

Projektkonzept - MidiMimik

Einleitung

Das Ziel dieses Projektes ist es, die Programmierung von Audio- und Videoinhalten praxisorientiert zu lernen und einen „Facial Expression Audio Synthesizer“ zu entwickeln. Wichtig dabei ist die Verwendung der Programmiersprache C++, denn auch diese soll erlernt werden. Das Projekt findet im Rahmen der Vorlesung AV-Programmierung statt und ist Teil der Prüfungsleistung. Es soll ein „Facial Expression Audio Synthesizer“ erstellt werden, also eine Software, mit der ein Nutzer später durch Verändern seiner Gesichtsmimik Sounds und Effekte steuern kann.

Das Projekt wird nach Fertigstellung im Rahmen einer Ausstellung zusammen mit den Projekten, die in ITS entstehen werden, präsentiert. Hierbei werden neben Studenten auch diverse Firmen eingeladen, um gegenseitig Kontakte zu knüpfen und den neuen Master „digital reality“ vorzustellen.

Dieses Projekt trägt den Titel „MidiMimik“, was die Funktionsweise der Software perfekt widerspiegelt.

Ziel

Das Ziel dieses Projektes ist eine Software, bei der das Gesicht als Interface dient. Die Augen sind „Auswahlknöpfe“, der Mund ein stufenloser Regler. Die Ausgangsposition ist eine Person, die mit geschlossenen Augen und geschlossenem Mund vor einer Webcam steht. Es ist noch nichts zu hören, auf einem Bildschirm, der vor der Person steht, wird das Gesicht dieser Person gezeigt. Sobald die Person beide Augen öffnet, fängt ein Musikstück an zu spielen. Durch kurzes Schließen des rechten Auges kann die Person nun durch die einzelnen Spuren des Stückes wechseln. Es taucht dann ein Icon für die ausgewählte Spur über dem Auge auf, am Klang ändert sich erst einmal nichts. Gleiches gilt für das linke Auge mit dem Unterschied, dass nicht durch einzelne Spuren des Songs, sondern zwischen allen bereitstehenden Effekten gewechselt wird. Auch hier taucht nun ein entsprechendes Icon über dem linken Auge auf, um die Auswahl zu visualisieren, da sich am Klang ja bisher nichts geändert hat.

Durch Öffnen des Mundes kann die Person nun den ausgewählten Effekt auf die ausgewählte Spur anwenden. Bei geschlossenem Mund ist kein Effekt zu hören. Je weiter dieser geöffnet wird, desto höher ist der Effektanteil. Wenn der Nutzer mit geöffnetem Mund eine neue Spur auswählt, bleibt der Effekt auf der vorherigen Spur so erhalten.

Auch wenn ein neuer Effekt mit geöffnetem Mund ausgewählt wird, wird der vorherige Effekt beibehalten. So lassen sich mehrere Effekte auf die selbe Spur anwenden.

Da der Spaß des Nutzers auch ein wichtiges Ziel ist, wird das Öffnen des Mundes, und somit der Effektanteil durch lustige Objekte, die aus dem Mund herausfliegen, visualisiert. Die Software ist nur mit einer Person gleichzeitig nutzbar, aber auch für andere bei der Ausstellung anwesende Personen wird das Zusehen durch die Visualisierung und den verändernden Klang interessant sein.

Anforderungsanalyse

Das Programm sollte final verschiedene Anforderungen erfüllen. Zunächst muss ein Gesicht erkannt werden. Der nächste Schritt ist es, beide Augen zu erkennen. Da die Steuerung über das Schließen der Augen erfolgen soll, muss außerdem erkannt werden, ob die Augen offen oder geschlossen sind. Dabei muss zusätzlich zwischen beiden Augen differenziert werden können.

Für die Effektsteuerung mit dem Mund muss dieser erkannt werden. Außerdem soll das Programm zwischen einem geschlossenen und einem offenen Mund graduell unterscheiden können. Hierbei könnte eine anfängliche Kalibrierung für jeden Benutzer erforderlich werden, da nicht jede Person den Mund gleich weit öffnet. Außerdem muss berücksichtigt werden, wie weit das Gesicht von der Kamera entfernt ist, da auch so beeinflusst wird, wie weit offen der Mund erscheint. Als Abhilfe könnte man hier den Abstand der beiden Mundwinkel erfassen und mit dem Abstand des geöffneten Mundes ins Verhältnis setzen.

