InterFACE: new faces for musical expression

Deepak Chandran
Center for Computer Research in Music and Acoustics
Stanford University
deepak@ccrma.stanford.edu

Ge Wang
Center for Computer Research in Music and Acoustics
Stanford University
ge@ccrma.stanford.edu

ABSTRACT

Iis an interactive system for musical creation, mediated primarily through the user's facial expressions and movements. It aims to take advantage of the expressive capabilities of the human face to create music in a way that is both expressive and whimsical. This paper introduces the three virtual instruments that make up the InterFACE system: namely, FACEdrum (a drum machine), GrannyFACE (a granular synthesis sampler), and FACEorgan (a laptop mouth organ using both face tracking and audio analysis). We present the design behind these instruments and consider what it means to be able to create music with one's face. Finally, we discuss the usability and aesthetic criteria for evaluating such a system, taking into account our initial design goals as well as the resulting experience for the performer and audience.

Author Keywords

NIME, human computer interaction, laptop instrument, facial recognition, CCRMA

CCS Concepts

•Human-centered computing → Gestural input; HCI design and evaluation methods;

1. INTRODUCTION

Humans value expression, intentionally through the music we make, or semi-intentionally in communicating with our face. Indeed, the musculature of the human face permits precise facial movements and commands a large set of expressions from a small set of facial features. The motivation for this work began as we were wondering how these features can be used as a musical interface, and how amusing its interactions might be both for the user and the audience. We considered how such a face-based interface might provide performers with primary and auxiliary channels of control over musical parameters, while at the same time, providing the audience with realtime visuals that narrate how facial expressions transform into musical elements.

In this paper, we introduce *InterFACE*, a minimal laptop application that uses facial tracking, computer vision, and the microphone in real time for musical purposes. InterFACE is made up of three virtual instruments that can be played individually or layered one on top of the other create well defined beats. The virtual instruments are: (1)*FACEdrum*, which is a drum machine that triggers samples and effects based on facial movements (2)*GRANNYface*, a sampler with simple granular synthesis features and (3)*FACEorgan*, a novel laptop instrument partly inspired by the playability of a mouth organ. InterFACE takes advantage of standard sensors found on most modern laptops (in particular, camera, keyboard, and microphone) as well as computer vision and facial tracking algorithms. The challenge of this work lies in using these technologies, and more importantly, finding suitable and satisfying mappings that bring the system and player together in an expressive interaction loop.

2. RELATED WORK

An early on inspiration for the InterFACE was the the Laptop Accordion[7] which took the approach of building backwards from the hardware. The FACEorgan is a cousin of the Laptop Accordion that follows a similar design principle of co-opting the laptop to resemble a traditional musical instrument while being creative with the existing features of hardware to create a unique but familiar interface for musical expression.

Fiebrink et. al[3], in their work, provide an argument for using the laptop itself as an interface by mentioning that while custom interfaces are incredibly powerful tools for creating music, they often require an overhead in terms of setting up time, not-so-cheap electronic components and may also present players with a rather steep learning curve. Traits like the ability to rapidly prototype and reduced overhead costs(time+money) are the need of the hour while crafting pieces for laptop orchestras.

The idea of creating music from facial gestures is not an entirely new one. Sonifier of Facial Actions[5], abbreviated SoFA, is an application created in Max/MSP runtime that divides the face into several zones and uses optical flow calculations to detect movement in those zones which are then used to trigger MIDI events. The Mouthesizer[4] was created as an interface to control audio effects using the mouth. Simply watching a guitar player use the Mouthesizer to control the frequency of a wah-wah effect is an enthralling visual in itself. The Mouthesizer makes this possible by requiring the player to wear a head mounted CCD camera.

Another interface that deserves a mention is the eyeHarp[10] which is a gaze controlled digital music instrument that allows people with severe motor disabilities to learn and compose music simply with their eyes.

Dahl et. al[1] argue that the use of metaphor is a powerful one in understanding new and unfamiliar musical interfaces, while also providing the audience with a visually engaging performance.

Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). Copyright remains with the author(s).

NIME'18, June 3-6, 2018, Blacksburg, Virginia, USA.

