Music Moves

Wim Looman, Richard Green
University of Canterbury

***Abstract*—This report**

I. Introduction

A lot of different methods for producing movement from sound have been researched. Some of these involve using the hands to "shape" the music, for example having a 2d pointer (which could easily be generated from a hand tracking algorithm) that affects the music based on a set of simple gestures [1]. Others attempt to provide a more traditional musical interface, albeit without the actual instrument. This can be a simple interface such as a DJ's deck [2], or a more complex interface like a full piano [3] or guitar [4].

Whilst the first approach will probably lead to the most valuable research, it would also require a deep understanding of music theory. Therefore the second approach was the one chosen for investigation. One instrument that hasn't been widely studied for video based playing is the air drum. It would have to be wondered why; as the drum simultaneously provides both a complex interface, the 3d position of the imaginary sticks, and a simple interface, only around 3-10 different drums/cymbals to hit.

Similar to previous research in this area [4], hand tracking was decided on as the simplest method of decoding the input video. This would then produce two sets of (x, y) coordinates per frame, one for each hand. These would be fed into a neural network, along with the positions for the previous set of frames. This would allow the neural network to detect the changes in hand position corresponding to the hitting of the imaginary drum set.

A. Hand Tracking

There are a few possibilities for the hand tracking, with or without marks. With marks we could have the user wearing a pair of brightly coloured gloves, or holding a pair of brightly coloured drumsticks. Without props we could attempt an implementation of the Kalman filter to track the front/tops of the users fisted hands.

To make this simpler the mark based tracking was chosen, specifically a pair of brightly coloured gloves. This will limit the subjects clothing in that it cannot contain the same colour as the gloves. However it greatly simplifies the hand tracking to just following the specific colour of the gloves.

The tracking will involve two stages, target acquisition & learning and the actual tracking. During acquisition and learning the user will start with their hands behind their back, then hold them in a few positions in front of the camera. This will produce a series of images which will be automatically processed to attempt to compensate for the lighting conditions. Once an accurate enough colour sample has been achieved the actual tracking will be started.

B. Neural Network

To simplify the creation of the neural network it was decided that this should be a separate program. Writing the network in Erlang would be much easier than using C, and by having a separate program the training would be simpler to setup.

The neural network would have a large input layer, four nodes per frame of the video. To provide context it is assumed that at least a second of video will be required, at say 30 frames a second this means the input layer would be at least 120 nodes. The actual length of the context will be part of the optimisation of the neural network.

C. Sound Output

The sound output will once again be a seperate program. This will take in the output of the neural network as a stream of drum/cymbal hits, it will then play back a pre-recorded clip for the specific drum/cymbal.

II. Results

III. Conclusion

References

[1] K. Jensen, "Aspects of the multiple musical gestures," *Journal on data semantics*, vol. 3902, pp. 140–148, Jan. 2006.
[2] B. Yetton and R. Green, "Virtualmix - virtual dj interface using computer vision," 2010, produced for COSC428.
[3] R. I. Godøy, E. Haga, and A. R. Jensenius, "Playing 'air instruments': Mimicry of sound-producing gestures by novices and experts," *Journal on data semantics*, vol. 3881, pp. 256–267, Jan. 2006.
[4] Y. Pan and R. Green, "A mark-less air guitar game through computer vision," 2010, produced for COSC428.