Responding to human full-body gestures embedded in motion data streams through machine learning and memory

by

John McCormick

B.A. (Arts)
Grad. Dip. (Media Studies)
M.A. (Animation, Interactive Media)

Submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

Deakin University
July 2014

(Thesis submitted for examination – August 2014)
DEAKIN UNIVERSITY

ACCESS TO THESIS - A

I am the author of the thesis entitled

Responding to human full-body gestures embedded in motion data streams through machine learning and memory

submitted for the degree of Doctor of Philosophy

This thesis may be made available for consultation, loan and limited copying in accordance with the Copyright Act 1968.

'I certify that I am the student named below and that the information provided in the form is correct'

Full Name: John McCormick

Signed: Signature Redacted by Library

Date: 20 August 2014
DEAKIN UNIVERSITY

CANDIDATE DECLARATION

I certify the following about the thesis entitled
*Responding to human full-body gestures embedded in motion data streams through machine learning and memory*

submitted for the degree of Doctor of Philosophy

a. I am the creator of all or part of the whole work(s) (including content and layout) and that where reference is made to the work of others, due acknowledgment is given.

b. The work(s) are not in any way a violation or infringement of any copyright, trademark, patent, or other rights whatsoever of any person.

c. That if the work(s) have been commissioned, sponsored or supported by any organisation, I have fulfilled all of the obligations required by such contract or agreement.

I also certify that any material in the thesis which has been accepted for a degree or diploma by any university or institution is identified in the text.
I certify that I am the student named below and that the information provided in the form is correct

John McCormick

Full Name...........................................................................................................

Signed ..................................................................................................................

Date.................. 20 August 2014.................................................................

Signature Redacted by Library
Acknowledgments

I would like to acknowledge the unwavering support of my principal supervisor Professor Kim Vincs, and co-supervisors, Professor Saeid Nahavandi and Associate Professor Douglas Creighton. They have been incredibly generous with time, suggestions and trust. For all the support from everyone at Motion.lab and Centre for Intelligent Systems Research I say thank you. Thanks to all the artists and animateurs I have had the pleasure to collaborate with, too many to mention, for the influences they have exerted on my work. To Adam Nash a great friend and inspiring artist. To my enormous family, of course, and to my partner Sothy who, as partners do, keep one sustained to completion.

I would like to extend my gratitude to Steph Hutchison, “the dancer” in this thesis. Untiring, generous and intelligent, the process was made much easier with such a professional accomplice. And I would like to thank the agent for emerging when I needed it. There was no guarantee that it would respond to my invitation, but I’m very glad it did.

John McCormick
July 22 2014
Melbourne Australia
# Table Of Contents

- Acknowledgments ................................................................................................................................. i
- Table Of Contents ................................................................................................................................. ii
- Table Of Figures ..................................................................................................................................... v
- Table Of Equations ............................................................................................................................... xvii
- Guide to Reading ................................................................................................................................... xviii
- Video Documentation ............................................................................................................................ xviii
- Publications ............................................................................................................................................ xix
- Abstract ................................................................................................................................................ xx
- Introduction .......................................................................................................................................... 1
- Background .......................................................................................................................................... 10
  - Developing an Agent Performer ........................................................................................................... 14
  - Memory .............................................................................................................................................. 15
  - Embodied Experience through sensory apparatus ............................................................................. 17
  - Relationship between the agent and its environment ....................................................................... 19
- Scaffolding to support the agent dancer relationship ............................................................................. 22
  - Artificial Neural Networks ................................................................................................................. 27
  - Artificial Neural Networks for detecting behaviour ........................................................................... 28
  - Movement Generation ....................................................................................................................... 31
  - Agents in Dance Performance .......................................................................................................... 35
# Table of Contents

Experiments ........................................................................................................................................................................................................................................... 39  
Self-Organising Maps ........................................................................................................................................................................................................................ 39  
SOM Algorithm .................................................................................................................................................................................................................................. 41  
Experiment 1: Creating associative memory of movement performance using Self-Organising Maps .......................................................................................................................... 44  
Unity Game Engine ........................................................................................................................................................................................................................... 49  
Experiment 2: “Hello World” Colour clustering with SOM .................................................................................................................................................................. 50  
Experiment 3: Interactions with associative memory of movement .................................................................................................................................................. 51  
Standardised Data Structure ............................................................................................................................................................................................................. 56  
Recording and Pre-processing the data for SOM training ................................................................................................................................................................. 57  
Creative Process .................................................................................................................................................................................................................................... 60  
Studies ........................................................................................................................................................................................................................................... 62  
Agent following the dancer from memory ................................................................................................................................................................. 62  
The agent learns to generate independent movement ............................................................................................................................................ 64  
Self-Organising Synaptic Map ........................................................................................................................................................................................................... 65  
Agent and Dancer moving independently ................................................................................................................................................................. 68  
Agent generating movement from movement seeds supplied by the dancer .................................................................................................................................... 68  
Combining a Self-Organising Synaptic Map (SOSM) and Hidden Markov Model (HMM) ................................................................................................................. 70  
Using SOSM and HMM for full-body gesture recognition .................................................................................................................................................. 72  
Agent recognizing short dance “gestures” ................................................................................................................................................................. 74  
Performance .......................................................................................................................................................................................................................................... 78  
Grevillea Crystalis Incarnadine .......................................................................................................................................................................................................... 79  
Recognition ........................................................................................................................................................................................................................................ 82
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrumental</td>
<td>86</td>
</tr>
<tr>
<td>Verbose Mode</td>
<td>88</td>
</tr>
<tr>
<td>Discussion</td>
<td>92</td>
</tr>
<tr>
<td>Spotting gestures in continuous movement</td>
<td>97</td>
</tr>
<tr>
<td>Movement generation variability</td>
<td>98</td>
</tr>
<tr>
<td>Learning paradigm for robot movement</td>
<td>98</td>
</tr>
<tr>
<td>The agent interaction as a source for choreographic invention</td>
<td>99</td>
</tr>
<tr>
<td>References</td>
<td>100</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>1</td>
<td>Lines from Sentient Space 1.0. 2004. Company In Space. Dancers Hellen Sky and Ruth Gibson performed using an Animazoo Gypsy and Polhemus Ultratrack motion capture systems. The coloured volumes connected between the data points of their virtual skeletons. The volumes of colour grew and shrank with the velocity of the dancer's movements. The environment was projected in Stereoscopic 3D using passive stereo filters and glasses. Performed at the Virtual Reality Theatre, Centre for Astrophysics and Supercomputing, Swinburne University. Images: John McCormick.</td>
</tr>
<tr>
<td>2</td>
<td>Lines from Sentient Space 1.0. 2004. Company In Space. Lines used a fairly direct mapping of the dancer's movement to the projected volume to visualise the movement. The dancer’s combined data from the two motion capture systems was combined to affect the volumetric display. The volumetric animation had a certain amount of elasticity in its movement, as it was not an exact reflection of the dancer’s combined movements. Combining the dancer’s movement data extended the notion that the volume reflected the connection between the dancer’s virtual skeleton representations.</td>
</tr>
<tr>
<td>3</td>
<td>Paper from Sentient Space 1.0. 2004. Company In Space. The sheets of transparent paper responded to the velocity of the dancer’s movements. Each alternating sheet was influenced by one or the other dancer. The sheets were flung out with large movements and gently oscillated with less vigorous movements. They also had their own elasticity, however they returned to their stacked starting points in the absence of any movement. The two skeletons can be seen in blue and red, however they weren’t visible in the performance.</td>
</tr>
<tr>
<td>4</td>
<td>Paper from Sentient Space 1.0. 2004. Company In Space. The scene was projected in Stereoscopic 3D to give a greater sense of depth to the objects and to attempt to envelop the dancers in the projected environment.</td>
</tr>
<tr>
<td>5</td>
<td>2 Avatars. 2002. Monaco Dance Forum. 2 Avatars was developed as part of the Motion Capture Tech laboratory “Real Time and Networked: Sharing the Body” facilitated by Scott deLahunta. Dancers Nik Haffner and Thomas McManus each wore Gypsy exoskeleton motion capture systems. One of the dancers was able to control the flight of the other dancer and in doing so the second dancer’s avatar would stretch and flow in response. Thus they both had their own intrinsic movement, however there was also an external relationship governing their duet that was a product of the transfer of movement data from one body to another.</td>
</tr>
</tbody>
</table>
Figure 6. 2 Avatars. 2002. Monaco Dance Forum. .................................................................................................................................................................................. 6

Figure 7. 2 Avatars. 2002. Monaco Dance Forum. Dancers Nik Haffner and Thomas McManus in Gypsy exoskeletons. ........................................................................................................ 6

Figure 8. Mesh from Sentient Space 1.0. 2004. Company In Space. The view of the avatar was from the perspective of the second dancer. A virtual camera attached to the second dancer’s head joint allowed her viewpoint to be displayed. During the performance she was focused on the other dancer and their relative relationship and proximity were visible on-screen. While this relationship was very simple, it illustrated a theme found in much of the work of this period, the use of data to explore relationships between participants in a digital environment. ...................................................................................................... 7

Figure 9. Nova. 2004 Company In Space. A simple particle system visualisation of the dancer’s movement. It was produced using a Gypsy exoskeleton motion capture system. ............................................................................................................................................................................................................................... 8

Figure 10. Escape Velocity. 1998. Escape Velocity was performed mainly as a telematic performance over ISDN videoconferencing. It was performed as a distributed performance on various occasions between Melbourne and cities in Australia and internationally including from Siggraph Orlando Fl. to the portico of the Melbourne Town Hall. The work was choreographed by Hellen Sky and performed by Hellen Sky and Louise Taube. Escape Velocity was one of the earliest CIS works investigating the mediation of personal relationships through technology and the changes in view of self and others afforded by this transmission of personal effect. It also demonstrated an ongoing pre-occupation with the enmeshing of one set of information with another to create a new artistic statement. .......................................................................................................................................................................................................................... 10

Figure 11. Escape Velocity. 1998. Company In Space. Escape Velocity was also performed as a single site work. Here the dancers Sky and Taube move through a field of lasers that when interrupted, emitted sound, adding to the soundscape. .......................................................................................................................... 11

Figure 12. Trial By Video. 1998. Company In Space. Trial By Video looked at the immediacy and availability of data available through digital networks and the effects this can have on an individual. The range of new “pressures” impacting on our lives is increasing, mostly benign, many positive, some not so. The increased scope for various forms of surveillance and personal intrusion afforded by emerging networks became one of the main themes of the work. Here Taube is contained within a cage of light. As her movements cut through the beams, short sentences, her inner thoughts, are emitted as a multi-layered soundscore. The cage becomes an instrument of expression for Taube as her normal voice is silenced. ........................................................................................................................................................................ 12
Figure 13. CO3. 2001. Company In Space. CO3 investigated the nature of the post-human condition, the emerging cyborg, as we get more enmeshed with the digital objects and artifacts that co-create our existence. Performed by Hellen Sky in a Gypsy exoskeleton motion capture system. In the first performance at the Astor Theatre in Melbourne, Sky was joined by Keith Roberson in Florida, who intervened in her actions using a Polhemus Fastrack tracking system. The interactions meant her form wasn’t always intact, often becoming a hybrid with the environment she inhabited. CO3 was performed in various configurations and contexts, always site-specific, and adaptable to the physical environment it inhabited. At the ICA in London, room-length metal walkways criss-crossed the space, around and under which the audience wandered to view the performance.

