©Samuel Swope. 2023

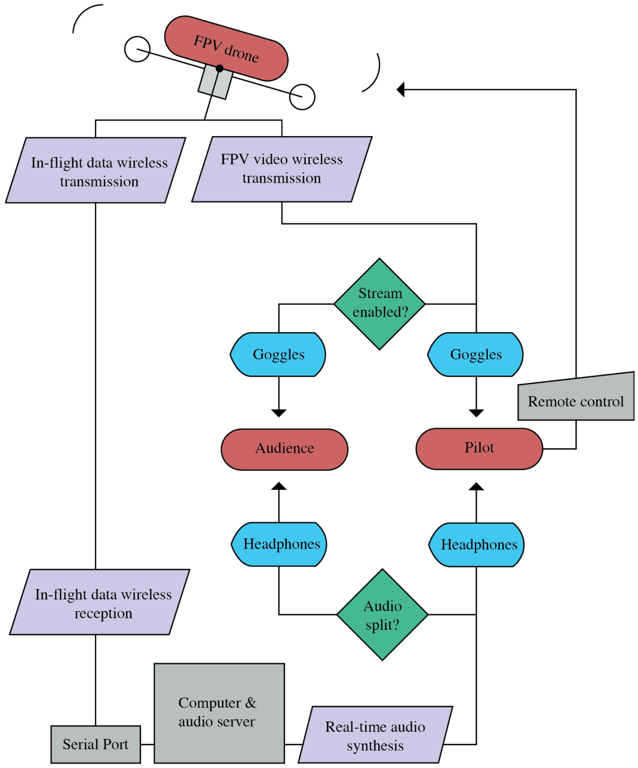

Figure 4. Prototype flow diagram chart. ©Samuel Swope. 2023

Prototype

Transmitter and Receiver The first prototype for this study was developed to investigate the research concept, evaluate quality, and establish a foundation for future engineering iterations. The prototype consists of two digital devices: a transmitter and a receiver.

The transmitter is encased in a thermoplastic polyurethane (TPU) 3D print and is designed to piggyback on an existing AXIS Flying CineWhoop 3” FPV drone with a DJI O3 Air Unit. This piggyback configuration allows the transmitter to be mounted on top of the FPV drone, similar to how a pilot would attach an external camera (figure 5). The transmitter can be powered either by an external 3.7V LIPO battery or by routing power from the FPV drone's LIPO battery through a 36V DC to 5V DC buck converter. The choice of power source provides versatility or reduces the overall payload, respectively. The transmitter includes a XIAO ESP32S3 microcontroller (MCU), a 2.4Ghz antenna with an IPEX connector, and a GY521 module with an MPU6050 sensor.

The receiver, also encased in a TPU 3D print, consists of a XIAO ESP32S3 microcontroller and a 2.4Ghz antenna with an IPEX connector. The receiver is connected to a MacBook Pro laptop computer via a USB serial port.

In terms of video and audio hardware, the pilot uses DJI Goggles Integra, which cover both eyes and display high-definition video wirelessly streamed from the DJI O3 Air Unit. Simultaneously, the pilot wears headphones connected to the laptop via the headphone port to receive the audio feed.

The software used for this setup is SuperCollider, a real-time audio synthesis and generative sound art program. SuperCollider consists of three major components: scsynth (a real-time audio server), sclang (an interpreted programming language), and scide (an editor for sclang with an integrated help system). The data from the transmitter is wirelessly streamed to the receiver and then mapped by SuperCollider for audio control using Open Sound Control. The transmitter's data, which mirrors the flight controller data of the FPV drone, is generated by custom code compiled within the Arduino IDE. The code is tailored to match the processes and algorithms of the FPV drone in either angle or acro flight mode.

Figure 5. Transmitter piggyback configuration and other components of the system at Nam Sang Wai. ©Samuel Swope. 2023

Algorithm Development for Sound Art Output To create the aural output, the pilot hears sound generated from the drone's aerobatic maneuvers, such as roll, pitch, yaw, angle, and acceleration. These parameters are mapped to control various audio parameters, including frequency, pitch, amplitude, duration, and position. The drone's proportional-integral-derivative loop (PID) controls audio processing techniques such as granular synthesis, effects, and filters.