Damit das Programm einwandfrei läuft ist außerdem eine Berechnung in Echtzeit nötig. Aufgrund der Wahrnehmung des Menschen, sollte die Latenz nicht größer sein als 150ms. Eine weitere Anforderung ist es, dass die Musik gleichzeitig abgespielt wird, während das Videobild analysiert wird. Diese sollte flüssig laufen, auch wenn währenddessen Parameter ausgewählt werden.

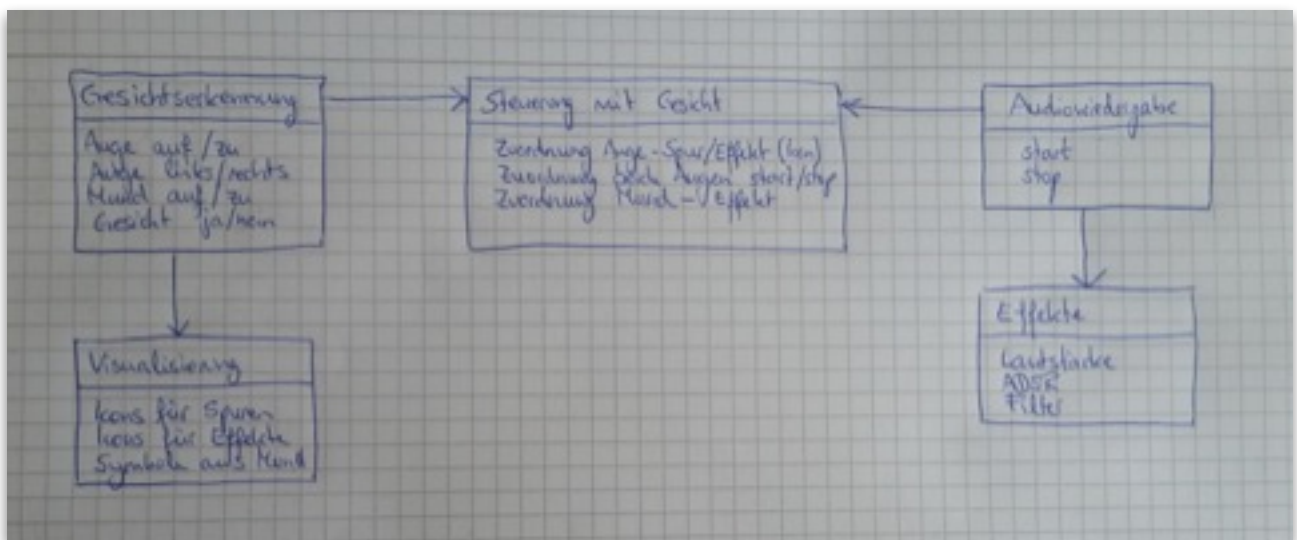
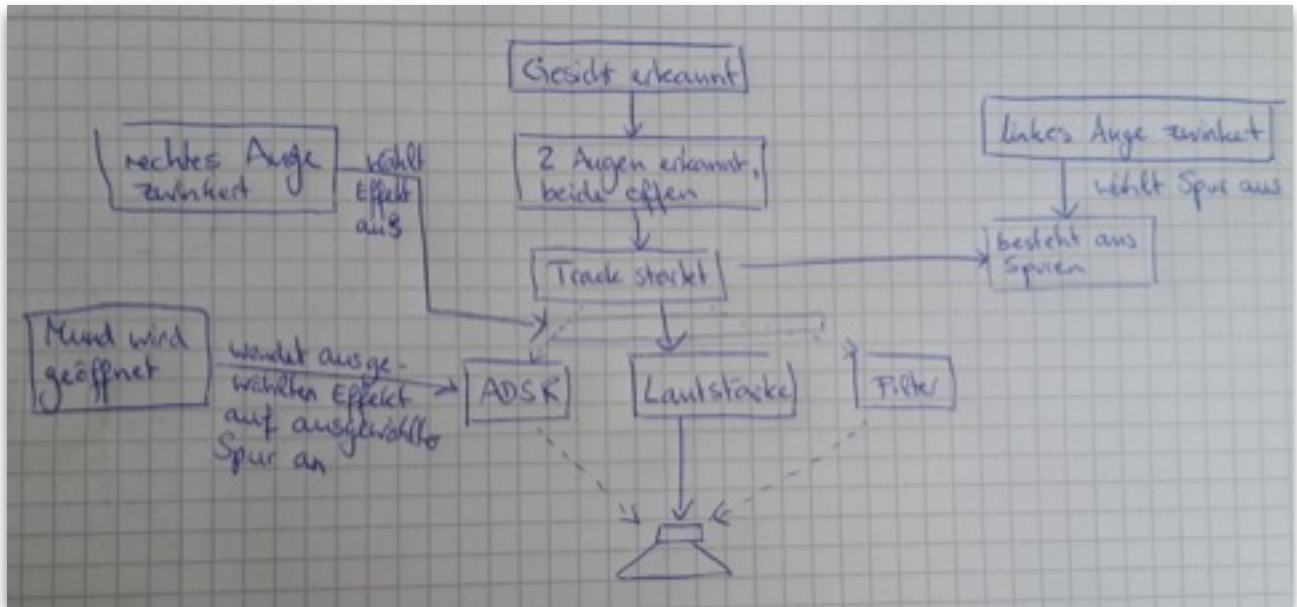
Technische Rahmenbedingungen