Figure 14. Ways To Wave. 2008. Squaretangle. Ways To Wave, by Adam Nash and John McCormick used a physical sculpture of a lotus to interface into a 3D audio-visual environment in Second Life. It was presented as part of 01SJ in San José, California. By moving the petals of the lotus and changing its shape, gallery patrons could “play” the online soundscape. This wasn’t a linear relationship, the flower adjusted parameters of the online soundscape, which resulted in incrementally changing patterns of sound.

Figure 15. Ways To Wave. 2008. Squaretangle. The online physical environment was a 3D volumetric space, which reacted to both the movement of the petals of the lotus in the real gallery space, and also to the movement of the avatars of online users in second life. Thus the gallery attendees and the online players could collaborate to co-create the soundscape and visual layer.

Figure 16. Ways To Wave. 2008. Squaretangle. Adam Nash pictured, “playing” the lotus flower in the San Jose Museum of Modern Art. The results of the interventions are streamed into second life where they enact changes in the scene pictured on the wall behind Nash. The actions change the shape and colour of the virtual scene and aid in the evolution of the resultant soundscape.

Figure 17. ASXeDancer. 2010. Squaretangle. Motion capture data used to animate characters in Second Life. The data was designed to take over visitor’s avatars when within a certain proximity, a kind of “Dance of Death” a Dance Macabre. Adam Ramona seen dancing.

Figure 18. ASXeDancer. 2010. Squaretangle. ASXeDancer was only ever realised as a prototype, yet it contained themes that continued into future work using motion capture data to allow connections between performing agents. The work was meant to be fun yet somewhat disconcerting. The visitor’s bodies being taken over by an unseen stream of data, events proceed on a trajectory of their own, independent of the “central” object. In a simple way it looked at hidden power structures and asked “where does agency lie?”
Figure 19. Façade. 2010. Squaretangle. Ars Electronica Futurelab. Façade used a motion capture system to allow the artist to paint the façade of the Ars Electronica build using movement. Here can be seen the Optitrack Arena software used to manage the motion capture system (lower half) and the façade simulation software used to interface to the external lighting of the building. ................................................................. 19

Figure 20. Façade. 2010. Squaretangle. Ars Electronica Futurelab. Author in motion capture suit used with the Optitrack motion capture system. The data was streamed to unity where it was filtered and mapped to the Ars building façade topography. ................................................................. 19

Figure 21. Façade. 2010. Squaretangle. Ars Electronica Futurelab. The main Ars Electronica building with a stylised representation of the human performer being animated in real-time across the façade. Each window in the building is embedded with multi-colour LED lighting which is programmable in real-time. The motion capture data can thus be visualised using the building as a 3 dimensional display. ................................................................. 20

Figure 22. Façade. 2010. Squaretangle. Ars Electronica Futurelab. Gesture recognition of simple gestures drawn with the hands using the motion capture system. Here a circle shape is one of the gestures identifiable by the system. The gesture recognition was written in Unity and the gesture visualisation sent to the façade of the building. ........................................................................................................................................................................................................................ 21

Figure 23. Façade. 2010. Ars Electronica Futurelab. Freeform painting using the motion capture data. Façade was an example of data derived from human movement re-emerging again in physical space. Most of the work prior to this tended to use video projection onto flat surfaces and screens. While the building surfaces are a type of screen, the ability to paint data to such a large physical object highlighted the potential for data to be transformed from one physical state to another. ........................................................................................................................................................................................................................ 22

Figure 24. Façade. 2010. Ars Electronica Futurelab. More free-from painting of the façade using motion captured movement. ................................................................................................................................. 23

Figure 25. Façade. 2010. Ars Electronica Futurelab. The lower section of the building pulsed through colours in time with the soundtrack played along with the performance. Frequency analysis of the soundscore was used to generate the colours. ........................................................................................................................................................................................................................ 23

Figure 26. Reproduction 1. 2010. Squaretangle. Adam Nash & John McCormick. Reproduction was our first major foray into generative agent performers. The entities were spawned from code when the simulated world was created, and then evolved according to simple rules. They are performing audio-visual creatures that have a collective song that changes over time. This first version presented at Neutral Ground Gallery in Saskatchewan saw the 9 species of entities spawn and then seek out other members of their species to reproduce. They also sought out other species to feed on their colour to gain the
proportions necessary for their own survival. Their behaviour in earlier iterations was somewhat unpredictable. Sometimes we would leave the gallery and return to find one species had multiplied to the point of bringing the computer to its knees with their sheer numbers. At other times, they would go exploring thousands of virtual kilometres away, leaving the gallery totally bare. The Reproduction environment was projected onto all four walls and ceiling of the gallery using a single projector and hemispherical mirror. Warp mapping of the scene was generated in the application to unwrap it on the gallery surfaces.

Figure 27. Reproduction 1. 2010. Squaretangle. ................................................................. 24
Figure 28. Reproduction 1. 2010. Squaretangle. ................................................................. 25
Figure 29. Reproduction 1. 2010. Squaretangle. Differe...
Reproduction 2. 2011. Squaretangle. Screen Space Gallery Melbourne. Reproduction 2 is also a step towards performing agents who have autonomy to pursue their own goals, affording the potential for emergent behaviour. This hands-off approach to designing performing agents would influence future research into digital performing agents. While much of the behaviour of the entities is programmed, the behaviours are specific to particular purposes, in this case their relationship to human species members. The rules do not govern their song generation yet song structure still arises as the entities pursue their own compulsions.

Reproduction 3. 2012. Squaretangle. Reproduction 3 saw our favourite performing entities inhabiting a 10 metre dome. Humans entering the dome were again tracked using the kinect sensor, and the entities were projected onto the entire surface of the dome. The entities could be viewed from both inside and outside the dome.


Reproduction 3. 2012. Squaretangle. Werribee Piazza. Presented outdoors in the Werribee Piazza, passers-by along the main retail strip of Werribee were attracted to the song and shifting colours of the entities. The outdoor public version provided some surprising challenges with a number of patrons bringing their children in prams and their pet dogs into the installation. The installation tended to follow the adults in these circumstances, however the informality of the environment and patrons was very refreshing for the entities and they responded well.

Reproduction 4. 2013. Squaretangle. Adam Nash, John McCormick, Stefan Greuter. Interactive Entertainment Conference, RMIT. Reproduction 4 was presented on an 8 stereo screen octagonal display. This time visitors could move around the installation and the entities would swim out from the centre to greet them with their characteristic songs. Each of the eight screens had their own kinect, which were calibrated to the same coordinate mapping in the entities world. Thus visitors to all of the screens occupied the same virtual arena in which the entities performed. Adam Nash seen here being sung to by a group of entities of his species. His attention to the entities and movement in relation to them elicit changes to their sung responses. Adam and the entities (and visitors when they interact) enter into a performing partnership from which the soundscape emerges.
Figure 40. Dancer Steph Hutchison with reflective markers attached to her body and costume for use with the Motion Analysis motion capture system. Normally the dancer would wear a black lycra suit to which the markers would be attached. Here the dancer is in a more minimal costume in order for her movements to be more clearly seen by the live audience. Her movement data is being streamed to the Unity game engine to visualise her movement as the blue lightning character. ...................................................................................................................................................................................................................................... 39

Figure 41. Dancer Erin xxx in a black lycra suit with attached reflective markers using the Optitrack motion capture system. In the initial tests I used both the Motion Analysis and Optitrack systems in order to test that the neural networks could work with different systems data., which was the case. In this short work, Erin is represented by a humanoid avatar while the stars circling her later exhibit flocking behaviour, using the dancer as their epicentre. .......................................... 40

Figure 42. SOM structure. Each recorded motion input frame has 79 components or vectors. The layer represents the 1600 neurons, a 40 x 40 map. The weights of the neurons are adjusted with a decreasing neighbourhood radius and by a decreasing amount to produce the final output SOM. .......................................... 41

Figure 43. SOM neighbour distance map. The grey dots are the neurons. The closeness of the neuron’s weights to that of its neighbour’s is shown going from black to yellow, black being closer, yellow further away. ........................................................................................................................................................................ 41

Figure 44. Final mocap suit and marker setup for the Motion Analysis system used for most of the experiments and performance. ........................................... 44

Figure 45. The markers and skeleton in Cortex, the software used with the Motion Analysis system. The skeleton is driven by the markers which represent the markers on the dancer’s body. .......................................................................................................................................................................................................................................................... 44

Figure 46. markers in Arena the software used with the Optitrack motion capture system. Like the Motion Analysis system, the data can be recorded for offline processing and playback as well as streamed live to other applications for real-time use. .......................................................................................................................................................................................................................................................... 44

Figure 47. SOM Neuron Hits. Showing the number of frames of data each neuron has encapsulated by being the Best Matching Unit (BMU) If a neurons weights are closest to the input data that neuron claims the data and moves its weight a predetermined amount towards the input data. Frames of data equating to movement postures that are similar will be added to the same neuron. In this test, some neurons have hundreds of hits, some have none, depending on the nature of the data itself. .......................................................................................................................................................................................................................................................... 45

Figure 48. SOM Hits. Showing eight frames of data after being presented to the trained SOM for testing. Four of the frames of data are of one posture and the other four are another different posture. We can see that the frames of like movements are captured by the same neurons. .......................................................................................................................................................................................................................................................... 47
Table of Figures

Figure 49. SOM Weight Positions. A 2-dimensional representation of the first two vectors of the 79 inputs. The green line is the input data and the grey points are the neurons with connections in red. The grey neurons have moved to closely match the input data. ................................................................. 48

Figure 50. SOM Colour Clustering. A test of the SOM algorithm in Unity. A 100 x 100 map of neurons is initialised with random weights representing Red Green and Blue (RGB) colour values, three vector weights per neuron. ......................................................................................................................... 49

Figure 51. SOM Colour Clustering. After training the SOM has managed to cluster like colours into different regions of the map with no supervision or labelling of the data. ................................................................................................................................. 50

Figure 52. SOM training using motion capture data in Unity. This 3-dimensional representation shows the input vectors (white) and neuron weights (red). Over 500 iterations the red spheres will gradually move out towards the input vector positions. ......................................................................................................................... 51

Figure 53. SOM training using motion capture data in Unity. The red neuron weights gradually become closer to the input weights. .................................................. 52

Figure 54. SOM training using motion capture data in Unity. At the end of training the SOM neuron weights will closely match the topology of the input vectors from the motion capture data. The offset between the red and white spheres is artificial to aid seeing the separate spheres. In reality the red spheres will more closely overlay the white spheres. ................................................................................................................................. 53