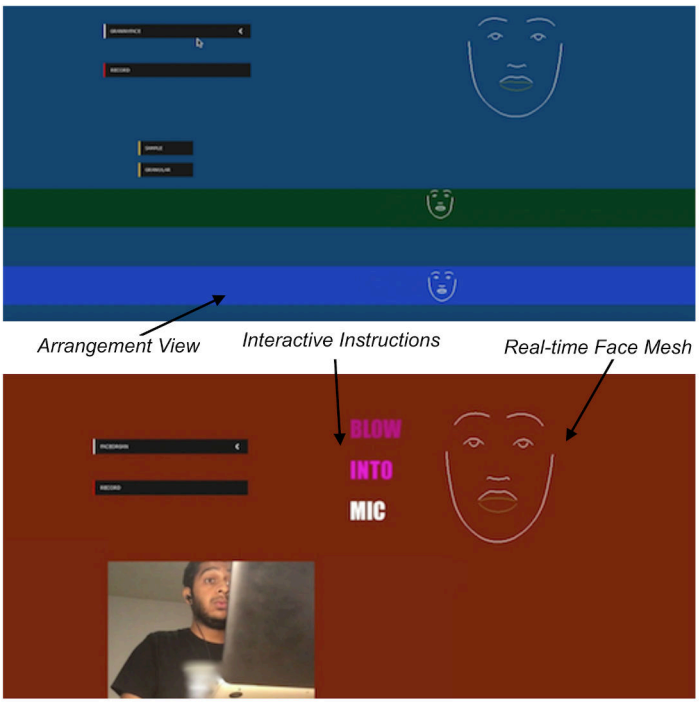

Figure 1: InterFACE in action

Siegel's[8] performance of "Two Hands (not clapping)" provides a fantastic case for performances that are mediated through computer vision. As part of this solo performance, Siegel uses hand gestures to play the role of a conductor. The webcam input is divided into several sections, each of which are mapped to different sounds.

3. DESIGN

InterFACE's design arose out of a curiosity to use one's face to make music in an expressive and amusing way. As musicians and instruments designers, we asked ourselves what are things we'd value in such an interactive system? Synthesizing our motivations for InterFACE and our sensibilities as musicians, we've come up with the following list of things we value for an interface like this. These values serve both as guiding principles as well as qualitative measures of goodness in our evaluation.

3.1 Physicality / Playability

As humans, we value the physical uses of our body, its nuances, and its range of expressions. We want to be able to make use of physical gestures, in terms of the movements of not only facial elements, but also the head, the movement of the laptop as a physical thing, as well as whistling- and, as the player's bandwidth allows, several of these in tandem. When possible, we want to take advantage of commonly used gestures in movement (e.g., nodding one's head). Furthermore, we want to craft action-to-sound mappings that not only produce meaningful sound from one's physical movements, but also encourage the player to move in intentioned and nuanced ways in order to control the sound.

3.2 Whimsicality

In his paper on the design principles of computer music controllers[10], Perry Cook proposed the principle, "Funny is often much better than serious." (Cook 2009). There is something authentic in that statement, in that there are aspects of the new interfaces we design which are often amusing and whimsical. For example, interfaces like Sonic Banana[9] and Laptop Accordion[7] afford nuanced playing, while clearly embracing whimsicality (e.g., of musically manipulating a yellow garden hose or playing a laptop sideways like an accordion). We believe these are more than incidental features, but constitute an essential part in the aesthetics of the instrument. Similarly, we believe that an interface that requires a player to shape their face not as "communication" so much as for the purpose of controlling sound holds potential in this regard, especially given how sensitive we are to changes in facial expression.

3.3 Performance Aesthetics

The overall aesthetics of the interface has a great deal of influence on how it's played as well as the way the audience experiences the performance. We recognize the value of an interface lies both in what it allows a player to do musically, as well as in how its overall aesthetics produce an aesthetic engagement with everyone involved in the performance. For something like InterFACE, part of its overall aesthetics has to do with the core elements (face, head, whistling) and modalities (gestural, audio, visual) involved. While this is often hard to characterize fully in words, the questions of "what is it like to play?" or "to experience as a performance" are important considerations for any interface.

4. InterFACE

Due to the unfamiliar nature of having to create music using one's face, we decided on a fairly minimal layout. The UI elements appear one after the other in a manner that

steers a new user to quickly familiarize themselves with all 3 virtual instruments of the InterFACE, while at the same time allowing experienced users to layer the instruments in any manner they prefer. The UI elements include:
a) a face mesh that displays an outline of the player's facial expressions in real time
b) multiple moving recorded face meshes that correspond to individual tracks similar to an arrangement view in a DAW
c) interactive mini-instructions
d) buttons to switch between instruments and start/stop recording
The combination of quirky 2-3 word instructions along with special indicators on the face mesh serve the purpose of guiding a new player to distort their faces in a manner that maps to the sound parameters for the instrument they're playing. The face-to-sound mappings that are outlined below were a design challenge in itself, and were chosen both for playability (as related to the specific metaphor or interaction) and for visual expressiveness / whimsicality.