Das Projekt soll an einem Windows Rechner oder Laptop in C++ programmiert werden. Hierfür wird die Programmierumgebung QT genutzt, in Kombination mit openCV. Falls somit die geforderten Ziele bezüglich der Latenz bei der Videoanalyse nicht erreicht werden können, werden zusätzlich noch Haar-Kaskaden eingebunden. Somit wird eine schnellere Videoanalyse ermöglicht. Die Audiodaten werden als MIDI-Steuerdaten programmiert.

Technisches Konzept

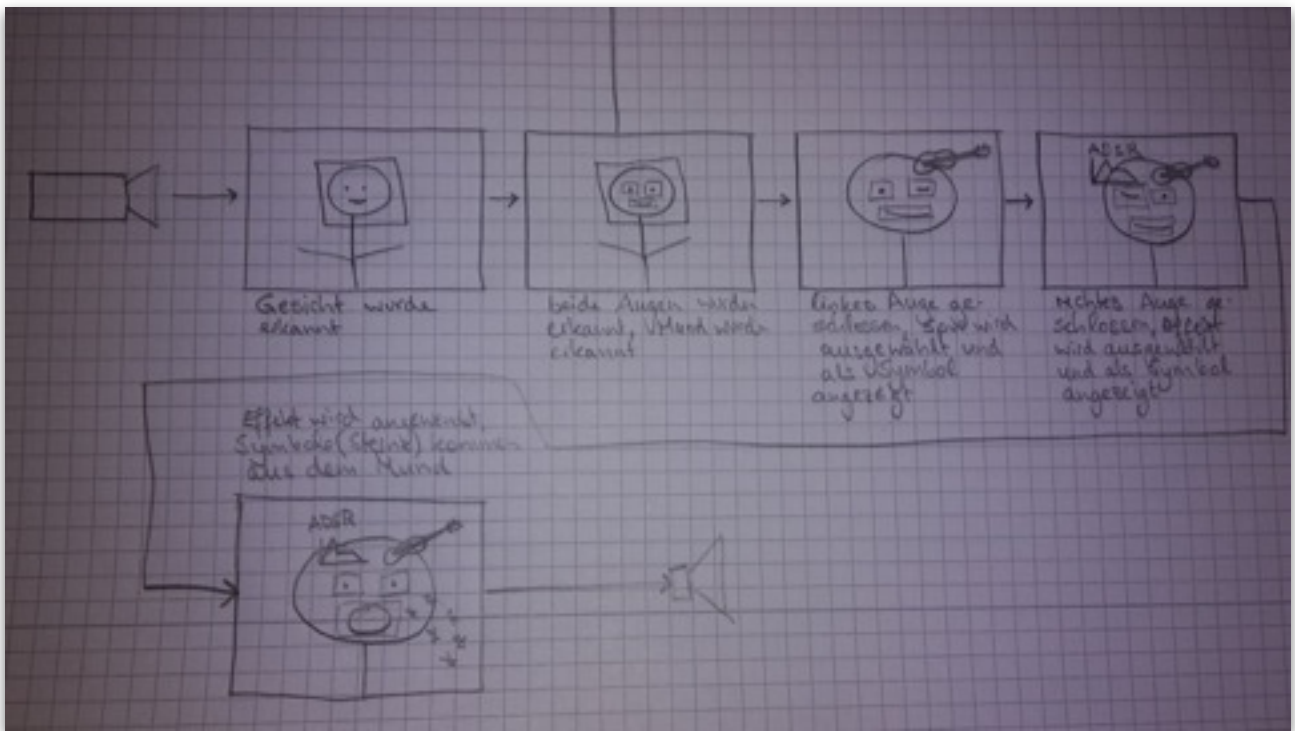
Das Projekt ist in mehrere Teile unterteilt: Gesichtserkennung, Audiowiedergabe, Audioeffekte, Kombination aus Audio und Video und die Visualisierung.