Figure 55. Testing recorded motion data into the trained SOM. Recorded motion is streamed from Cortex to Unity. The data steam is used to animate the white avatar. The current position and joint rotations of the white avatar are then passed to the SOM and the neuron with the best match has its weights used to animate the red avatar. With the test data the matches are usually very close as the test data was recorded using the same dancer performing the same sequences on the same day. This showed the SOM’s ability to follow the movements of the dancer from moment to moment, a form of immediate recognition. ................................................................. 54

Figure 56. Testing recorded motion data into the trained SOM. The SOM was often close in matching the movement of the incoming stream from the test motions, but was also often seen matching with a movement that was different yet still related and plausible. ................................................................................................................................. 55

Figure 57. The standardised skeleton used throughout the experimentation and performance in Unity. This allowed the live and pre-recorded data streams as well as the SOM neuron weights to have the same structure and be applied to the avatars interchangeably. ................................................................................................................................. 56

Figure 58. Live capture and streaming of the dancer’s movement in Cortex ................................................................................................................................. 57
Table of Figures

Figure 59. The streamed motion data displayed on a humanoid avatar in Unity. ................................................................. 57
Figure 60. Rows of data used for SOM training. The first 3 numbers are XYZ positions, each subsequent 4 numbers are quaternion rotations. ...................... 58
Figure 61. Motion Analysis HTR data structure. .................................................................................................................. 58
Figure 62. Process for training the SOM.......................................................................................................................... 59
Figure 63. Mapping SOM neuron weights to the avatar’s skeleton. Left, the trained SOM, centre as close up of the neurons with larger ones representing neurons with more hits, right one neuron’s weights transferred onto the avatar’s skeleton in order to animate it. .................................................. 59
Figure 64. Process for creative development of the agent learning for rehearsal and performance. .............................................. 60
Figure 65. Testing the agent’s ability to follow the dancer’s movement after training.......................................................... 61
Figure 66. The agent was able to follow the dancer successfully when the dancer worked within the range of movement she had passed to the agent to learn. Anything outside the vocabulary it had learnt caused more unpredictable behaviour. ................................................................. 62
Figure 67. Agent (red avatar) following the dancer (white avatar). ............................................................................................ 63
Figure 68. The agent would sometimes perform movement that was similar but not identical.................................................... 64
Figure 69. The SOM layer the agent uses for postural generation. This is the more regular form of vanilla SOM. In order to string the postures into sequences, some form of temporal awareness is required. Tracing the pathways between neurons as they fire proved to be one way to enable temporal connections between the neurons. ................................................................................................................................. 65
Figure 70. Tracing temporal connections between neurons as they fire. This concept I modelled on synaptic plasticity. As the neurons fire a connection is made from one neuron to the next. A second map contains these connections. If a path from one neuron to another is fired more that once the strength or weight of that synaptic connection is strengthened accordingly ........................................................................................................ 66
Figure 71. Using the Self Organising Synaptic Map (SOSM) to enable the agent to generate its own movement based on the movements it has learnt from the dancer. The dancer’s avatar (white) is animated by the live motion data stream, the agent’s avatar (red) is animated continuously from its SOSM. .............. 67
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>72</td>
<td>Using the SOSM to allow the agent to generate movement based on the movement performed by the dancer. The agent looks at the current movement of the dancer, as in the following task, but then allows its SOSM to take over and generate movement by following the links in the synaptic layer of the SOSM.</td>
</tr>
<tr>
<td>73</td>
<td>Process for training the Hidden Markov Model (HMM). The HMM is used to recognise short movement sequences performed by the dancer.</td>
</tr>
<tr>
<td>74</td>
<td>The dancer recording movement to train and test the Hidden Markov Model.</td>
</tr>
<tr>
<td>75</td>
<td>Motion capture software to record HMM phrases.</td>
</tr>
<tr>
<td>76</td>
<td>Movement played back onto the standard avatar to record the motion data to be used for the SOSM and HMM. 12 trials of each of the seven sequences were recorded. Two of each were used for creating the SOSM, then eight of each to create the HMM. The other four were used for testing the HMM.</td>
</tr>
<tr>
<td>77</td>
<td>Test results from the Hidden Markov Model Gesture Classifier. Using the test recordings of the dancer’s movement “gestures”, the HMM was able to correctly classify the movements with 100% accuracy. Four of each of the eight full-body dance phrases were presented to the HMM which was correct on all occasions.</td>
</tr>
<tr>
<td>78</td>
<td>Grevillea Crystals Incarnadine. The agent (red avatar) attempts to follow the movements of the dancer (white avatar).</td>
</tr>
<tr>
<td>79</td>
<td>Grevillea Crystals Incarnadine. The closeness of the agent in being able to follow the dancer is visualised as a crystal flower. The differences in their respective joint positions are represented by the length of a petal for each joint.</td>
</tr>
<tr>
<td>80</td>
<td>Grevillea Crystals Incarnadine. Later the joint rotations are also used to change the amount the petals curl and bend.</td>
</tr>
<tr>
<td>81</td>
<td>Grevillea Crystals Incarnadine.</td>
</tr>
<tr>
<td>82</td>
<td>Grevillea Crystals Incarnadine.</td>
</tr>
<tr>
<td>83</td>
<td>Recognition. The agent (red) and dancer (white) performing individually. The dancer is tracked by kinect to animate her avatar. The agent is dancing continuously thanks to its SOSM.</td>
</tr>
<tr>
<td>84</td>
<td>Recognition. When the dancer is on stage the morphing creature uses her movement to animate itself. When the dancer leaves the performance area, the agent’s data is used by the morphing creature to animate itself. The agent and dancer work together to provide a continuous stream of data to the creature.</td>
</tr>
</tbody>
</table>
Figure 85. Recognition. As well as providing movement to quickly train the agent how to dance, The artist’s iris textures were also used to clothe the creature to give it identifying features based on the human’s. .......................................................... 84

Figure 86. Recognition. The iris textures mapped to the morphing creature. Along with the movement given over the to the agent to learn, the iris textures gave an immediately organic quality to the creature, familiar yet alien. The performance version of Recognition also used echocardiogram recordings of heart activity to further accentuate the transformed human origins of the agent. .......................................................... 84

Figure 87. Recognition Installation version at Cube 37, Frankston Arts Centre. The creature was projected onto the glass front of the gallery. A kinect sensor behind the glass tracked interacting patrons and drew on their movement to animate the creature. The installation used the same learning agent as in the performance. ......................................................................................................................... 85

Figure 88. Recognition installation at Cube 37. The installation ran through the night and when there were no humans present the agent would dance in their stead using the movement it had learnt from dancer Steph Hutchison. ......................................................................................................................... 85

Figure 89. Instrumental. The agent creates movement phrases based on the current movement of the dancer. .......................................................... 86

Figure 90. Instrumental. The pathway of the agent is drawn as a trail of crystal that becomes a musical instrument. .......................................................... 86

Figure 91. Instrumental. Crystal spheres falling form above create a soundscape of bells as they collide with and rebound through the structure created from the agent’s path. ......................................................................................................................... 87

Figure 92. Verbose Mode. The agent and the dancer both have particle physics based avatars. The dancer’s is animated from her movement through the motion capture stream, the agent’s from her voice using frequency analysis. ......................................................................................................................... 88

Figure 93. Verbose Mode. The dancer performs particular full-body gestures that the agent has learnt to recognise through training of a combination of SOSM and HMM. When the agent thinks it has recognised a gesture it interjects commentary on the dancer’s performance. This commentary ranges from playful, to ironic to sarcastic, depending on the particular phrase the agent decides to speak. ......................................................................................................................... 89

Figure 94. Verbose Mode. There is a conversational quality to the sequence of events, a to-and-fro as in a verbally sparring couple or good friends camaraderie. .......................................................... 90

Figure 95. Verbose Mode. The agent was able to correctly recognise the dancer’s gestures around 90% of the time in performance. When it was wrong the dancer derived some amusement from its incorrect guess, which in no way detracted from the work. ......................................................................................................................... 91
Table of Figures
Table Of Equations

Equation 1. SOM Euclidean distance function where V is the current input Vector and W is the current neuron’s Weight vector....................................................... 42

Equation 2. SOM exponential decay function to determine which neurons surrounding the Best Matching Unit also have their weights adjusted. ......................... 42

Equation 3. The amount by which the neurons have their weights adjusted. where W is the neuron weight, t is the time step, θ influences the amount the weight is adjusted by the learning rate, diminishing with the distance from the best matching unit, and L is the Learning rate, the amount the weight is ultimately adjusted, which is also a diminishing amount of the difference between the neuron’s weights and the input values................................................................. 43

Equation 4. where dist is the distance of the neuron from the BMU and σ is the width of the neighbourhood function from Equation 2............................................. 43

Equation 5. HMM where x is the sequence of observations, y is the sequence of states, p_{yt|yt-1} is the probability of being in the state yt given the previous state was yt-1 and p_{xt|yt} is the probability of encountering the observation xt if we are currently in state yt........................................................................................................... 70

Equation 6. Hidden Markov Model Summary where the model λ is defined by the total number of states n, the matrix of Transition Probabilities A, the matrix of Observation Probabilities B, and the initial state probabilities describing the probability of starting in each of the different states in the model. .......................... 71

Guide to Reading
Guide to Reading

The relationship between the main text of this exegesis and the accompanying images differs between chapters. In the first two chapters, Introduction and Context, the images at times, serve as examples to illustrate the main narrative. But largely, they create their own historical narrative of works I have engaged in that have led to the current research into human / agent duet performance. Themes from previous works have heavily influenced the current directions I have taken, and I have included images of older works to illustrate that journey. The subsequent chapters describing the actual experiments and performances undertaken in this research contain a more direct connection between the images and the text, with the images largely illustrating the points made in the main body of writing.

I am indebted to Hellen Sky (Company In Space) and Adam Nash (Squaretangle) who were the main collaborators on many of the earlier works illustrated in the first two chapters.

Video Documentation

Video documentation of the experiments and performance associated with this research may be found at:
http://www.johnmccormick.info

Documentation of the final performance Emergence may be found at:
http://www.johnmccormick.info/category/emergence/

A simplified explanation of components of the research can be found at:
http://www.johnmccormick.info/research/
Publications
The following peer-reviewed publications resulted from this research project.


Abstract

This research project investigated the possibility for a digital dance environment to have a truly inter-dependent relationship with a live dancer. I wanted to develop a performing agent capable of dancing with a human dancer in a manner reminiscent of another human dancer. For the Digital Performing Agent to be able to perform live with the human dancer, the agent needed to be able to contextualise the movement the dancer was performing and to have a suitable movement vocabulary with which to contribute to the performance.

In this thesis I will discuss my research into the use of Artificial Neural Networks (ANN) as a means of allowing a software agent to learn a shared vocabulary of movement from a dancer. In particular I used a form of ANN called a Self-Organising Map (SOM). The agent was able to use the learnt movements to form an internal representation of what the dancer was performing, allowing it to follow the dancer, generate movement sequences based on the dancer’s current movement and dance independently of the dancer using a shared movement vocabulary.

