# RECOGNITION AND RESPONSE TO HUMAN MOVEMENT CONTAINED IN MOTION CAPTURE DATA USING THE SELF ORGANISING MAP

John McCormick

Neural Networks have been used successfully for recognition of human gestures in many applications including analysis of motion capture data. This paper investigates the potential for using the same methods for both recognition and synthesising responses in relation to movement contained in motion capture sequences.
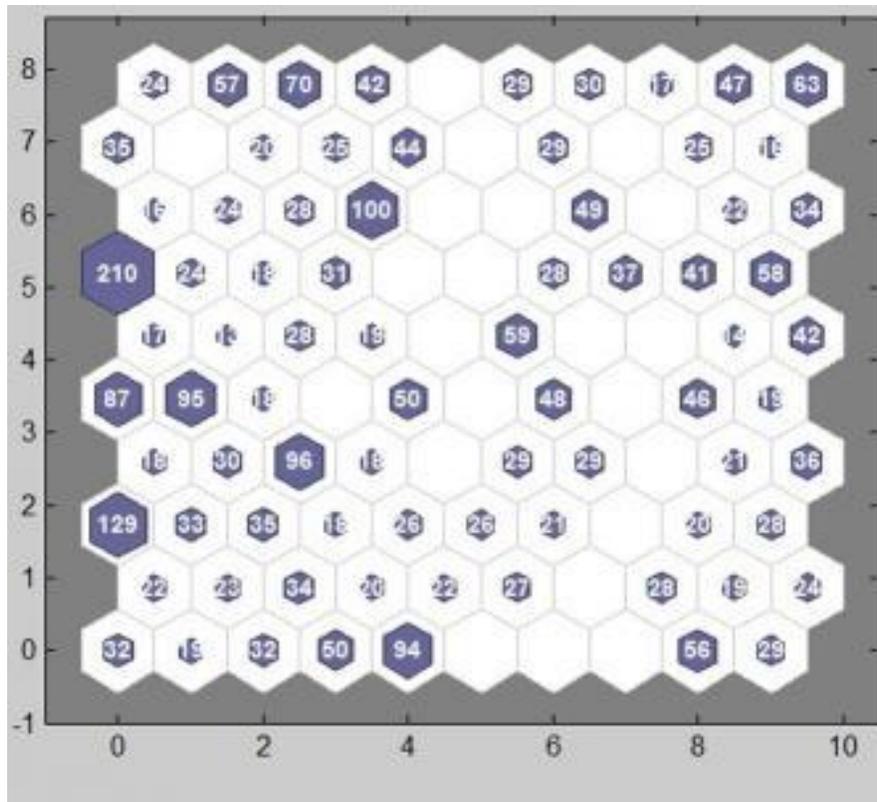
Representation of the trained Self Organising Map with neurons containing clusters of similar patterns derived from the motion capture data.

This research arose from questions regarding the nature of collaboration and the use of immersive digital sound and visual environments as a component of live dance performance incorporating real-time motion capture data. Human collaborators are able to make use of their experiences and memories to respond to developmental concepts and synthesise possibilities in relation to a new artwork. If the software environments or agents were to be considered part of the collaborative process, what traits would be beneficial to them?   Some form of memory would be useful, so the agent would have references to apply to incoming stimulus, or some substance with which to synthesise possibilities.

In contemplating the software agent as collaborator, even in a very limited sense, this research considered models of software that attempted to mimic human brain functions. Artificial Neural Networks (ANN) attempt to model the behaviour and capabilities of neural networks in the brain, and have been popular in the area of machine learning, including the field of gesture recognition. While there have been many successes in the area of gesture recognition using artificial neural networks, [1] [2] in human communication that recognition may often be a precursor to a response and a solution was sought that offered possibilities of both recognition and response especially in relation to human movement contained within motion capture data. There are many types of ANN however this research currently employs a particular type of ANN, the Self-Organising Map (SOM) or Kohonen Feature Map, after Teuvo Kohonen who first described them. [3] [4]

The Self Organising Map is an unsupervised form of neural network in that there is no ideal output suggested to the network, only the input data is provided. Furthermore the input data is not necessarily labelled in any manner so it is up to the SOM to find any patterns within the data and to group these into classes.

Sequences of movement were captured using both Motion Analysis and Optitrack optical motion capture systems to determine if the method was system independent. Both systems used multiple cameras to record the positions of reflective markers attached to the body of the dancer. The data was produced to represent both limb position as defined by marker positions and a hierarchy of joint rotations. This allowed testing of the SOM with the most popular representations of motion data, i.e. position data or joint rotation data. The SOM chosen was represented as a 10 X 10 array of neurons giving 100 neurons competing to classify the samples of motion data. The motion sequences were around 3000 frames long and each frame was treated as a sample by the network. Within each frame there were either 34 marker positions (the number of markers on the dancer's body) or 19 joint rotations, and each of these represented by a vector (x,y,z), so a total of 102 position values or 57 joint rotation values. Each frame containing all of these values was presented to every neuron and the one deemed to have the closest match (Best Matching Unit, BMU) is the winner and the map is adjusted accordingly. A weighting for the winning neuron and a decreasing number of neighbouring neurons are adjusted and over many iterations a weighting map is formed that increasingly matches the topography of the input data. The final map can be visualised in a number of ways, but perhaps the most pertinent to this paper is in the form describing the number of clusters or hits each neuron achieved.

Figure 1 shows the number of frames each neuron gathered as a class or cluster with similar data patterns. The patterns here, being frames of mocap data, could be considered dynamic postures extracted from the movement sequence. To test the trained network, mocap data representing a limited number of movements of known composition and length was introduced and the resultant neuron map compared to the map of the trained network. For example the main sequence contained a few hundred frames of the dancer in T-Pose (standing with feet together and arms out to form a T shape) at the start of the sequence. A short sequence of mocap data containing only T-Pose data when presented to the trained network resulted in all the frames stimulating the same neuron that contained the T-Pose samples from the original sequence. This pattern was seen when presenting other short, known movement postures to the trained network. The classification or recognition of the movement data was seen in both position and rotation datasets, though the resultant maps were different in the distribution and number of hits each neuron accumulated.

The SOM proved to be a robust method for classifying motion captured movement. It was able to create a map of movement frames which could be used to classify or recognise further incoming motion data.

More importantly for this research, it was able to create a map that could be treated as a type of memory of the dance as represented by the motion data. Traversing the map in different ways could lead to responses that are inherently related to the memory of the performed movement, but with the potential to create variations on the movement as responses to incoming motions. This is possibly analogous to the process displayed by human performers when improvising or developing movement and it is this re-synthesis or traversing of memory in order to produce movement responses that is the current stage of this research.

The results have pointed to a number of further possibilities relating to live performance. Multiple maps representing different components of the performance; movement, sound, images, could be trained and then traversed simultaneously during the performance. The use of multiple maps may be analogous to the processing of specialised information by different parts of the brain and could be used with some higher function logic to co-ordinate the synthesis of the multiple elements.

References and Notes:

1. G. Caridakis et al., "SOMM: Self Organizing Markov Map for Gesture Recognition," in *Pattern Recognition Letters* 31, no. 1 (2010): 52-59.
2. A. Shimada and R. Taniguchi, "Gesture Recognition Using Sparse Code of Hierarchical SOM," *Proceedings of 19th International Conference on Pattern Recognition (ICPR 2008)*, Tampa, FL (2008).
3. T. Kohonen, *Self-Organization and Associative Memory* (New York: Springer-Verlag, 1989).
4. T. Kohonen, *Self-Organizing Maps* (New York: Springer, 1997).