Auch bei der Programmierung werden diese Teile in verschiedenen Klassen programmiert, um das Programm möglichst übersichtlich zu halten.



Bedienkonzept

Das Programm soll möglichst komplett mit dem Gesicht gesteuert werden können. Die Programmoberfläche ist dabei nur das Videobild, das die Kamera liefert. Um sich zurechtzufinden und zu wissen, welche Parameter ausgewählt sind, werden die passenden Symbole über den Augen eingeblendet.



Eventuell wird es noch ein Menü geben, das vor dem eigentlichen Start zu sehen ist. Hier kann dann zwischen verschiedenen Musikstücken per Mausklick ausgewählt werden. Außerdem ist die Kalibrierung des Gesichts möglich, damit die Gesichtserkennung perfekt arbeiten kann.

Zeitplan

- Gesichtserkennung:
 - Gesicht erkennen und alles andere weiß (15.11.2016)
 - Augen erkennen (Unterscheidung zwischen links und rechts / auf und zu) (22.11.2016)
 - Mund erkennen
 - Mundhöhe erkennen (Abstand zwischen den Lippen) —> evt. Grauwertbild?
 - *Optional: Erkennung zwischen Kopf und Bildrand*
- Steuerung mit dem Gesicht:
 - Track durch Öffnen beider Augen starten und durch Schließen stoppen (29.11.2016)
 - Tonspuren mit dem linken Auge wechseln (6.12.2016)
 - Effekte mit dem rechten Auge wechseln (6.12.2016)
 - Stärke des ausgewählten Effekts durch Öffnen des Mundes einstellen

- Visualisierung (24.1.2017)
 - Icon für aktive Tonspur taucht über dem linken Auge auf
 - Icon für aktiven Effekt taucht über dem rechten Auge auf
 - Beim Öffnen des Mundes erscheinen je nach Intensität lustige Objekte
- Audio
 - MIDI-Track kann abgespielt werden (22.11.2016)
 - Lautstärke lässt sich steuern (29.11.2016)
- Effekte (17.1.2017)
 - ADSR Hüllkurve
 - Filter (Grenzfrequenz verschieben)
 - Hall
 - Special Effekt: Komisches Geräusch (z.B. „Kikeriki“)

Datum	Gesichtserkennung	Steuerung mit dem Gesicht	Visualisierung	Audio	Effekte
15.11.16	Gesicht erkennen und alles andere weißen				
22.11.16	Augen erkennen (links/rechts, auf/zurück)			MIDI-Track kann abgespielt werden	
29.11.16		Track durch Öffnen/ Schließen der Augen starten/ stoppen			Lautstärke lässt sich steuern
6.12.16		Tonspuren/ Effekte mit den Augen wechseln			
13.12.16	Prototyp				
20.12.16	Mund erkennen		Icons werden über Augen angezeigt		Geschwindigkeit lässt sich verändern
10.1.17	Mund öffnen erkennen		Objekte fliegen aus Mund		ADSR Hüllkurve
17.1.17					Filter
24.1.17	Ausstellung				

Teamplanung

Die Teamplanung ist für dieses Projekt sehr wichtig. Viele Dinge können parallel laufen (siehe Zeitplan), sodass es sich lohnt, die Aufgaben untereinander aufzuteilen und so effizienter arbeiten zu können.

Deshalb sieht die Gruppenaufteilung so aus: Die MS-ler werden sich um die Programmierung kümmern.

Die Unteraufteilung sieht folgendermaßen aus:

Nicolas Graf: Gesichtserkennung

Dennis Kruschel: Steuerung mit dem Gesicht

Tobias Lamping: Audio und Effekte

Da für die Visualisierung das Erstellen von kleinen Grafiken nötig ist, werden sich die MT-ler Nikolas Baumgartner und Antonia Schwab darum kümmern. Außerdem sind sie auch dafür verantwortlich, einen Überblick über das Gesamtprojekt zu behalten und den Zeitplan einzuhalten. Außerdem werden sie den Ablauf und Aufbau der Ausstellung planen, nötiges Equipment ausleihen und das Plakat erstellen.

Trotz dieser Aufteilung soll eine möglichst enge Zusammenarbeit aller Gruppenmitglieder stattfinden, regelmäßige Treffen sind geplant.