PhotoSYNTH3 #InterFACE

Zenon Olenski
Inventor & Developer
designerzen.com
Manchester / London, UK
designerzen@gmail.com

ABSTRACT

In this paper, we introduce and describe a novel approach for controlling machines using body parts without the need for touch, and present a fully working implementation in the form of a musical instrument that exemplifies how this intuitive form of input can be highly expressive yet still retain fine control, subtlety, and finesse. As a series of input mechanisms, this paper suggests new methods for human-machine interaction where *accessibility, expression and fidelity are the primary concern*s.

1. INTRODUCTION

By piggybacking onto a person's own natural movements using motion-tracking in real-time, *we can reduce the time required to learn a complex skill to mere moments whilst still retaining full expression and precise control*. This project demonstrates how to operate an interface by mapping simple, obvious, movements onto a complex control system in an intuitive and ergonomic way, and how, by observing subtle emotional cues and facial expressions, we can comprehend intention and mood using a traditional understanding of human psychology, and modify the data and output accordingly.

Expressions, movements and emotions can now control and manipulate data, software and even hardware.

This virtual interface has no moving parts, requires no specialist hardware and can run on most modern computers and mobile devices with dedicated GPUs.

Using Machine Learning we can create models that observe the human form allowing them to understand basic intentions. The face as a control surface is an instantly familiar interface - we can convert smiles into musical notes and winks into audio controls.

As a result of feedback from younger users, the demonstration software is now referred to as the, *"Smile Powered Synthesizer",* but its original strapline was, *"Instant Musical Superpowers"*.

2. REQUIREMENTS

Apart from a modern machine with a *powerful GPU;* a screen or projector, and a webcam are the only hardware requirements. A mouse, keyboard or touch screen interface is required to initiate play but only a face and mouth is necessary once the application has begun.

Web Audio Conference WAC-2022, July 6–8, 2022, Cannes, France

3. TECHNICAL IMPLEMENTATION

3.1 Capturing face position

This project takes advantage of *getUserMedia* to stream live video from the web camera onto a Canvas where algorithms work in real-time to discern landmarks on any detected faces. When a face is identifiable to the machine it will try its best to follow the contours and facial features through time - making assumptions and predictions that connect together positions over frames - providing an understanding of where the user is in relation to the machine at any point in time and what orientation their body is in.

3.2 Converting face position to music

This positional data is analyzed and facial markers are located, mapping out the entire face and its features. Triangulation allows us to calculate the shape of key facial characteristics and determine both their proportional sizes and relative distances in order to extract useful information such as mouth shape, head angle, eye positions and whether the eyelids are open or closed..

From this data, basic emotions are inferred, and winks, smiles and frowns are calculated in realtime to give us a series of useful metrics. These calculations are accelerated and stabilized by *WebGL* and as much multithreading as possible is provided via *WebWorkers*.

3.3 Generating Audio

WebAudio provides both sound synthesis and sample playback. This allows for considerable musical range by allowing previously recorded instruments to form the basis of the new sounds - an Oboe therefore sounds like an actual oboe. We also generate sounds using procedural code for the beats, synths and accompaniment; providing more range, configuration and potential expression.

The percussion is created using a range of *WebAudio Nodes*, predominantly *Oscillators*, *Buffers* filled with noise and Filters such as *BiQuad* to shape it.

All sounds are also passed through some extra dynamics to breathe life into the audio and bring unity between the different parts - *Convolvers*, *Compressors* and *ImpulseFilters* work together to provide reverb, environment and to dampen the silence.[1]

3.4 Recording Audio / Video / MIDI

Various approaches are taken to record the performance resulting in different media types : audio files, videos, pictures and MIDI performances.

[1] A common audio trick to smooth chaos and soften noises

Audio is pre-mixed through *GainNodes* then recorded in a background *Worker* using *MediaRecorder* and also passed through an *AnalyzerNode* in order to read the FFT data, resulting in both an audio recording and its associated visual waveform.

Video is recorded using captureStream from the Canvas and camera video stream. Photographs are individual Canvas frames.

