

INTRODUCTION

In this technical sketch we introduce a new interaction paradigm: using facial gestures and expressions to control musical sound. The mouthesizer, a prototype of this general concept, consists of a miniature head mounted camera which acquires video input from the region of the mouth, extracts the mouth shape using a simple computer vision algorithm, and converts shape parameters to MIDI commands to control a synthesizer or musical effects device.

FACE AND GESTURE MUSICAL INTERFACES

Leading the edge of musical technology innovation are the efforts to interface expressive gestures to sound synthesis and modulation.^{1,2} The face is a salient source of non-verbal communication conveying information about attention, intention, and emotion through facial expressions and gestures. The lower face is also critical to speech production through the action of the lips, tongue, and jaw. It therefore seems natural to us to link expressive action of the face to the control of musical sound.

TECHNOLOGY OF THE MOUTHESIZER

Our system consists of a wearable, lightweight, head mounted camera (Figure 1). The open area of the mouth is segmented by intensity and color thresholding and the largest blob is selected. The mouth opening is not a surface so the segmentation algorithm is robust over a wide range of illumination conditions. Statistical shape analysis is used to extract parameters proportional to the width and height of the mouth opening, which are normalized and output in MIDI format to a synthesizer-effects unit. The system runs at 30 fps on a notebook computer. Experimentation has revealed several intuitive shapesound mappings. The video demo at the site below implemented mouth-controlled guitar effects of wahwah and distortion (Figure 2). Performers report that the mouth controller is more intuitive and versatile than a foot pedal and that it feels good to use. Another example uses mouth aspect ratio to audio morph between three formant filters corresponding to vowel sounds [i], [a], and [o]. These audio effects can be applied to synthesizer patches or analog signals. Future work will consider other regions of the face and further explore the rich space of gesture-sound mappings. The research is targeted both at musical performers, who have their hands busy playing an instrument, and the handicapped, who may lack control of their limbs due to spinal damage, but who retain control of their facial muscles.

A short video is available at:

<http://www.mic.atr.co.jp/~mlyons/mouthesizer.html>

References

1. Wanderley, M., & Baffier, M., eds. (2000). Trends in gestural control of music, CD-ROM. IRCAM, Paris, 2000.
2. Cutler, M., Robair, G., & Bean. (2000). The outer limits: a survey of unconventional musical input devices. *Electronic Musician*, August 2000: 50 - 72.



Figure 1. The head mounted video camera with views of the vowels [i], [a], and [o]. The segmented area is shown in red.



Figure 2. Guitarist Ichiro Umata using the Mouthesizer.