

1. Einleitung

Teil des Kurses Audio-Video-Programmierung ist es ein Projekt, welches Videoverarbeitung, und eine Webaudiooberfläche zu verbinden. Diese nutzt die aus dem Video gewonnen Informationen zur Steuerung einer Audioausgabe in Form von Geräuschen und Tönen. Im Bereich Videoverarbeitung werden wir in unserem Fall besonders den Aspekt des Motion-Trackings betrachten.

Unser Projekt ähnelt einem Videoexperiment.

Unser Ziel ist es eine Art Musikinstrument zu bauen, welches mit dem eigenen Körper gespielt werden kann. Unsere anfängliche Idee war es Mimik/Gesichtsausdrücke hörbar zu machen. Allerdings überschreitet vermutlich Gesichtstracking unsere Fähigkeiten. Daher haben wir nach einer etwas passenderen Variante gesucht, welche uns umsetzbar erschienen. Dadurch sind wir zu Ausdruckstanz gekommen, wobei wir uns auf die Bewegung des Kopfes konzentrieren wollen. Eine Inspiration für dieses Projekt war unser ITS Projekt von vor zwei Semestern, in welchem wir ein Theremin gebaut haben. Auch ein Theremin wird ohne Berührung und nur durch Position eines Körperteils gespielt. Im Fall eines klassischen Theremins sind dies Hände. In unserem Projekt wird es der Kopf sein, welcher durch seine Positionen auf der X- und Y-Achse das Instrument steuert. Auch unser IST-Projekt hat nicht die klassische Mechanik eines Theremins genutzt, welches die Positionen der Hände mit Hilfe eines magnetischen Felds ortet, sondern die Positionsbestimmung mit Hilfe von Ultraschall-Abstandssensoren umgesetzt, welche von einem Raspberry Pi in Lautstärke- und Tonhöhenwerte umgerechnet und wieder ausgegeben wurden. Und dieses Mal wollten wir als weiteren Ansatz eine Ortung der Position durch Videoverarbeitung nutzen. Eine weitere Inspiration waren Spiele wie Yasuhati oder Chicken Scream. In diesen Spielen steuert der Spieler seine Figur mit Hilfe von Lautstärke. Je lauter das Geräusch, desto höher springt die Figur. Wir fanden es interessant diese Idee umzudrehen, das wenn man höher springt ein lauterer Geräusch ausgegeben wird, sodass eventuell das Springen der Spielfigur durch das Springen des Spielers gesteuert werden kann.

Ein Ziel unseres Projektes ist unsere persönlichen Programmierfähigkeiten zu verbessern.

2. Projektziel

Unser Endprodukt soll durch die Bewegung des Nutzers steuerbar sein. Dabei kann seine horizontale Bewegung im Blickfeld der Kamera die Tonhöhe steuern und seine vertikale Bewegung die Lautstärke steuern. Im zugehörigen Audiointerface lässt sich aus verschiedenen Soundsamples auswählen oder eigenes Zusammenstellen. Das können Instrumente wie ein Klavier oder eine Blockflöte, einfache Töne wie Sinus oder auch Geschrei sein. Der Nutzer kann sich auch aus Sinus-, Dreieck- und Rechteck-Schwingung ein eigenen Ton zusammenstellen, gesteuert durch Regler im Audio Interface. So kann er beispielweise lauter angeschrien werden je höher er springt.

3. Anforderungsanalyse

Unsere Anwendung sollte möglichst in Echtzeit auf Bewegung reagieren. Die Latenz sollte also möglichst klein sein. Es sollte sich die Tonhöhe und Lautstärke verändern und dieses sollte möglichst passend zu den Bewegungen des Menschen passieren. Die Webaudiooberfläche sollte möglichst einfach und intuitiv zu bedienen sein und ein optisch ansprechendes Design aufweisen können. Außerdem sollte erkannt werden, wenn sich kein Mensch im Kamerablickfeld befindet, sodass nicht zufälligen Geräusche ausgegeben werden.

