SoundSync

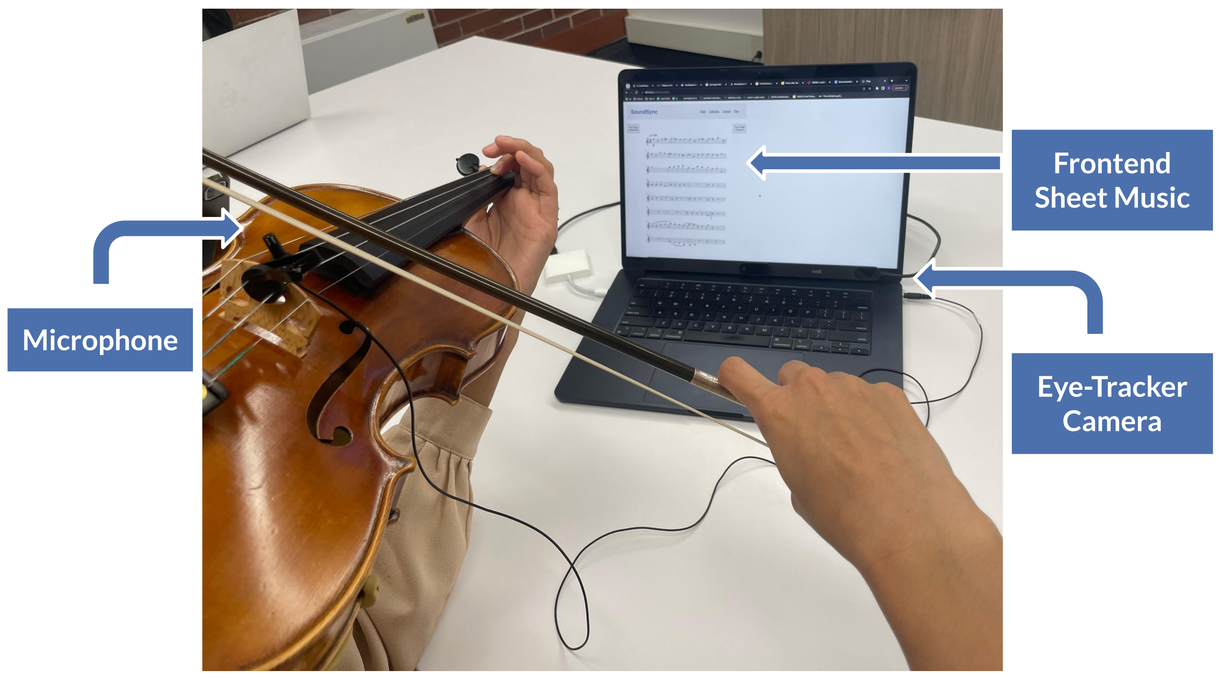
B6: Sanjana Shriram, Rohan Raavi, Caleb Lille
18-500 Capstone Design, Fall 2023
Electrical and Computer Engineering Department
Carnegie Mellon University

Product Pitch

SoundSync is a novel system for automatically turning digital sheet music for musicians. While playing, musicians have to turn pages, resulting in a loss of focus and inability to make music while using their hands to flip a page. Musicians may have the choice of having a professional page turner or foot pedal, however, both are ultimately disruptions and may fail. SoundSync aims to seamlessly and autonomously turn digital pages in real time using eye tracking and audio processing to ensure focused playing for the duration of a practice session.