In order to provide the agent with the ability to generate movement from the SOM, I made some temporal enhancements to the standard SOM based on the concept of synaptic plasticity. This enhanced version I have termed a Self-Organising Synaptic Map (SOSM), which proved capable of movement generation as well as movement recognition. By combining the SOSM with a Hidden Markov Model (HMM) the agent was able to recognise short full body movement phrases and respond when the dancer performed these phrases.

I considered the relationship between the dancer and agent as a means of supporting the agent’s learning and performance, rather than developing the agent’s capability in a self-contained fashion. To this end I employed concepts found in the area of distributed cognition as a framework for the inter-dependent relationship between the agent and dancer. This prompted the use of unsupervised learning techniques for developing the agent’s performance capabilities. The agent exhibited emergent movement behaviour in the performance with the dancer as a result of the methods used in this research.
Introduction

The dream of artificially intelligent agents with human-like capabilities has been persistent within many areas including robotics, entertainment, service industries and education. In dance there are also those aspiring to the creation of intelligent systems that can be truly responsive to human performers in a manner that shows some understanding of their aspirations in performance. There are many notable examples of a live dancer’s movement being used as the source for projected 3D graphics. Some allow the digital environment to make programmed choices based on the dancer’s movement. However, few use artificial intelligence to inform the choices available to the digital environment. Where there have been examples of using artificial intelligence techniques to create performing agents for dance performance, these agents have tended to use the dancer’s movement as the source for projected graphics rather than focus on creating a two-way performing relationship with the dancer. This thesis describes a unique development of a performing agent that can learn from a human dancer and effectively act as a performing partner with the dancer in a live performance. This work introduces a novel application of statistical-learning-based, real-time recognition and generation of full-body motion for real-time dance performance.

Many artists working with dance have aspired to the creation of responsive systems that enact change according to the dancer’s performance. Usually this manifests as visual or auditory...
Introduction

visualisations or sonifications of the dancer’s movement where the flow of information is predominantly from dancer to digital environment. This research goes beyond this largely one-way action by providing the dancer with responses upon which she too can enact thereby developing a truly interactive performance. The premise of this research was that the agent should be able to be treated by the dancer as she would another dancer, that they should be co-creators of the performance. Adhering as closely as possible to this premise required close analysis of the relationship developed between collaborating dancers throughout the process of conception, development and performance; a model of a creative process. The model I present is of a truly interactive dance performance containing non-linear choreographic structures. Linear, pre-determined choreography can make it difficult for the dancers to change track mid performance in response to interactions with their performance partners. Therefore semi-improvised structures allowing freedom to engage with the developing relationship between the performers are employed. Within this work the dance is based on structured improvisation to allow scope for both the dancer and agent to interact with each other to create the performance.

To pursue a collaborative creative model that is dance based, a means of sharing a movement vocabulary between dancer and agent during live performance was needed. In this project, machine learning techniques were developed to allow the agent to learn a shared vocabulary of movement from the dancer for use in the live performance. This movement vocabulary then became the basis from which the agent could develop its own representation of what the dancer was performing, and from which it could respond with appropriate movement stylistically in

Figure 2. Lines from Sentient Space 1.0. 2004. Company In Space. Lines used a fairly direct mapping of the dancer’s movement to the projected volume to visualise the movement. The dancer’s combined data from the two motion capture systems was combined to affect the volumetric display. The volumetric animation had a certain amount of elasticity in its movement, as it was not an exact reflection of the dancer’s combined movements. Combining the dancer’s movement data extended the notion that the volume reflected the connection between the dancer’s virtual skeleton representations.
keeping with the shared creative performance. This raised the two main capabilities the agent required in order to effectively perform with the human dancer in a meaningful way; it must be able to recognise what the dancer is performing and also be able to respond in an appropriate manner based on the shared movement vocabulary it has learnt.

The use of a learning paradigm informed the choice of machine learning techniques utilised for the agent. Drawing on the analogy of two dancers working together, a common scenario is one dancer creating movement sequences and passing them on the other to learn. In contemporary dance it is not unusual for the receiving dancer to learn the movement and to display personal idiosyncrasies in the subsequent execution of the movement. A degree of personalisation is inevitable due to the different morphologies and backgrounds of different dancers. Movement is rarely learnt unchanged from the original. The fact that I am using human models of activity and creative process led to the use of learning structures that are themselves inspired by human biology and learning mechanisms. Artificial Neural Networks (ANN) are biologically inspired computational models that seek to emulate the neural structure of biological organisms, and have been used successfully for machine learning and pattern recognition. As such they are an appropriate choice for learning human movement and recognising patterns inherent in the movement data. I have chosen an unsupervised form of learning for the Artificial Neural Network whereby the movement data is presented to the network for learning without any labelling or directed goals; the network must find inherent patterns within the data completely undirected. Just as a human dancer must learn to cope with processing new movement, so too should the agent,

Figure 3. Paper from Sentient Space 1.0. 2004. Company In Space. The sheets of transparent paper responded to the velocity of the dancer’s movements. Each alternating sheet was influenced by one or the other dancer. The sheets were flung out with large movements and gently oscillated with less vigorous movements. They also had their own elasticity, however they returned to their stacked starting points in the absence of any movement. The two skeletons can be seen in blue and red, however they weren’t visible in the performance.
and just as one cannot tell exactly how the movement will be internally integrated by the dancer, so too with unsupervised learning, the agent’s ANN will assimilate the movement data in a unique manner.

In order for the agent to achieve both recognition and response of the dancer’s movement I found that a combination of machine learning techniques produced the best outcomes. Combining Artificial Neural Networks with a Hidden Markov Model enabled the agent to both recognise the dancer’s movements and perform with the dancer using the movement it had learnt.

The main conduit for the transfer of information to the agent for both learning and interaction with the live dancer is motion capture data. In this research, optical motion capture systems are employed, which provide sub-millimetre accurate descriptions of the positions and rotations of the dancer’s limbs. High quality motion capture was used to represent the dancer’s movement because it gave a reproducible, predictable and relatively unambiguous depiction of the dancer’s movement. There are many other ways in which the dancer might be represented such as marker points, point cloud and path trajectories. However in order to fulfil the desire to give the agent and dancer similar characteristics, a skeletal representation was employed as it made reading the movement clear by presenting a humanoid shape and structure. Using a humanoid skeleton is a typical method of visualising human motion data and was thus used for the agent as well, allowing both agent and dancer’s movement to be recognised in a typically human manner. This allowed the movement performance of the agent to be directly compared to the performance of the dancer. It also allowed the dancer to interact with the agent in a familiar manner, she could view the
agent’s movement in human-like terms, aiding the interactive relationship between them.

The development of a performing agent capable of meaningfully interacting with a dancer in a live performance is a great challenge. The agent is required to have the capability of recognising what the dancer is doing and responding in an appropriate manner that allows the dancer to be reciprocally stimulated. Aside from the technical challenges of designing such a system, the question arose as to whether there was an appropriate conceptual framework that might guide the agent’s development and allow it to be supported within the performance relationship with the dancer. The related areas of embodied and distributed cognition and cognitive scaffolding provide a structure through which to conceptualise the interwoven relationship between the dancer and agent throughout the creative process. While this research was not directly focused on proving a theory of cognition for the agent, it seemed highly appropriate to postulate a framework that might aid in the design of both the agent and the performance.

In the final outcome of this research I present four live performances between dancer Steph Hutchison and a performing agent dancer. The dance works illustrate the developing capabilities of the agent and the potential for emerging choreographic relationships between the dancer and agent. In the first work *Grevillea Crystalis Incarnadine* the agent follows the dancer as she performs, based on dance sequences the agent has learnt. While this was the first and perhaps most simple choreographic structure for the agent, it nevertheless required the agent to have a continuous internal representation of what the dancer was performing at any moment. This

![Figure 5. 2 Avatars. 2002. Monaco Dance Forum. 2 Avatars was developed as part of the Motion Capture Tech laboratory “Real Time and Networked: Sharing the Body” facilitated by Scott deLahunta. Dancers Nik Haffner and Thomas McManus each wore Gypsy exoskeleton motion capture systems. One of the dancers was able to control the flight of the other dancer and in doing so the second dancer’s avatar would stretch and flow in response. Thus they both had their own intrinsic movement, however there was also an external relationship governing their duet that was a product of the transfer of movement data from one body to another.](image-url)
internal representation of movement was not exactly like the dancer’s movement but it was the closest match from what the agent remembers of the dance it had learnt. The agent had no language with which to describe or associate to the movement. It had only its continuously stimulated neurons, which formed an internal representation of what the dancer was doing in response to the dancer’s incoming movement data. This deep, neuronal connection between the dancer and agent is the basis for the subsequent choreographic investigations. This approach is indicated by theories of distributed and scaffolded cognition, whereby the agent is able to make use of its environment to aid its functioning in the world. In this case the agent’s environment is the movement data it is “sensing” through the motion capture system.

The second work I present, Recognition, sees the agent and the dancer performing together, yet independently. The agent continuously performs its own dance using the movement vocabulary and style it has learnt from the dancer. This work provides proof of the agent’s ability to synthesise movement using what it had learnt from the dancer. Unlike most previous work into movement synthesis, which focuses on creating variability in style, this work seeks a uniformity of style within the performance, so that the movement of both dancer and agent follows similar trajectories. There is always room for improvement, and the agent is still learning. Nevertheless this dance demonstrates the agent’s ability to synthesise movement from its artificial neural network in a way that is consistent in style to the human dancer.

The next dance, Instrumental, takes the capability generated in Recognition further, adding the
ability for the agent and dancer to more fully interact in the generation of their collaborative performance. The dancer can provide movement motifs as starting points from which the agent generates dance material based on the shared vocabulary it has previously learnt from Steph. This work draws upon the learning metaphor that new learning builds upon prior learning. In *Instrumental* the agent finds new ways to make use of prior learning by essentially accessing its memories in a non-linear fashion to produce movement appropriate to the “seed” provided by the dancer’s current movements. The ability for the agent to feed off the dancer’s movement, and for the dancer in turn to have a recognisable human-like form and quality of dance to respond to gives this work a unique collaborative characteristic in the shared generation of the live performance.

In the final performance work, *Verbose Mode*, I wanted to extend the analytic capabilities of the agent further to encompass recognition of longer movement phrases. In *Instrumental*, the agent has a representation of what the dancer is performing at any moment. However, it is very short term whereas the ability to recognise longer movement “gestures” would allow a form of interaction between agent and dancer motivated by a deeper understanding of the dancer’s intent. Using the existing Artificial Neural Network alone proved unsatisfactory to the task, yet I was hoping to be able to further build on the past learning rather than use an entirely new technique. By combining the Artificial Neural Network with a Hidden Markov Model, the agent was able to learn to recognise patterns of movement as they stimulated a pathway through the neural network, and thus recognise the movement patterns these stimulated pathways represented. The result was the capability to recognise specific dance phrases and to respond to these upon identification.
It also added incrementally to the learnt capabilities of the agent.