4.1 FACEdrum

FACEdrum is a drum machine that uses specific facial movements for creating beats. It offers two modes of operation: sequencer and effects.
To create beats, the player simply bobs their head to the groove in their head, while the mapping described in Figure 2 takes care of triggering the appropriate drum samples. The effects mode currently supports Delay and Swing. The swing effect creates minor shifts in the tempo which can be used to create more 'organically' timed beats, while the delay effect can be used for stuttering drum sounds that is popular in modern electronic music.

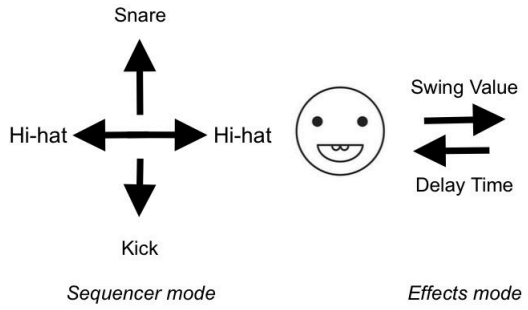

Figure 2: Parameter Mapping for FACEdrum

4.2 GrannyFACE

By picking an audio sample and making minor tweaks to certain synthesis parameters, Granular Synthesis provides an effective method for a musician to create sounds ranging from ambient soundscapes to rhythmically distorted noises. GrannyFACE is a sampler with granular synthesis features that provide the user the ability to design interesting sounds and create loops of the sounds they just created. As shown in the parameter mapping(Figure 3), each section of the player's face maps to a specific parameter of the underlying granular synthesis engine. In an effort to create a satisfying mapping in which the size of the input has a proportional effect on the output (and for maximal whimsicality), the GrannyFACE asks the player to distort their faces to the maximum to get the most interesting of sounds. For example, the mouth width is mapped to be to be inversely proportional to the grain length, i.e. as the player opens their mouth the grain length reduces exponentially and the original sample becomes less recognizable(and more distorted).

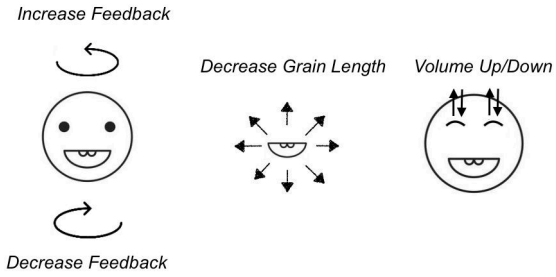

Figure 3: Parameter Mapping for GrannyFACE

4.3 FACEorgan

Among the three virtual instruments of InterFACE, the FACEorgan is perhaps the fullest embodiment of the design principles/values we've outlined earlier combining the physicality of a traditional instrument, the whimsical nature of whistling (or singing falsetto) into the computer, while controlling timbre using one's eyebrows. The FACEorgan requires the player to hold the laptop as in Figure 4. By simply whistling into the laptop microphone, a pitch-tracker follows the corresponding pitch and controls a tone generated using FM synthesis. The player can perform slight pitch bends by simply tilting the laptop up(pitch up) or down(pitch down). A more experienced musician would immediately use to this to the effect of performing a vibrato by repeatedly tilting the laptop up and down in a jerking motion. Since it is expected that the player wouldn't always hold the laptop at a constant level with respect to their face, the same note played multiple times would sound ever so slightly different each time (similar to a fretless guitar). Because we were utilizing the FM synthesis to create the tones, it made sense to give the player a finite amount of control over the timbre. As the player raises their eyebrows, the number of partials in the tone increase and a *shriller* sound is created. The charm of playing (or watching someone play) the FACEorgan is in the surprise and challenge of having to play an instrument while using physical faculties in unconventional combinations (e.g., whistling, while raising eyebrows, and physically moving the laptop).