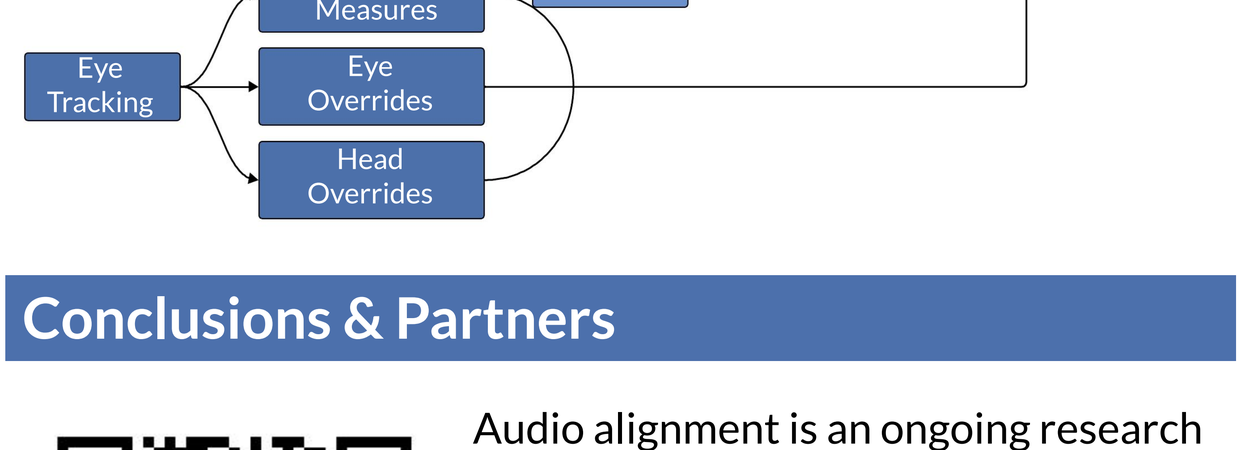
System Description

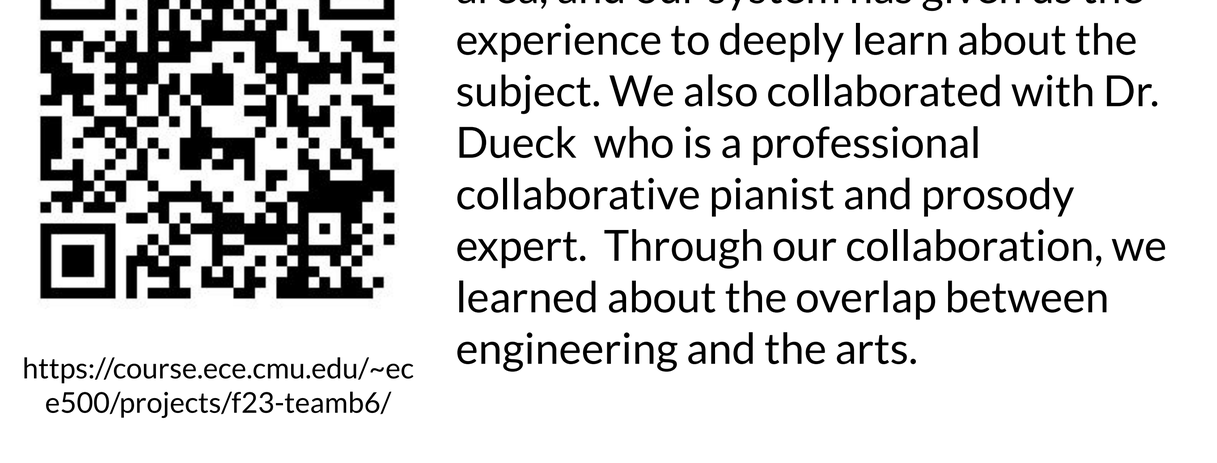
The system incorporates a combination of eye tracking and audio alignment to pinpoint where the user is playing to determine when to flip the digital page. The Tobii Eye-Tracker 5 Camera feeds eye gaze data into a heuristic model. There are 3 algorithms used to align audio: Midi Align, Harmonic Midi Align, and Dynamic Time Warping. We parallelized these algorithms to balance the tradeoffs between latency and accuracy. Data from the eye tracker is fed into a rules-based heuristic function that computes a predicted position in the music. Additionally, override features are tracked using eye position and head tracking. All signals from the audio and eye algorithms are fed into our Monitor Signals function which weighs the confidence of the signals to output a single cursor value and page flip boolean. Lastly, the frontend shows the sheet music and turns the page automatically as a user plays.



System Architecture

Our system architecture has 3 parts: a setup stage, the backend processing, and the frontend. Users begin by uploading music into MuseScore which is a free online tool that helps them export a standardized version of their sheet music and a MIDI file. After that, users will calibrate their eyes and consent to having their audio and eye tracking data recorded. In the backend, two parallel processes are running - a Python program aligns audio and a cpp subprocess processes eye tracking data to send a page turn signal. The frontend polls for updates from the backend every 20ms and uses jQuery to update a real-time cursor and flip the page.





System Evaluation

Three methods on the right demonstrate the tradeoff between speed and accuracy for audio alignment algorithms. MIDI Align utilizes chorma vectors to align. MIDI Harmonic Align adds an extra step of harmonic-percussive separation to MIDI Align. Dynamic Time Warping finds a warping path between two audio sequences.

Audio Alignment Tradeoffs

Algorithm	Latency	Accuracy
MIDI Align	158 ms	Not robust with incorrect notes, supports missed notes
MIDI Harmonic Align	763 ms	Filters percussive noise, not robust with incorrect notes, supports missed notes
Dynamic Time Warping	500 - 1000 ms	Very robust, needs longer note sequences

Conclusions & Partners

https://course.ece.cmu.edu/~ece500/projects/f23-teamb6/

Audio alignment is an ongoing research area, and our system has given us the experience to deeply learn about the subject. We also collaborated with Dr. Dueck who is a professional collaborative pianist and prosody expert. Through our collaboration, we learned about the overlap between engineering and the arts.

Electrical & Computer ENGINEERING
Carnegie Mellon