This thesis is divided into 5 sections. Chapter 1 contextualises the main areas of research underpinning the performances. There are a number of domains within which this research operates, three of which are primary to the work: Machine Learning and Agent-based Artificial Intelligence, Theories of (Artificial) Cognition and Extended Dance Performance. In this chapter I contextualise this project within previous work undertaken in these areas that are applicable to the ultimate goal of creating highly interactive dance performance using an intelligent agent performer.

Chapter 2 documents the initial studies where machine learning techniques were used to build the agent’s abilities, as well as develop the underpinnings of the human – agent performance relationship. The main focus is on the solutions to the technical challenges of creating an agent performer and how it evolved through the iterative process of rehearsal with a human performer. In Chapter 3 the four performance works are described and the practical application of the agent is documented. The transition from generating the basic performance capability between an agent and human performer where both are visualised as humanoid avatars, to looking at the relationships that have emerged and how these might, in turn, be visualised as 3D environments are investigated. Chapter 4 draws together the results and conclusions of this research including the successful artistic and technical outcomes. Some ethical considerations are also discussed in this chapter. Chapter 5 concludes with a description of tangential investigations that have arisen while undertaking the research, which indicate future directions for both artistic works and further...
development of performing agents.

The methodologies employed in this research are interspersed within Chapter 2, Experiments and Chapter 3, Performance. I have chosen to introduce the methodologies chronologically as they developed in response to the experimental creative process rather than as a discrete chapter. I feel this describes the research journey better and elucidates the connections between the choice of technological methods and the ongoing artistic development. This approach also shows how the results of one experiment led to incremental changes enabling the next experiments.

The main methodologies for developing the performing agent are described in Chapter 2, Experiments. Chapter 3, Performance then looks at how the accumulated capabilities of the performing agent fared in live performance, where the exigencies of the dancer’s and agent’s choices come in to play.
Background

I have been engaged within the area of technologically mediated movement performance for a number of decades. Throughout that period I have been involved in the creation of many artworks, working with the artistic groups Company In Space and Squaretangle, that place the performer within a digitally mediated experience. There has always been the question for me regarding the nature of the relationship between the human in and with emerging augmented environments. Many of the works I have been involved in creating probe this issue in different ways. Escape Velocity (1998) (Figures 10,11) was a duet between Hellen Sky and Louise Taube, which looked at the nature of personal interaction and representation within shared environments covering extreme distances both nationally and internationally. (1) Trial By Video (1998) (Figure 12), another intercontinental work investigated the level to which external forces can exert influence and control through electronic communication mediums. (2) CO3 (2001) (Figure 13) and Sentient Space (2002) (Figures 1-4) used motion capture to explore the nature of embodied interaction in shared virtual worlds. (3) Ways To Wave (2008) (Figures 14-16) used physical sculptural instruments to interface to online environments to create a symbiotic synaesthetic experience between online and offline participants. (4) The Reproduction series (2010 -2013) (Figures 26–39) saw evolving, musical, generative entities that created multi-species chorales in response to the presence and interactions of human members of their species. (5) Throughout these works there is an increasing sense of the environment having greater capacity for autonomous action in relation to human presence.
One view of these performances is that these environments extend the human capability. They are an extension of ourselves into digitally enhanced territory. The data is still our data and it is achieving our own purposes in digitally augmented space. Many interactive artists seek to go beyond this direct mapping of cause to effect in their artworks by introducing the capacity for ordered or even chaotic unpredictability in their work. I have always thought of the relationship between human performer and augmented environment as an entering into somewhat unknown territory. There was always some uncertainty regarding the relationships formed within the augmented realm. The performer’s data was not always mapped directly into the augmented space; it entered the space, yet was subject to the uncertain interactions of other people’s data, and to uncertain intentions and the changing properties of the environment itself. Yet, as in life, we could see ourselves enacting our part within these augmented realms, and outcomes were still connected to our actions.

Performances such as CO3 and Sentient Space which incorporated live motion capture of dancers within a projected stereo 3D environment, raised questions regarding the nature of collaboration and the use of immersive digital sound and visual environments as a component of live dance performance derived from the real-time motion capture data. Within these performances a small finite set of responses was programmed into the software environment and any changes to these possibilities often required intervention from a human technical operator. Typically the motion data was used as a direct source for visualisation without any analysis by the software environment. While this led to many satisfactory results within the performances, it raised the possibility that the
software environment could potentially have a greater capacity for interpreting and responding to the dancer's movements.

Motion capture was also used by other groups in a dance context during this period. Lintermann, Haffner and McManus used a magnetic capture system to explore the transformation and visualisation of body movement with their *Time Lapses* project at ZKM in Karlsruhe. (6) Gibson/Martelli used motion capture to explore camouflage and disguise in *Summerbranch*, (7) and the relationship with virtual representations of their real world environments in *Swan Quake*. (8) I was fortunate to have worked on the motion capture programming in *Summerbranch* and *Swan Quake*. Troika Ranch have used their own Midi-Dancer system as well as single camera tracking methods to create complex interactions with their visual and sound environments in works such as *loopdiver* and *16 [R]evolutions*. (9)

The performances I worked on tended to be highly collaborative affairs with input from choreographers, dancers, composers, musicians, programmers and 3D designers. With the software environment becoming an extremely important part of the visual and aural component of the performance, the notion of the software environment as a collaborative partner in the creative process began to develop. The most appropriate context for this shift to software as collaborative partner seemed to be the intelligent software agent paradigm. (10) Attributes associated with software agents include; reaction to and tight coupling with the environment, autonomy, goal-orientation and persistence, and these characteristics distinguish the agent from a normal running...
program. These attributes also seem desirable in relation to the goal of enabling the software environment to take a more active role in the collaborative, creative process. The quality of autonomy implies a capacity for self-direction in the software agent, which in turn raises the notion of some form of intelligence however rudimentary.

This line of thinking led me to search for models of intelligence that might be useful in guiding the development of an autonomous, synthetic agent. The extensive literature on the philosophy of mind provided many possibilities, however one in particular, the area of situated cognition (11, 12), seemed to offer a challenging paradigm within which to explore the development of an intelligent agent within a live performance context. Situated cognition, with its premise of extremely tight coupling of cognitive processes to the environment, seemed potentially aligned with both the desire to more closely couple the dancer and performance environment and the development of a synthetic agent which could also respond intelligently to its environment of which the dancer is a major part.

While the predominant focus of this research was on the development of enhanced capabilities for new artistic performances, this development is pursued through the interdisciplinary exploration of three key areas; the development of an intelligent software agent (science and engineering), the creation of new performance possibilities (creative arts practice) and the application of situated approaches to cognition (philosophy of mind). These fields of enquiry are complimentary and in combination, offered a potential means of successfully achieving the goal of developing an

Figure 13. CO3. 2001. Company In Space. CO3 investigated the nature of the post-human condition, the emerging cyborg, as we get more enmeshed with the digital objects and artifacts that co-create our existence. Performed by Hellen Sky in a Gypsy exoskeleton motion capture system. In the first performance at the Astor Theatre in Melbourne, Sky was joined by Keith Roberson in Florida, who intervened in her actions using a Polhemus Fastrack tracking system. The interactions meant her form wasn’t always intact, often becoming a hybrid with the environment she inhabited. CO3 was performed in various configurations and contexts, always site-specific, and adaptable to the physical environment it inhabited. At the ICA in London, room-length metal walkways criss-crossed the space, around and under which the audience wandered to view the performance.
intelligent performance environment that builds on past experiences of artistic creation.

Artificial intelligence techniques have been used to visualise, sonify, and respond to dancers’ movement in performance for many years. (13-17) I wanted to explore whether it was possible to develop a performance agent that could participate in some way within the choreographic process as well as within the performative outcome. The development of the synthetic agent within a predominantly artistic setting offered the opportunity to undertake research in which methods can be explored without having to adhere to a purely practical outcome or having to be concerned for the potential safety of the participants as might be the case in a medical, engineering or scientific setting. The use of a live performance environment also provided a real-world setting within which to test intelligent system models. In a live performance setting the agent must be able to perform "in the wild" in an uncontrolled environment with its concomitant uncertainties. With a live audience in attendance, the agent also needs to work well enough to make a satisfactory artistic statement.

Developing an Agent Performer

The move to an agent-oriented approach within the performance environment is an evolution that draws on advances in machine learning that are rapidly finding their way into our everyday interactions. The use of search engines, such as Google image search, are prime examples. My own research is a continuing response to the possibilities available to deepen the augmented relationship with hybrid spaces, to use them for the purposes of making artwork, yet to also open these possibilities up for scrutiny and analysis. I see the roles of the arts in this context as to aid
understanding, to open methods up to questioning and to propose alternative applications of these technologies.

To this end, the goals of the agent developed in this study are the recognition of, and response to, human movement data. The agent takes human data and learns its salient features in an attempt to better understand the human’s motivations, a form of data mining. These goals necessitate a consideration of how an agent might need to be conceptualised, if it were to be able to interact with a human performer. For this to occur, it is not sufficient to generate a system that can respond to motion data per se. The meaning of the motion data in a human performance, and specifically a dance context, must be considered part of the interactive process in any system involving a human. To this end, I examined a number of concepts related to the cognitive aspects of movement and dance, as a means of developing a framework for creating the agent. These considerations form the basis of the discussions below.

Memory

Human collaborators are able to make use of their experiences and memories to respond to developmental concepts and synthesise possibilities in relation to new artwork. If an emergent software agent were to be considered part of the collaborative process, what traits would be beneficial to them? Some form of memory would be essential to imbue the system with references to apply to incoming stimulus, or some substance with which to synthesise possibilities.

Matt Carter in Minds and Computers writes:
... embodied experience was a necessary condition for the development of semantics, which, in turn, are necessary for having a mind. Consequently, if we want to develop an artificial intelligence it must, in the first instance, be connected to the external world in the relevant ways. In other words, it must enjoy sensory apparatus which mediate the relations between it and the external world. Furthermore, our embryonic artificial intelligence must then be able to gather a weight of experience, through which it will be conferred with mental representations. (18) p.206

Memory, the weight of experience, is seen by Carter as a fundamental building block upon which mental representations may be constructed. Some form of memory is crucial for the agent to be able to contextualise the human dancer’s movement and to perform its own movement that is consistent with the style and intent of the shared performance with the dancer. Without a form of memory, the sharing and retention of movement potentials is not possible. In humans, memory implies learning, and in particular unsupervised learning in that the associations formed in response to a particular event differ from person to person. Different people will remember the same event with varying details.