5. IMPLEMENTATION

InterFACE is built entirely in C++ and utilizes the openFrameworks toolkit for GUI elements. The audio engine for InterFACE is implemented using the Synthesis Toolkit. In particular, the FM synthesis sounds were produced using Maximillian, an audio processing library that works especially well with openFrameworks. Pitch detection is computed using the classic YIN algorithm[2], which provides a simple yet effective method using autocorrelation for monophonic pitch estimation.
There are two types of user movements that InterFACE tries to detect. The first is pronounced movements like the player moving their heads. These movements are detected using openCV with the Haar-cascade detection algorithm[11]. We used the pre-trained classifier XMLs for face, eyes, smile, etc. that openCV provides which provided good results for our purpose. To detect subtle facial expressions like eye-

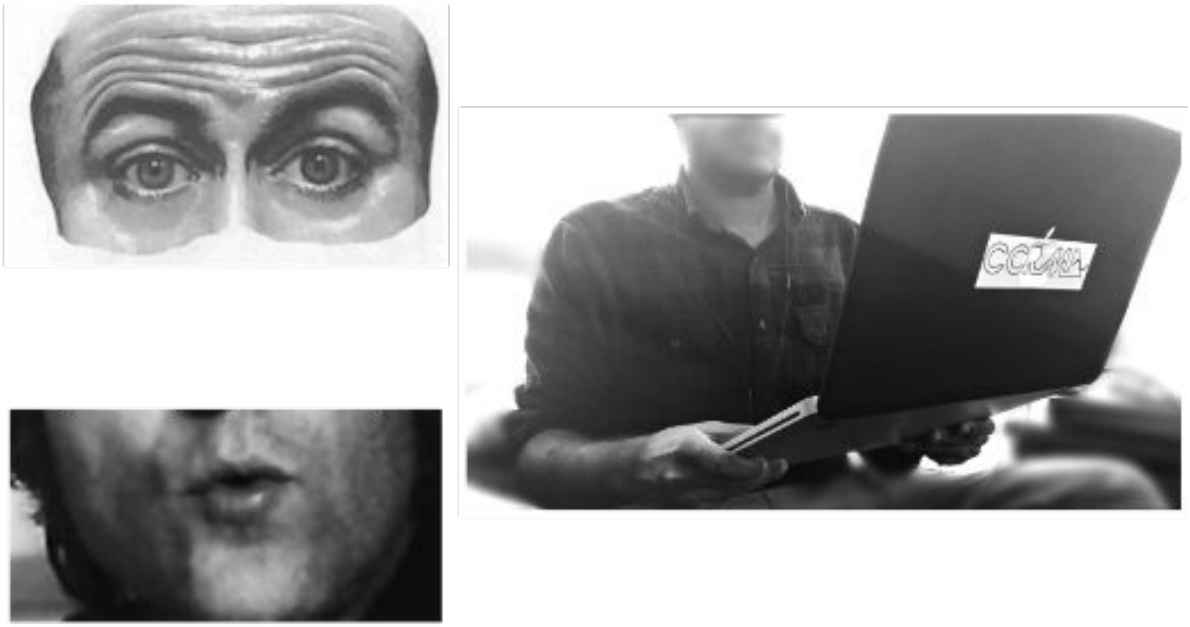

Figure 4: FACEorgan is played by whistling, moving your eyebrows, and tilting your laptop

brow movement, we used a version of the ofxFaceTracker[6] extension with few of our modifications that took care of the complexities of tracking facial expressions allowing us to focus on playability. The facial mesh is a 70 point vector that gets it shape from the values returned by ofxFaceTracker.

6. EVALUATION

In the evaluation of InterFACE, we first define how we measure success of its design as related to the core values we articulated in the design section. Here we qualitatively analyze and assess InterFACE as a system with respect to these considerations. It is worth noting that InterFACE was designed not to solve a problem, but it seemed interesting and potentially amusing in itself. That is to say, the values of whimsicality and overall playful aesthetic are things that matters in important ways, in addition to playability.

6.1 Physicality

While performing with InterFACE, it becomes quite clear that some mappings work better than others. For example, asking the performer to nod their head to create a beat in FACEdrum seemed like a fun use of the natural human movement of grooving to the beat. However, due to the latency of the face tracking algorithm, it is difficult and sometimes even frustrating to create beats that are timed accurately. This is also due to mapping a somewhat continuous gesture to a discrete sound trigger.

On the other hand, the instruments that don't require as much movement on part of the performer, like GrannyFACE and FACEorgan, tend to lend themselves to better playability. The head movements in GrannyFACE present a straightforward correspondence (if not an apt metaphor) to controlling the parameters of granular synthesis. The eyebrow movement in FACEorgan are a satisfying (auxillary) gesture to control the timbre of the generated tone, controlled by the primary input of whistling.