Performances can be saved as well as loaded as MIDI files.

Once recorded, sonic media can be looped and redubbed.

3.5 Connecting to other Equipment

WebMIDI is used to communicate to external music equipment via the MIDI 1.0 protocol which, once a compatible device has been connected, will sync tempo and send all notes and modifiers as they are being played by the user.

If multiple people are playing, each person can control their own MIDI instrument which remains in sync with all others.

4. DEMONSTRATION

As a good representation of the flexibility and fidelity of this approach in commanding machines, an "extended-reality" accessible musical synthesizer - *completely controllable by the face* - has been made publicly available and can be accessed online at https://interface.place.

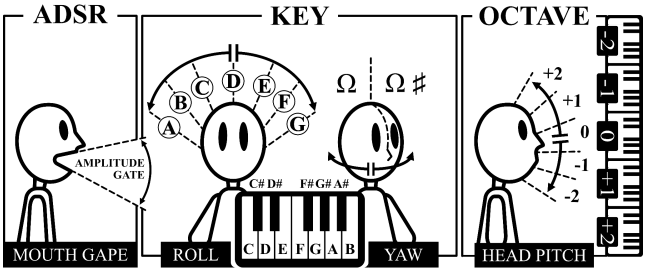
Other public demonstrations are available in the references.

5. OPERATION

This software was designed to be face controllable but as a web security requirement, one physical touch is necessary before the camera is granted permission and for this reason the player selection screen requires mandatory user interaction with a mouse or touch screen.

5.1 Controls

Once the software has been started, the face is the primary control mechanism for each user. By altering the angle of the head and the openness of the mouth and eyes, the player has full control of any digital musical instrument that communicates via MIDI as well as the internal digital synthesizer that has many built-in sounds, effects and instrument packs.



5.1.1 Mouth

The mouth acts as a classic amplitude gate where the more open the mouth is, the louder the sound. Closing the mouth stops the sound entirely. This gesture is based on a simulation of singing.

5.1.2 The Head

In order to control the key and octave of the sound, the angle of the head is analyzed, turning the head into a rudimentary joystick.

The yaw and roll of the head controls which key to play and the pitch of the head decides the octave.

Tilting and rotating the head left and right will select which key on the keyboard will be played and rocking it up and down will select which octave those keys will be played from.

5.1.3 Eyes

As a way of adding extra control to the sound once it has been played, the eyes change the stereo panning to the direction where the user is looking; so if the user looks left, the stereo pan plays more on the left channel than the right channel.

Closed eyes also offer a useful control mechanism and by closing both eyes for one bar, a person can change which instrument they are playing, for example.

5.1.4 Emotions

By observing facial landmarks it is possible to recognise certain obvious physical emotions such as happiness, anger and depression. This is used to add colour to the sound.

5.1.5 MIDI

Once connected, any MIDI instrument will stay in key and sync with the user's action.

6. ACKNOWLEDGMENTS

Thanks to the TensorFlow team at Google for their Machine Learning models, to Drake Music for their testing and feedback, the MIDI Association for their inspiration, and Hackoustic for their support. Audience of the Future, Creative Council and CreaTech for promoting the concept in the UK and SXSW / Music Maker Festival / Music Hack Day and all others who have documented the project as it has developed over the years.

7. REFERENCES

[1] Z.Olenski. 2022. *PhotoSYNTH > InterFACE in 90 Second (*MIDI Association Innovation Awards, *28 May 2022)*. Video Introduction. https://youtu.be/-DSDIET5MJo.

[2] Z.Olenski. 2021. *GIVE A MAN A RECORD, THEY DANCE FOR THE DAY, GIVE THAT MAN A SYNTHESIZER AND THEY DANCE FOR A LIFETIME (*SXSW, *18 March 2021)*. Technology Talk. SXSW2021. https://youtu.be/gTwf6ii6Lak.

[3] Z.Olenski. 2021. *InterFACE Introduction* (UKRI CreaTech, London 2021) Video Demonstration & Tutorial. https://youtu.be/2xarDjv1fx8