In dance, embodied experience and hence memory is embedded within the morphology of the human body. Memory in dance is procedural, in the sense that, like expert movement in other elite professions, once learned, complex dance movement phrases are performed without conscious cognitive awareness (19). Memory, in this case, is enacted only through moving one’s body. Erin Manning (20) argues that dance movement is also inherently relational, proceeding from a ‘pre-acceleration’ that defines intentionality in relation to the world and to other people as well as...
trajectory. She describes dancing a duet with another person as … *not a learning by heart. It is not a choreography. It is improvising with the already-felt.* (20) p.380. Manning’s argument suggests that the procedural nature of dance memory does not imply that dance performance is fixed by the past, but rather that the body memory of past movement is brought to bear on the present moment. This process is constituted in terms of felt and experienced physical morphology and structural (skeletal) organization because memory encompasses the sensation of movement rather than simply a linguistic encoding of the pathways of joints and limbs in space.

For the performing agent I began looking for a way to enable it to learn and remember aspects of the movement provided to it from a human dancer. Artificial Neural Networks seemed to offer the desired characteristics for the agent’s memory. They are modelled on the neural structure of biological neural systems, they are capable of both supervised and unsupervised learning, they can contain persistent memories and they can respond to incoming information, in other words they can be interacted with.

Carter also introduces two other key concepts, that of embodied experience through sensory apparatus and the fundamental relationship between the agent and environment.

**Embodied Experience through sensory apparatus**

What exactly might the concept of embodied experience mean to the developing performing agent? Anderson in his essay on Embodied Cognition applied to Artificial Intelligence, (21) points out that there has been a profound shift in large sections of the AI community from Descartes’s “thinking
Chapter 1: Context

thing”, the bounded, subjective human being, to a more Heideggerian approach with agency and interactive coping being the main contributors to cognition. I’ll expand further on different theories of cognition that might be useful as a framework for the agent’s development later in this chapter. However, it is useful to first look at what the agent has at its disposal that might contribute to its sensory experience of the world.

The main sensory apparatus at the agent’s disposal is the motion capture system. The motion capture data contains a representation of the dancer’s movement in the form of positions and rotations of the dancer’s skeleton. It is a simplified representation of the moving body and can be used to visualise the postures of the body and the flow of movement over time. In the dualist Cartesian view, the motion capture sensors would be viewed as merely an input mechanism to the mind of the agent. While the sensors allow the agent’s physical extension into the world, they are only conduits for information passed to the symbolic representation contained in the neural substrate. An embodied view would see the motion capture system as the connection to the external world in the manner of the body, sensing and contributing to interaction with the world, creating meaning through this interaction.

While the motion capture data stream that the agent senses through its sensory system, the camera and network of the motion capture system, might appear superficially to be simply a flow of numbers representing joint angles and positions, it is also an emission of the dancer’s physical intelligence honed over years of training. The data is indicative of the decisions and choices made

Figure 18. ASXeDancer. 2010. Squaretangle. ASXeDancer was only ever realised as a prototype, yet it contained themes that continued into future work using motion capture data to allow connections between performing agents. The work was meant to be fun yet somewhat disconcerting. The visitor’s bodies being taken over by an unseen stream of data, events proceed on a trajectory of their own, independent of the “central” object. In a simple way it looked at hidden power structures and asked “where does agency lie?”.
by the dancer as she navigates the parameters of her own performance in the creation of the
original movement to be shared with the agent, as well as her reactions to performing with the
agent in their duets. This is a rich source of performative possibilities that the agent can draw upon
without having to contain them all within itself. Imagining the motion capture system as the sensory
apparatus of the agent which embeds it in the environment of data emanating from the dancer
offers the agent potentially enormous reserves of support if applied through the right framework.

Relationship between the agent and its environment

Mind, Body and Environment feature in varying roles in the Cognitivist and more recently,
Embodied approaches to cognition. The cognitivist approach, with its Cartesian underpinnings,
highlights a linear sequence of events consisting of sensory input, cognitive processing and finally
action. As Stewart et al. put it in their introduction to Enaction, (22)

In this scheme, “cognition” is thus sandwiched between two layers—sensory input and
motor output—which are not themselves considered as properly cognitive.

In the area of Artificial Intelligence, cognitivist thought has given rise to the Computational Theory of
Mind. This theory was put forward predominantly by Hilary Putnam (23-25) and elaborated by Jerry
Fodor. (26, 27) The cognitive component between input and output is characterised by symbolic
representation manipulated by formal rules, not unlike a computational machine. This explanation
implies that the agent’s cognitive faculties are bounded and self-contained within the agent’s neural
substrate. The body is the means of informing the mind about the state of the environment and the

Figure 19. Façade. 2010. Squaretangle. Ars Electronica Futurelab.
Façade used a motion capture system to allow the artist to paint
the façade of the Ars Electronica build using movement. Here
can be seen the Optitrack Arena software used to manage the
motion capture system (lower half) and the façade simulation
software used to interface to the external lighting of the building.

Figure 20. Façade. 2010. Squaretangle. Ars Electronica Futurelab.
Author in motion capture suit used with the Optitrack motion
capture system. The data was streamed to unity where it was
filtered and mapped to the Ars building façade topography.
body's internal states. The body is also the conduit for action in the environment. The body is viewed as a mechanistic structure supporting the separate, cognitive brain. This view of cognition has also had an influence on how human cognition has been viewed. However there have been profound changes with the adoption of embodied and situated approaches to human cognition and AI research is being increasingly influenced by these more recent embodied theories.

Situated Cognition takes a different view of the relationship between mind, body and environment. The distinction between mind and body are broken down and replaced by a concept of practical engagement with the environment. As Clarke states, (28)

...cognition must be understood in terms of how it functions under the pressure of real-time interaction with the environment.

The situated approach to cognition has paved the way for a diverse range of explorations into the connection between action, the environment and cognition. Erin Manning's "movement-becoming-thought" (20), the "enaction" of Stewart et al. (22), Alva Noe's "actionism" and "presence-to-mind" (29), situated cognition (12, 28, 30), the scaffolded mind, (31-33) the extended mind, (11, 34-38) are all in their own way, expressions of situated cognition.

Rupert distinguishes between the main forms of Situated Cognition as Embodied, Embedded and Extended. (36) Rupert states that for the Embodied view,

... human cognition bears a privileged relationship to distinctively bodily processes. Our fundamental concepts pertain directly to the physical body; other concepts are, in some important way, an extension of those fundamental bodily concepts.

With emphasis on

...the context-dependant effect of sensory experiences and motor routines on cognitive
Manning and Noe are notable examples of Embodied proponents. They make heavy use of a first person, phenomenological viewpoint in their investigation of mind. An agent’s interactive activities within its environment leads to perception.

The Embedded view of cognition Rupert summarises as having a heavy reliance of human cognition on the environment, with cognitive processes depending in surprising and complex ways on the use of external resources. These external resources often take the form of tools that extend the cognitive reach of the user. While these tools can extend the cognitive reach, in the embedded view they are not seen as part of the actual cognitive core.

The concept of extended cognition differs from that of embedded cognition in the consideration of what actually constitutes a part of the cognitive process. While an embedded cognition approach would draw the line at the boundary of the human organism, extended cognition views many tools as a functional part of the cognitive process and would thus include them as a primary cognitive element even though they are not part of the biological organism.

While the notion of interactive engagement of the agent with its environment is common to the embodied, embedded and extended explorations of situated cognition, there is some disagreement as to what actually constitutes a cognitive process, that is, where are the boundaries of cognition. Clark and Chalmers (11, 28, 39) have been vocal champions for the theory of extended mind.
Scaffolding to support the agent dancer relationship

Extended cognition has often focused on the use of tools as cognitive extensions, as exemplified in Otto’s notebook, a thought experiment described in Clarke and Chalmer’s seminal paper on the Extended Mind. Otto’s notebook, suffering from Alzheimer’s Disease, resorts to using a notebook as extended memory. According to Clarke and Chalmer’s concept of functional parity; the notebook is performing the same functions that if it was contained within neural substrate, would certainly be considered part of Otto’s cognitive system and should thus also be considered as part of Otto’s cognitive apparatus. As Theiner points out, this is an individualistic view and he proposes that looking towards cognition across social groups could offer more insights into the mechanisms of extended cognition. The investigation of cognition across social groupings might also provide insights applicable to our dancer – agent relationship.

As I am dealing with the emerging relationship between a human dancer and a digital performing agent, the notion of cognitive extension through tools is less appealing than theories that might take into consideration the relationship between two minds. A tool, no matter how sophisticated its construction, is fundamentally a new version of the kind of one-way mapping of human data to visualization system that has been accomplished many times in dance technology works. To achieve genuine interaction in which the agent can play an active and directed role, a means of exchange is needed. From the perspective of the agent, the human is a fully cognizant agent and offers a potentially deeper relationship than any single tool.

Figure 23. Façade. 2010. Ars Electronica Futurelab. Freeform painting using the motion capture data. Façade was an example of data derived from human movement re-emerging again in physical space. Most of the work prior to this tended to use video projection onto flat surfaces and screens. While the building surfaces are a type of screen, the ability to paint data to such a large physical object highlighted the potential for data to be transformed from one physical state to another.
The core of this research is the development of the relationship between a human and digital agent beyond the notion of the ‘system as a tool’, and I am therefore interested in frameworks that might help define this relationship as well as describe potential methods for creative development on the part of the agent and dancer.

Sutton et al., (38, 42) Barnier et al., (43) Harris et al. (44) describe research into socially distributed cognition in the form of shared remembering particularly between long term couples. They point out that:

...autobiographical memory is not simply about accurately recalling episodes from one’s past; it is also enmeshed in our broader ongoing cognitive lives. In general, we live these cognitive lives, and engage in the activities that constitute them, in the company of others. (42) p.35.

The case of semi-improvised performance by two dancers (or dancer and agent) might be seen as a case of collaborative autobiographical memory where the shared movement vocabulary is collaboratively revisited in performance. Sutton et al. use the terms distributed and scaffolded cognition somewhat interchangeably.

The concept of scaffolding has been used in theoretical biology, amongst other fields of research, as a neo-evolutionary framework to describe how social and cultural influences can affect the organism as well as evolutionary forces. (45-47) Evolutionary forces traditionally operate over generational timeframes, whereas cultural changes affecting a human can occur in a single lifetime.
as well as in generational timeframes. Scaffolding, as proposed by Theiner, Li and Capra in the context of evolutionary biology, exhibits many of the traits of extended cognition however it focuses on wider dynamic forces over civilizational, individual and momentary time scales and both genetic and non-genetic transmissions. (31, 40, 45)

Sterelny points to the scaffolding of young children’s minds by their mother’s. Like Clarke and Chalmer’s example of Otto’s notebook they are: reliably and easily available, routinely used, substituting for imperfect memory and trusted by default. (32)

The case of child to mother could be likened to that of experienced dancer and emerging agent. In performance the dancer can be relied upon as a trusted source of knowledge, substituting for imperfect memory of the agent. Sterelny also points out that the scaffolded support the mother provides the child has itself been shaped by others of the mother’s generation and indeed previous generations. In terms of our dancer and agent, the agent can draw not only on the dancer’s movement but also on her cultural inheritance.

