Facing the Music
A Facial Action Controlled Musical Interface

Michael J. Lyons and Nobuji Tetsutani
ATR Media Integration and Communication Research Laboratories
2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, Japan
+81-774-95-1433, {mlyons,tetutani}@mic.atr.co.jp

ABSTRACT
We describe a novel musical controller which acquires live video input from the user's face, extracts facial feature parameters using a computer vision algorithm, and converts these to expressive musical effects. The controller allows the user to modify synthesized or audio-filtered musical sound in real time by moving the face.

Keywords
Musical Controllers, Facial Action Recognition, Affective Computing.

INTRODUCTION
Music is a universal human communication system by which people transmit feelings and abstract thoughts through a medium which anyone, anywhere, can understand. Musical instruments may be thought of as interfaces for converting musical intention into sound patterns via muscular action. The rapid evolution of technology is opening up new opportunities for creating musical interfaces that make increasingly greater use of the human capacity for the fine control of temporal sequences of muscular action [1].

MUSICAL CONTROLLERS
The development of *controllers* to modulate musical tone, has been important in the evolution of acoustic and electroacoustic instruments. Addition of the sustain peddle to the piano, for example, greatly enhanced it as an instrument for expressing mood and emotion. Controllers may be actuated by any part of the body. The degree of muscular control afforded to a body part is related to the amount of neural tissue devoted to that part, which can be gauged by the size of corresponding areas in somatosensory and motor cortices. The somatosensory and motor homunculi [2] depict the piecewise somatotopic mapping of the body in the cerebral cortex. Figure 1, an artist's rendition of the somatosensory homunculus, makes it clear that the hands and lower face account for a disproportionate share of cortical real estate. This is reflected in the important role played by the hands in motor tasks and the lower face in speech and facial expression The face is the single most important source of social information communicated between humans. The muscles of the mouth and face are involved in both verbal and non-verbal communication. Speech, singing, and facial expression all require the fine spatial and temporal control of a large number of muscles of the face and mouth. Hence it seems natural to make use of muscular action of the face for human computer interaction. Here, we report the development of a computer vision based system for converting facial actions to musical control signals

FACIAL ACTION DRIVEN MUSICAL INTERFACE
Figure 2 shows a general schematic of the interface.. A frame grabber digitizes the signal from a camera pointed at the face. Initially, two types of system were tested: one which tracks the face from a stationary, remote camera; and one in which a miniature camera is mounted in front of the face on a lightweight headset. The former configuration proved unsatisfactory: musical performance makes demands on tracking stability beyond the current state of the art. Moreover, musicians who tried the system were more comfortable with a "wearable" device which is proximate, under their control, and removable at any time.

Figure 1 Somatosensory homunculus representing relative sensitivity of various parts of the body surface. Adapted from [2] with permission.

Next, facial features are located in the image and shape parameters are extracted. The first implementation treats only the mouth region as it is the most salient feature for speech and expression. Color and intensity thresholding is used to extract the area of the mouth opening, which is not a reflective surface and therefore relatively insensitive to changes in lighting and view. Parameters proportional to mouth width and height are calculated from the second order statistics of the segmented region, and mapped to two MIDI control changes. The system runs at approximately 15 fps on a SGI O2 computer; latency is noticeable but tolerable. It is critical to chose a good mapping of shape to sound effect: arbitrary mappings are possible but are neither intuitive nor pleasurable to use. Below we give an example of an intuitive and easy-to-use mapping.

EXAMPLE: MOUTH CONTROLLED GUITAR EFFECTS

Guitar effects refers to analog or digital processing of electric guitar signals to modify the tone. The rock guitarist Jimi Hendrix was an effects virtuoso: his emotionally expressive use of *wah-wah* and *distortion* was partly responsible for his distinctive sound. Our sample implementation maps mouth height to the cut-off frequency of a sweeping resonant low-pass audio filter. Filtering the guitar signal with the sweeping resonant filter mimics the effect of opening and closing one's mouth while voicing the sound "ah": hence the onomatopoeic term *wah-wah* for this effect. Mouth width is mapped to the *distortion* level of an amplifier. Opening the mouth increases the non-linearity of an audio amplifier which clips the guitar signal waveform. Stretching the corners of the mouth apart in an emotional grimace produces a rough, dirty tone which is valued in rock and blues music. The audio effects are implemented using a Nord Virtual Modular synthesizer (purchased from Clavia DMI, Sweden). Audio effects are edited in software then loaded into a DSP module for real-time signal processing. Effects parameters are modified in real-time by MIDI control changes sent to the DSP system from the visual subsystem. In addition to the example described here, we have used the mouth controller to modulate techno loops and keyboard triggered synthesizer patches. A short video clip of a guitarist using the interface is available at: http://www.mic.atr.co.jp/~mlyons/mouthesizer.html.