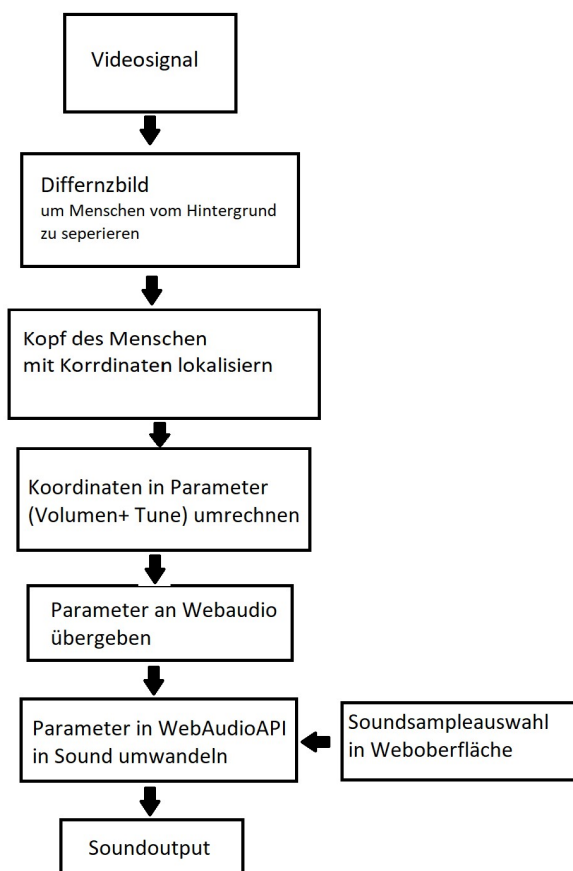
Zusätzliche Funktionen, auf welche wir uns nicht fokussieren werden, aber die man noch später implementieren könnte, wenn noch Zeit vorhanden ist und diese unsere Fähigkeiten nicht überschreiten, sind zum Beispiel das Video zusammen mit dem Ton aufzunehmen, um es auf Social Media zu teilen. Eine weitere extra Funktion wäre eine Art Singstar oder Guitar Hero ähnliches Spiel, bei dem man mit dem Kopf bestimmten Linien folgen muss, um ein Song zu spielen. Eine weitere Möglichkeit für ein Zusatzfeature wäre die Möglichkeit für zwei Menschen gleichzeitig mehrstimmige Geräusche zu erzeugen.

4. Technische Rahmenbedingungen

Unser Hardware-Set-Up besteht aus einem PC mit dem Betriebssystem Windows 10. Zusätzlich werden noch eine Webcam und Lautsprecher an den Computer angeschlossen, wenn sie nicht schon direkt verbaut sind.

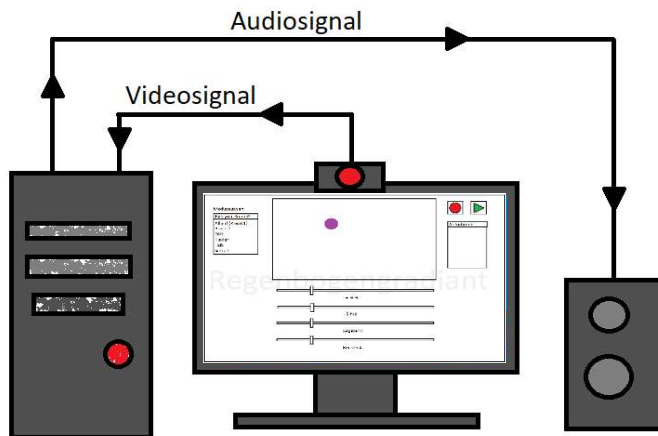
Auf der Softwareseite werden wir zum einen Python mit der Bibliothek openCV, welche zur Bild- und Videoverarbeitung genutzt werden kann, verwenden. Außerdem nutzen wir JavaScript und HTML bzw. die Web Audio API, mit welcher Audiosignale erstellt und ausgegeben werden können und ein grafisches Interfaces zur Bedienung unserer Anwendung erstellt werden kann. Diese Programmier- und Beschreibungssprachen wurden uns vorgegeben.

5. Technisches Konzept



Flussdiagramm

Das Videosignal der Webcam wird im Programm zunächst in ein Differenzbild umgerechnet, um den Menschen besser vom Hintergrund unterscheiden zu können. Dann wird die Position des Kopfes berechnet, in dem der höchste Punkt der sich bewegenden Person bestimmt wird. Diese Koordinaten werden dann entweder im Pythoncode oder im Webinterface in Parameter zur Tonerzeugung umgerechnet um mit diesen dann in der WebAudioAPI Geräusche/Sounds/Töne zu erzeugen. Auch werden bei der Sounderzeugung die Einstellungen im Interface miteingebunden.