Theiner points out that in the case of language or mathematic symbolic language, we do not create our own language but draw on other minds both present and past. (40) This applies at least to some degree to dance movement as well. In passing her movement to the agent for learning the dancer is passing on the result of a lineage of minds that will in turn support and enculturate the agent’s movement capabilities as well as provide a scaffold for further development. The movement
carries a history of thought and experimentation derived from earlier minds.

Research into the reciprocal interactions between brain, behaviour and sociocultural contexts has led Li to propose a framework termed “cross-level dynamic biocultural coconstruction”. The indications from this research are that the brain exhibits remarkable experience-dependent plasticity. (31) Li goes on to state:

...genetic activities and neural mechanisms themselves possess remarkable plasticity awaiting sociocultural contexts to exert reciprocal influences on them and to be the “coauthors” of mind and behaviour.

The implications of this approach for our agent – dancer relationship are quite profound. We need not think of our agent having to contain within itself all the capability necessary to be able to undertake the complex task of performing with the dancer. By applying the concept of distributed or scaffolded cognition, the agent can be seen to draw on the cognitive abilities of the dancer to “co-author” its collaborative behaviour within the performance.

While distributed or scaffolded cognitive theories seek to explain cognitive development in and across humans, we can make use of these frameworks to aid the development of our performing agent as it develops alongside the human dancer. Rather than view the agent as a self-contained entity with completely bounded capabilities, we can draw on the dancer-agent relationship as a cognitive system to allow the agent to quickly generate capability as a collaborative performer.
These differing approaches to cognition influence the way in which the agent’s capabilities are envisaged and consequently how they are designed. I believe the underlying assumptions regarding how the agent’s cognitive process might work have a huge impact on its resultant design and abilities. Looking at the different approaches, the question arises; could the use of a cognitive model for the agent assist in its development? If so, which would be the most appropriate to employ to best enable the development of the agent’s performance abilities?

I became most attracted to the concepts of distributed and scaffolded cognition. If these theories could guide the development of an agent that was inherently supported by its interactions with the dancer, it could potentially enable complex behaviours to emerge without continuous implicit direction from a software program.

I isolated two main capabilities as being essential for the agent to interact with the dancer, the ability to recognise appropriate traits in the dancer’s movement and the ability to respond in a meaningful manner. What this recognition and response would actually look like in performance was initially unknown. However I felt that developing these two abilities in some form was necessary. In order to achieve this the agent required some form of memory with which to, in Carter’s words, ‘be conferred with mental representations.’ (18) Memory is not considered as a storage and retrieval mechanism, but rather as a substrate of innate learnt potentials able to be influenced by the dancer – agent relationship to co-author the dance performance. Some form of persistent memory seemed essential if the agent was to have a reference point for the dancer’s
movement data. Also if the agent was to attempt co-creation of the performance, it could not come to the relationship empty.

**Artificial Neural Networks**

In contemplating the software agent as collaborator within an interactive performance, even in a very limited sense, I considered models of software that attempted to mimic human cognitive functions. Artificial Neural Networks (ANN) attempt to model the behaviour and capabilities of neural networks in the brain and have been successful 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, (48, 49) in human communication recognition is often a precursor to a response. In the context of this project, I needed to develop a system that could both recognise and respond to human movement via a motion capture data stream. There are many types of ANN. However I began to investigate a particular type of ANN, the Self-Organising Map (SOM) or Kohonen Feature Map, named after Teuvo Kohonen who first described them. (50, 51)

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 appropriate classes. This unsupervised form of learning allowed the agent to find its own associations in the movement data and created the conditions for agent behaviours to emerge from the learning process.
Artificial Neural Networks for detecting behaviour

Self-Organizing Maps (SOM) have been used successfully to infer the movement of organisms as they respond to their environments. Son et al. describe the use of Recurrent SOM for the detection of the movement patterns of Lumbriculus variegatus in response to the introduction of heavy metals to their environment. (52) The body of the blackworm was divided into 12 segments and the resultant 11 angles used to train the SOM to detect changes in body shape corresponding to changes in environmental toxicology. Liu et al. used SOM and Hidden Markov Model (HMM) to reveal state changes in the behaviour of Daphnia magna, a water flea, in response to the introduction of a pesticide Diazinon to their water environment. (53) The movement tracks of the D.magna were used to train the SOM resulting in several states. The speed changes associated with these states were verified by the HMM to show the effects of the introduction of toxic chemicals on movement behaviour.

The work of Son et al and Lui et al provides a model for the relationship between movement input and output that is relevant to this study. If considered as a self-contained system, the relationship between the organism and its environment is expressed via movement behaviour. Toxicity levels are a factor beyond the direct control of the individual organisms, and the organisms are reliant on a form of conflicted benevolence from humans in their struggle for survival. Changes in their movement behaviour tell a story of the organisms changing environment. The organisms are only viewed by the neural network in terms of their movement data - the localised movement of L. variagatus in Son et al. and the movement in space of D. magna in Liu et al. This ability for the
neural network to infer the relationship of the organism to its environment from the relatively simple movement data is a trait that could be of potential use for the agent in discerning pertinent traits within the dancer’s movement data stream.

Han and Cho made use of recurrent SOM and Markov Model for the prediction of user’s movement for location-based services. (54) GPS data of user locations were modelled using the SOM and Markov Model combination in order to predict the movement trajectories of users between different locations. This is useful especially in mobile applications for predicting the services a user will need in the present as well as in the future, as the user moves from site to site. The GPS data comprised relatively simple two-dimensional points representing the position of the user over time.

There is a large body of research concerning gesture recognition, much of it concerned with gestures of the hands. Caridakis et al describe the use of Self-Organizing Feature Maps and Markov Models for hand trajectory-based gesture recognition. (48) Movement trajectories derived from a video feed were passed through a SOM trained on a dataset of gesture points. The sequence of Best Matching Units (BMU) are then passed to the Hidden Markov Model to classify which class the gesture belongs to. These methods use visual tracking of features via video feed, extracting 2 dimensional or segment angles and passing these to the SOM for training. Caridakis et al used the 2 dimensional pathway of the hand to describe hand gestures. Similarly, Liu et al tracked the 2 dimensional path of D magna over time. Hidden Markov Models proved effective in movement classification when used in conjunction with ANN in the work of Liu and Caridakis.

Figure 32. Reproduction 2. 2011. Squaretangle. Screen Space Gallery Melbourne. The work used a spatialised soundscape enabled by Adam Nash. As the entities swim forward to greet new-coming human members of their species, their songs naturally came to the foreground and expanded into the space. When more than one new human visitor entered, multiple groups of different species might approach their respective new members and the space would be filled with the swirling melodies of the collective groupings. In a sense the form of the sound is emergent from the rules governing the behaviour of the entities in relation to the presence of human species members entering the space. The rules do not govern the specifics of how and when the entities will sing, they merely have their peculiar songs and go about their normal activities. The soundscape is a product of their behaviour.
Lee and Kim used a Hidden Markov Threshold Model to recognise gestures of the hand extracted from video feeds. The gestures comprised the pathway of the hand as it described particular spatial patterns. The gestures were designed to be used to control a powerpoint presentation on a computer. Ten gestures were defined for the ten most frequently use commands. The HMM proved successful in gesture segmentation, determining when a gesture began and finished, and in gesture classification.

As shown above, Hidden Markov Models feature prominently in the area of movement detection and human gesture recognition. While the gestures described in most research are relatively simple compared to the complex movement of a contemporary dancer, the combination of SOM and HMM have shown great promise as a foundation to begin developing the agents cognitive capabilities. However one of the main questions to be considered in their implementation was whether they would be able to handle the relatively large amounts of data found in the full-body motion capture stream as opposed to tracking a single point such as the hand.

Fdili Alaoui et al. took a different approach to gesture recognition, using movement quality rather than specific movements as the mode of interaction. Hand gestures captured with a Kinect sensor were clustered into three qualities: breathing, expanding and reducing. When participant’s gestures matched one of the three qualities, projected light responded appropriately to the quality of the gestures. This work provides an effective means of movement classification. However, in this
In my project, I was interested in using the movement in its source format, so as to generate a common means of exchange between performer and agent. Classification into specific movement qualities, as achieved by Alaoui et al. in this context, would not necessarily have facilitated a real-time exchange between dancer and agent, since dancers do not necessarily classify movement as part of the process of responding to it. (56) To the contrary, in most improvisational practices, movement responses occur much more quickly, and via much more nuanced movement exchanges, than would be possible if classifying and consciously naming a movement quality were made an integral part of the process.

Souza et al. used Support Vector Machines and Hidden Conditional Random Fields to learn and classify signed language from Brazilian Sign Language. (57) Both face and hand movements were incorporated into the system that could freely recognise a finite vocabulary of words. Again, while this approach provides an excellent basis for linguistic communication, it does not provide an appropriate method for enabling a movement-based exchange that is not predicated on pre-existing movement ‘vocabulary’.

**Movement Generation**

The examples described above investigated the potential for using neural networks for recognising behaviour in the test subjects. SOM in particular showed great potential for the recognition of both movements of organisms in response to environmental changes and human hand gestures. My investigation had the further requirement of providing a means by which the agent could respond to
the dancer in a human-readable manner. I was also interested in giving similar visual weight to the movement of dancer and agent, as much as possible treating them as equals in performance. This was an aesthetic choice as much as a functional one. To enable the work to function as an embodiment of real-time performer/agent exchange, I needed to generate a common modality through which the actions of performer and agent could be seen simultaneously, and perceived as part of a single work. To this end I decided to give both dancer and agent humanoid avatars so that their movement could be read in like terms. This would also allow the dancer to view the agent’s movement in familiar terms so as to allow her to respond as she would to another human dancer.

A form of SOM was used by Pierris and Dahl to allow a humanoid robot to learn and reproduce simple gestures. (58) An expert trainer manipulated the arm of a humanoid NAO robot to teach it the “looking at a wristwatch” gesture. The gesture was isolated to one arm of the robot. The robot was able to learn the gesture and reproduce it thanks to the trained SOM. This research was concerned with the development of biologically inspired robots that can learn and hence, adapt to their surroundings in a more flexible manner than robots with strictly programmed behaviours. While the SOM proved effective in training the robot’s arm, the system required a human manipulator to enable the learning process and was limited to the one arm. My system presumed the agent could learn unaided from the motion data and would learn full-body motions.

A robotic project with a different take on movement reproduction was undertaken by Masuda et al. (58) A humanoid robot was programmed with 40 whole-body movements, ten of each associated...
(according to the researchers) with four different emotions, pleasure, anger, sadness and relaxed. Human subjects were asked to categorise the movements according to the emotion they associated with the movements. An interesting feature of this study was the use of Laban Movement Analysis to form the feature set used to categorise the original movements. A high correlation was found (> 85%) between the emotions estimated by the human observers and the emotion estimated by the Laban Feature Set extraction. This research brings up interesting concerns surrounding human-agent interaction, the role of emotions and non-verbal behaviour in human agent communication, and self-perception in robots and other agents. However the movements used were very simple, such as raising both arms over the head and then lowering them. The movement used in my research required very complex, idiosyncratic, full-body movements, devised by a highly proficient contemporary dancer.