CONCLUDING REMARKS

We have demonstrated the first system that uses computer vision to convert facial action to musical effects. The use of facial musculature in speech and emotional expression makes musical control a natural application for this interface, however it might also be used as a general machine interface by people with severe spinal damage, whose cranial nerves, and hence facial control, are intact. Work is underway to use other facial features such as the eyes and eyebrows and to explore the rich space of possible facial action to musical parameter mappings.

ACKNOWLEDGMENTS

We thank Rodney Berry, Palle Dahlstadt, Sidney Fels and Ivan Poupyrev for helpful discussions and Kazue Shinozawa for drawing figure 2.

REFERENCES

1. Wanderley, M., and Baffier, M. (eds.). *Trends in Gestural Control of Music* CD-ROM, IRCAM, Paris, 2000.

2. Gillespie, B., Haptics., in *Music, Cognition, and Computerized Sound*, P. Cook (Ed.). MIT Press, Cambridge,MA,1999.

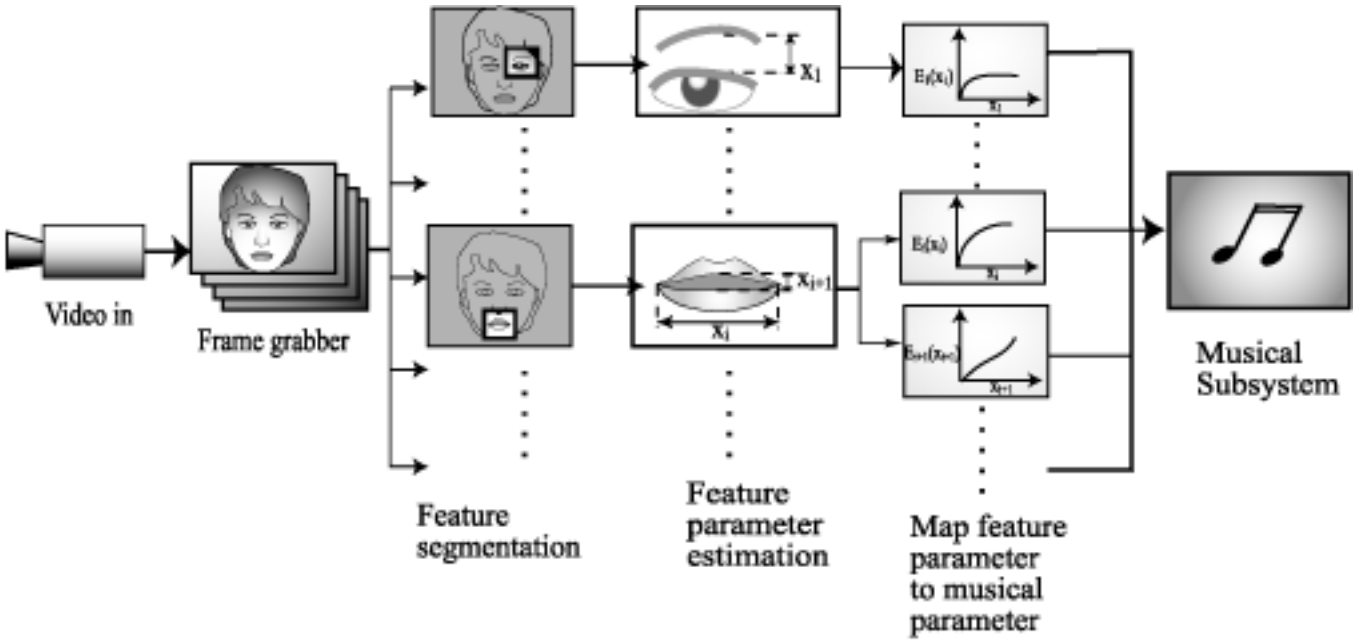

Figure 2 Schematic of the facial action driven musical controller.