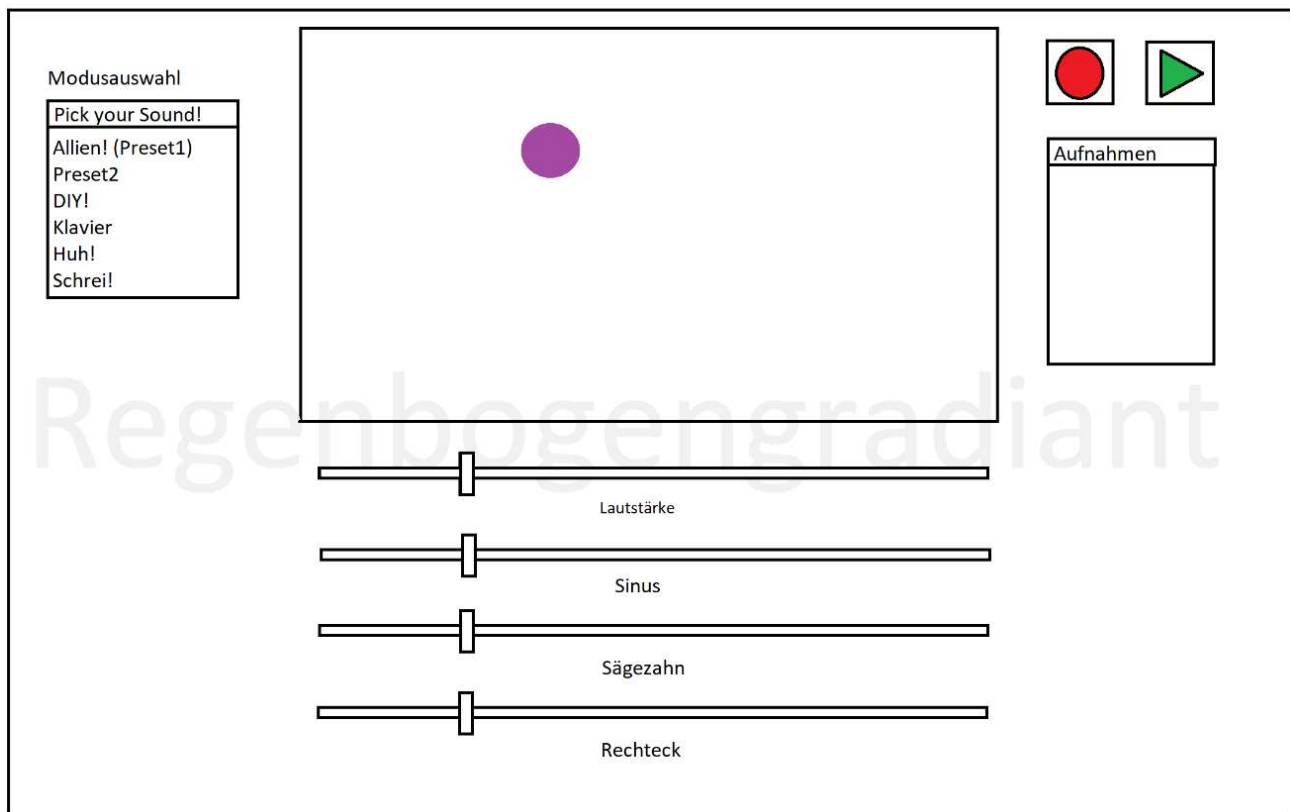


Der Signalweg ist ziemlich übersichtlich. Das Videosignal geht von der Webcam in den PC und das Audiosignal dann vom PC zu dem Lautsprecher.

Hardwaresetup & Signalfluss

6. Bedienkonzept

Unsere Anwendung kann durch Bewegung vor der Kamera und durch das User-Interface im Browser gesteuert werden. Es können Tonhöhe und Lautstärke durch Bewegungen gesteuert werden. Die Soundauswahl ist über das Interface möglich, in welchem es ein Menü zur Auswahl gibt und im Falle des selbst gebauten Sounds zusätzliche Schieberegler, die den Anteil an Sägezahn-, Sinus- und Rechteckschwingung steuern. Wenn wir die Aufnahmefunktion implementieren wird es dafür zusätzliche Bedienelemente zum Aufnehmen und Abspielen geben.



Interface-Entwurf

7. Zeitplan

Hier eine Auflistung, in welcher Reihenfolge wir die Arbeitsschritte angehen möchten:

- Webcam-Einbindung
- Kopf lokalisieren
- Erste Webinterface mit Steuerung von Lautstärke und Tonhöhe über Regler
- Verbindung der beiden Komponenten
- 5.11 Position des Kopfes wird erkannt & einfache Tonsteuerung durch diesen möglich
- Verschiedene Soundsample integrieren
- Verbesserung/Optimierung des Trackings und der Übertragung in Töne
- Interface überarbeiten/erweitern
- Evtl. Mehrnutzeroption integrieren
- Audioaufnahmefunktion integrieren
- Audio-&Videoaufnahme kombinieren und aufnehmen (wenn überhaupt möglich)
- 10.12 Möglichst flüssiglaufende Version präsentieren