Testing The testing of the system was conducted at a remote control vehicle playground in Nam Sang Wai, Hong Kong, where the audio and video signal processing proved consistently strong. Four unique sound designs were developed for real-time generative compositions during the study.

1. One single synthesizer is a band-limited sawtooth wave generator. This synthesizer is controlled by flight data, which is mapped to various parameters

such as frequency, detune, boost, amplitude, maximum resonance, and spread. The resulting sound from this synthesizer can be described as resembling the sound of wind or a culturally recognized and familiar sound commonly associated with flying.

2. A free interpretation composition incorporating various instruments, including a lead flute, saw bass, violin, hats, and other audio samples. Each instrument is controlled by a Pbind, which is a SuperCollider programming construct that combines multiple value streams into a single event stream. In this context, the in-flight data updates the arguments associated with each Pbind, allowing for dynamic changes in the composition.
3. A composition that employs granular synthesis techniques, utilizing a collection of thirteen samples extracted from an 808 drum kit and stored in a buffer. These samples are played in a sequential manner, creating rhythmic patterns. The in-flight data is responsible for controlling various parameters of the granular synthesis, including the duration and position of each grain, the playback rate, and the trigger rate. These controls allow for dynamic manipulation of the synthesized sounds in response to the changing flight data.
4. A composition featuring a drone arpeggio in the key of A minor, utilizing the notes A, C, E. The in-flight roll position dynamically controls which note is played, while pitch controls each note's amplitude and frequency, all set to a tempo of 140 beats per minute. Additionally, a collection of twelve sound effects is stored in a buffer, and their triggering is determined by flight maneuvers. An algorithmic control system governs the rate, amplitude, and frequency of these effects before they pass through a reverb, enabling responsive and dynamic manipulation of the audio output.

In each of the four sound compositions, the roll position of the drone was systematically mapped to the left and right audio channels (figure 5).

A skilled professional FPV pilot, experienced in racing, freestyle, and inspection, participated in this pilot study. The pilot was asked to fly freely while testing the prototype and all four sound compositions. Of the four sounds, the pilot preferred sound design number one and two. When describing number one, the pilot spoke phrases such as "flying in a spaceship, ear-catching, adrenaline rush, and can hear the wind" and stated that the sound is suitable for racing or freestyle. When describing number two, phrases spoken were "like a bird, can feel the position, much more synchronized, spatially interesting, humorous, and relaxing."

The pilot ended both flights early when testing design number 3 and quickly commented that the sound did not match the experience of flying. While number 4 was exciting and brought to mind the feeling of gameplay, possibly in a galaxy scenario, the triggering logic of the sound effects at certain instances was disturbing.

Overall, the pilot spoke about syncing throttle control with an additional audio parameter, and while the system did have an immersive psychological effect, some sound designs may better fit certain flying scenarios because of the visual and aural overlay. Improving latency inconsistencies between the video stream and audio stream would help the pilot feel the quad more. A more compact and adaptable transmitter would allow any pilot to mount it in different locations of any quad, for example, in the center or the tail. Lastly, the pilot commented that new generations of FPV pilots will appreciate this new system and suggested that synthesized and synchronized sounds of motors and propellers would help pilots, especially those who have grown accustomed to listening to the mechanical sounds of their quads from afar, adapt to the new spatial configuration.

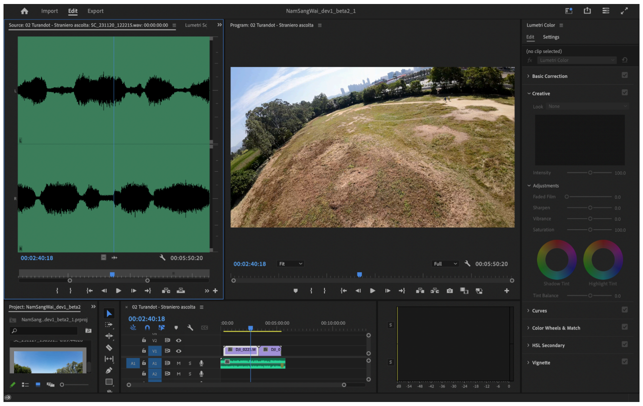

Figure 5. FPV drone footage with correlating left and right audio channels. ©Samuel Swope. 2023

Evaluation Evaluation of the system, technology, and sound design revealed qualitative findings that highlighted the need for further improvements in system range and latency. The elimination of the laptop computer would enhance the portability and versatility of the system. The introduction of sound art in the drone flying experience resulted in a qualitative shift in spatial understanding and the potential influence of sound design on flying style and emotion, or vice versa.