A few aspects of InterFACE are more nuanced. In FACEorgan, the subtle movements required to perform a vibrato are accurately detected by the underlying face tracking algorithm, giving the player fine-grain control over the rate and amount of vibrato.

From these observations, we can say InterFACE works well for mappings involving continuous movements of the face, while it doesn't perform as well for discrete mappings and lacks in its robustness from latency.

6.2 Whimsicality

Since all three virtual instruments require the performer to distort their face, often to a comical extent, the whimsical nature of InterFACE is quite apparent, especially to an audience viewing a performance. It would be fair to say that InterFACE is as much about the music being made as it is about the effort and mechanisms to make it. The aspect that quite possibly ties it all together is this: as amusing as the player facial expressions might be, it is clear (and uncanny) they are also purposeful, and have meaningful impact on the sound output. That the performer is able to make the most imaginative of sounds by putting in more effort, only makes it even more amusing.

The amusement and whimsicality factors of InterFACE are among its more successful aspects. We surmise that this is at least in part due to our involuntary sensitivity to faces, and possible to witness the cause and effect, respectively, of facial expression transformed into musical gestures.

6.3 Overall Aesthetics

In a way, InterFACE is as much about the music it makes (and is capable of making) as it is about the corresponding visuals of a player playing InterFACE. By focusing on controlling the sound with one's face, the performer simultaneously creates an audiovisual experience that appeal to both senses in the viewer. It is worthwhile to note that while the musical output of InterFACE varies greatly in quality, the overall aesthetics of the experience (as long as the player earnestly tries to the play the instrument) is nearly always interesting. And yet, there is a much greater overall sense of satisfaction when a performance is both visually and mu-

sically interesting.

7. CONCLUSION

As future work, we are considering a 'music video' feature that captures both the raw video of the player and the graphical mesh from face tracking, and mapping them onto grids of looping faces that are synchronized with the music. This may be followed by a 'share-with-friends' feature, perhaps initiating a InterFACE*off* battle between them. Additionally, we'd like to improve the system using faster and more responsive face tracking algorithms to support faster facial movements, and to improve/further simplify the graphical user interface, potentially supporting a better 'Live Performance' mode. As a more practical application, we are interested in exploring InterFACE as a system for sufferers of paralysis and other movement-impairing medical conditions to learn and make music.

In conclusion, the laptop, in and of itself, provides prospects for designing new interfaces for musical expression through artful interpretation of our physical gestures. InterFACE takes advantage of the native laptop inputs to create a real time face tracking system for music creation. We discussed the design ethos behind creating an instrument that relies on facial expressions (and other physical gestures) for controlling sound synthesis and musical processes, and presented three different instruments within InterFACE. We provided a qualitative discussion of the playability and aesthetics of the design choices in InterFACE, stemming from the design values we started with. Overall, it was an attempt to create something that values playability and physicality, while embracing the inherent playfulness of using one's face to make music.

8. LINKS

InterFACE homepage (with video demo): `http://ccrma.stanford.edu/~deepak/projects/interface/`.

9. ACKNOWLEDGEMENT

Many thanks to the students of the Music, Computing and Design course for feedback during the design phase.

10. REFERENCES

[1] L. Dahl and G. Wang. Sound bounce: Physical metaphors in designing mobile music performance. *New Interfaces for Musical Expression*, 2010.

[2] A. de Cheveigne and H. Kawahara. Yin, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America*, 2002.

[3] R. Fiebrink, P. R. Cook, and G. Wang. Don't forget the laptop: Using native input capabilities for expressive musical control. *New Interfaces for Musical Expression*, 2007.

[4] M. J. Lyons and N. Tetsutani. Facing the music: A facial action controlled musical interface. *Conference on Human Factors in Computing Systems*, 2001.

[5] K. K. Mathias Funk and M. J. Lyons. Sonification of facial actions for musical expression. *New Interfaces for Musical Expression*, 2005.

[6] K. McDonald. ofxfacetracker `github.com/kylemcdonald/ofxFaceTracker`.

[7] A. Meacham, S. Kannan, and G. Wang. The laptop accordion. *New Interfaces for Musical Expression*, 2016.

[8] W. Siegel. Two hands (not clapping) `https://vimeo.com/93591774`.

[9] E. Singer. Sonic banana: A novel bend-sensor-based midi controller. *New Interfaces for Musical Expression*, 2003.

[10] Z. Vamvakousis and R. Ramirez. Temporal control in the eyeharp gaze-controlled musical interface. *New Interfaces for Musical Expression*, 2012.

[11] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001.