Future iterations of the system could explore the addition of other sensors, such as time-of-flight modules for proximity-based sound and barometric pressure sensors for altitude-based audio. Optical flow sensors could also be incorporated to calculate movement and translate it into sound. Algorithms and sensors that detect variations in throttle would enhance the pilot's perception of the quad's power by providing auditory feedback from their throttle control.

The hardware testing and design phase of this research methodology provided valuable insights into the implementation of aural sensory stimulation for drones. The findings highlight the need for further research to refine the system, explore alternative hardware and software solutions, and expand the variety of soundscape and compositions. The distribution of the developed framework and the evaluation of future iterations of the system with professional FPV pilots remain crucial steps in advancing the understanding of the potential of aural output in creating an immersive control environment.

Conclusion

This research project aimed to enhance the embodied telepresence experienced by FPV drone pilots through the integration of real-time, data-driven sound design. By incorporating sound into the FPV drone control system, the project sought to address research questions related to shared experiences between body and machine, the impact of different sound designs, and the changes in the pilot's understanding of aerodynamic position and movement.

The study reviewed the current limitations of human-drone sensory overlay, emphasizing the need to expand beyond visual representation and explore the multisensorial aspects of aerial art. It also discussed the potential for drones to provide sensory experiences beyond vision alone and the importance of developing a methodology for aural sensory stimulation in drone operation.

A physical prototype was created to test the integration of sound into FPV drone control systems. The prototype consisted of a transmitter, receiver, and software for real-time audio synthesis. Testing and evaluation of the system revealed the need for further improvements in system range and latency, as well as the potential influence of sound design on flying style and emotion.

Future iterations of the system could explore the addition of other sensors and further refinement of the hardware and software solutions. The distribution of the developed framework and continued evaluation with professional FPV pilots would enhance future research and impact.

Overall, this research project has laid the foundation for integrating sound into FPV drone control systems and has opened up new possibilities for creating immersive and dynamic airborne art. By expanding the sensory dimensions of aerial art, this study contributes to the discourse on art and technology and enriches our understanding of human-drone interaction.

References

[1] Jack Burnham, “Systems Esthetics,” *Artforum International*, September, 1968.

[2] Valentino Braitenberg, *Vehicles: Experiments in Synthetic Psychology* (MIT Press, 1984), 51.

[3] Daniela Agosthino, Kathrin Maurer, Kristin Veel, “Introduction to the Sensorial Experience of the Drone,” *The Senses and Society*, Vol. 15, No. 3, 251-258, https://doi.org/10.1080/17458927.2020.1820195

[4] Svea Breaunart, “Shifting the pattern: Lateral thinking and machine vision,” *The Senses and Society*, Vol. 15, No. 3, 259-271, https://doi.org/10.1080/17458927.2020.1820186

[5] Steen Ledet Christiansen, “Unruly Vision: synesthetic space: drone music videos,” *The Senses and Society*, Vol. 15, No. 3, 286-298, https://doi.org/10.1080/17458927.2020.1814565

[6] Maximilian Jablonowski, “Beyond drone vision: the embodied telepresence of first-person-view drone flight,” *The Senses and Society*, Vol. 15, No. 3, 344-358, https://doi.org/10.1080/17458927.2020.1814571

[7] Oscar Liang, “Review: TBS Steele Earbud | FPV Headphone (August 19 2018)”, https://oscarliang.com/tbs-steele-earbud/

[8] John V. Pavlik, “ Drones, augmented reality and virtual reality journalism: Mapping their role in immersive news content,” *Media and Communication*, Vol. 8, No. 3, 137-146, https://doi.org/10.17645/mac.v8i3.3031

Bibliography

Jose Martinez-Carranza, “A review on auditory perception for unmanned aerial vehicles,” *Sensors*, Vol. 20. No. 24, https://www.mdpi.com/1424-8220/20/24/7276

Willoughby Sharp, *Air Art* (Kineticism Press, 1968), 5.