To create a system that linked human and agent in movement terms, I needed to generate an approximation of human movement. Generation of human-like movement has been approached with many methods. Brand and Hertzmann used a Stylistic Hidden Markov Model (SHMM) to create a style machine that could learn motion patterns from complex dance motion capture sequences. The style machine could synthesise new interpolations and extrapolations of the learnt movement styles. Style was learnt separately from the actual movement allowing, for example, an expert mover’s style to enhance the movement of a novice mover. However unlike Brand and Hertzmann, I did not attempt to distinguish movement structure from style. This was in keeping with a contemporary dance philosophy in which dance movements are understood to create meaning.
via the construction of ‘states of the body’ rather than via spatial patterns and gestures alone, and a distinction between structure and style does not arise. (60) Rather than attempting to create a ‘style generator’, I aimed to give the agent the means to respond in an aesthetically appropriate manner within the context of a contemporary dance performance. Variability in movement style was less important for me than the functional exchange between the dancer and agent in performance.

Hsu et al. used a learning style translation model based on time warping to be able to modify a library of walking motion clips to effect new stylistic interpretations of the recorded motions. (61) Taylor & Hinton used a Conditional Restricted Boltzmann Machine to learn movement style from a library of walking clips, being able to generate similar movement with variable styles based on the styles inherent in the original library. (62) These examples illustrate the potential for using a learning model for the generation of human-like movement.

These projects demonstrate successful techniques in movement recognition or generation. I wished to extend this capability by using SOM for both movement recognition and generation. I was interested to see if a single learning structure could allow the agent to perform both movement recognition and synthesis functions.
Agents in Dance Performance

In dance technology performances, a blend of movement recognition and generation has usually been considered necessary so that the machine behaviour can be aesthetically integrated into an overall performance event. While such systems typically employ some elements of both recognition and generation, given the focus on producing a satisfying performative outcome, movement generation has tended to be emphasised over movement recognition. In Apparition, Klaus Obermaier and the Ars Electronica Futurelab were concerned with the system as performance partner and the immersive kinetic space. (63) While remaining responsive to the dancer, the system behaves independently, having its own physics model. Thus the relationship is not direct but can change according to the desired interplay between the many layers of the performance. Frieder Weiss’ dance collaborations including those with Chunky Move also achieved a high correlation between projected environment and dancer.

The relationship of the digital pixel environment to the performer varies from being an illustrative extended motion of their movement, a visual expression of internal states, and also a self-contained animated habitat. (64)

These works used video camera based tracking to enable the system to follow and respond to the movements of the dancer, however the level of correspondence could change dramatically in order to achieve degrees of emotional or aesthetic projection. While these works aspired to, and succeeded in, creating a close correspondence between the dancer and projected environment, they weren’t predicated on an autonomous, artificially intelligent agent for their realisation.

Figure 39. Reproduction 4. 2013. Squaretangle. Adam Nash seen here being sung to by a group of entities of his species. His attention to the entities and movement in relation to them elicit changes to their sung responses. Adam and the entities (and visitors when they interact) enter into a performing partnership from which the soundscape emerges.
Seminal work in the area of performing agents has been undertaken by Marc Downie and Open
Ended Group. (65, 66) In *how long does the subject linger on the edge of the volume* …,(2005) a
software agent constantly updated its own visual expression according to the relationship of the
dancers in the performance space, who were tracked by motion capture. The geometric
appearance of the "Triangle Agent" was influenced by a physics model. The *Triangle Agent*
continuously updated its own image based on the spatial relationships of the dancers. It sketched
out and re-appraised these relationships from present and past. *The creature is a physically
simulated body in an environment with gravity and ground. If the creature is unbalanced, it falls
over, dragging its annotations with it, until it finds a new equilibrium and continues.*(66)

How long… is possibly the dance performance work that most closely embodies my aspirations for
the performing agent. It uses motion capture to track the positions of the dancers on stage and the
agent uses this data to construct its own body. Whereas *how long*… did not make use of skeletal
representations for the agent, I have chosen to use the same internal skeletal representation for
both the dancer and agent so as to give them a measure of equality in how they are viewed and
represented. The dancer and agent’s movement streams can then be used somewhat
interchangeably. Unlike Downie’s work where the projected image was the agent representation, in
this work the agent is given the same form as the dancer and the projections are a representation
of their combined relationship. This is in keeping with the objective of this research to give the
dancer and agent equal screen status in performance.
Another intriguing work, *Becoming* (2013) by Marc Downie and Nick Rothwell, sees a live Artificially Intelligent agent that tries to learn the movements present in an iconic 1980’s science-fiction film. The agent features an abstract body that draws on the movement, colour and geometry in the film. The agent was used in rehearsal by choreographer Wayne McGregor as an “eleventh dancer” to provide choreographic inspiration for a new work. The agent in *Becoming* was presented as an installation as well as for inspiration for the creation of dance work. However it wasn’t designed as a performing agent capable of interacting with a live dancer, something my agent would be required to do.

The above-mentioned works seek to do more than use the projected environments to further dramatic intent. They envisage an environment as a partner, capable of agency. Indeed OpenEndedGroup refer to many of their digital manifestations as agents. This body of work also sees a move towards the use of learning models for the agents, and the development of context and history in the digital components of the performance. The chosen visual manifestation of these works is often geometric in nature, fulfilling multiple roles as avatar, projected setting and human augmentation. While these works are inspirational for this project, they don’t allow full-body recognition of complex dance movement in performance, which my performing agent would require. My performing agent also needed to have a bodily representation commensurate with the dancer’s in order for the visual manifestations of dancer and agent to carry similar weight. Where the projected imagery is abstracted, it needed to use the movement of both agent and dancer and indeed their relationship, to inform the visualisations, rather than the agent visualising the dancer’s
movement, to realise my goal of creating a truly shared, interactive performance between dancer and agent.
Experiments

Self-Organising Maps

The desire to use Artificial Neural Networks in this research was prompted by the desire to create as ‘natural’ a performance ecosystem as possible between the agent and the human performer. By natural I mean that the creative process should be influenced by a comparable relationship between two human dancers. Of course this can’t be exactly emulated. However I envisaged that attempting to approximate this relationship would have benefits for both the agent and dancer. The agent could be rapidly developed from learning from the dancer, the dancer could use her experience of particular performance development structures to imagine how she could interact with the agent. I was attracted to learning models that could enable the agent to engage in autonomous sharing of movement material with the dancer. The type of Artificial Neural Network I began with and continued to use throughout the research was a Self-Organising Map (SOM), also known as a Kohonen Feature map, named after Teuvo Kohonen who first described them. (50, 51) SOM have a number of unique properties that make them useful for data visualisation. The most frequent use of SOM is as a dimension reducer allowing high-dimensional data to be visualised in lower dimensions, usually in two dimensions. For example, the motion data used in this research typically has 79 or more dimensions, equating to the number of floating-point numbers representing the joint rotations and global position of the dancer. In the tests using the

Figure 40. Dancer Steph Hutchison with reflective markers attached to her body and costume for use with the Motion Analysis motion capture system. Normally the dancer would wear a black lycra suit to which the markers would be attached. Here the dancer is in a more minimal costume in order for her movements to be more clearly seen by the live audience. Her movement data is being streamed to the Unity game engine to visualise her movement as the blue lightning character.
raw motion capture data formats, the Motion Analysis system (Figure 41) contained 161 dimensions to each frame of data and the Optitrack system (Figure 40) contained 99 dimensions. In the latter experiments I standardised the data format to 79 dimensions. This was based on the final skeleton having 19 joints with rotations represented by quaternions. Each joint rotation had four dimensions to represent the quaternion rotation and then the skeleton had three further dimensions to describe its global XYZ position, a total of 79 dimensions. A SOM is able to process this data and display key relationships within the data using a two dimensional map or grid. SOM is thus a form of data compression that allows complex datasets to be more easily visualised allowing the key relationships within the data to be better understood. SOM are also referred to as topology-preserving as topological relationships in the original data are maintained in the resultant map. SOM are also commonly used for clustering of data, a classic example being the clustering of iris data into categories according to the length and width of the sepal and petal of the iris flowers. A SOM is able to cluster irises with similar properties into distinct classes, aiding identification. (68)

SOM are also known for their associative memory properties, the weights of the neurons progressively become highly correlated to the input data. When new data with a similar feature space is introduced to the trained SOM, the inputs tend to become associated with the neurons with the most similarity. So a new movement posture will stimulate the neuron in the trained SOM that is the closest match, effecting a type of short-term, unlabelled memory. It is this associative memory characteristic of SOM that is used primarily in this research.
SOM Algorithm

Each neuron in the SOM has a series of internal weights corresponding to the number of components in the input data. For example, when representing colour data, a SOM has 3 weights, each representing red, green or blue colour components. Each neuron has an x and a y coordinate within the rectangular map allowing them to be uniquely identified. Once the neuron weights are initialised with random numbers, the procedure is as follows:

**SOM Algorithm**

1) An input vector comprising a frame of motion data (79 numbers) is presented to the network. This can be randomly chosen or could also be the next frame of data in the input sequence, as is the case with motion data.

2) Each neuron in the network is compared with the input and the closest match, determined by the closeness of it weights to the input values, is chosen, often designated the Best Matching Unit (BMU). Euclidean distance is a common measure of closeness.

3) The weights of the BMU are adjusted to become closer to the input data and neurons within a diminishing radius are also adjusted to a lesser degree the further their distance from the BMU.

4) These steps are repeated for the number of iterations chosen.
Euclidean distance as a measure of the closeness between a neuron’s weights and the input vector is given as

\[ \text{Dist} = \sqrt{\sum_{i=0}^{n} (V_i - W_i)^2} \]

where \( V \) is the current input Vector and \( W \) is the current neuron’s Weight vector. Simply put, this says for each piece of input data from 0 (the first) to \( n \) (the last), subtract the current neuron’s Weights from the input Vector’s values, Square the result and then take the square root as the final result. The neuron with the lowest final result becomes the Best Matching Unit. The same equation can be applied to very simple data with only a few dimensions, as well as high dimensional data with tens or hundreds of dimensions.

For example, if the input colour is blue (0, 0, 1) as RGB, and the current neuron weights are (0.1, 0.4, 0.5) this is described by the following code snippet:

\[
\text{distance} = \sqrt{(0 - 0.1)^2 + (0 - 0.4)^2 + (1 - 0.5)^2}
\]
\[
= \sqrt{0.1 + 0.16 + 0.25}
\]
\[
= \sqrt{0.51}
\]
\[
\text{distance} = 0.714
\]
In this example there are three values in the input vector and neuron weights, however if there are more dimensions, such as in the motion data streams, the additional values are simply added to the top line above and included to provide the final value.

The radius of the neighborhood determining which neurons around the BMU also have their weights adjusted is determined by the following equation:

Equation 2: \( \sigma(t) = \sigma_0 \exp \left[ -\frac{t}{\lambda} \right] \)

where \( \sigma_0 \) denotes the width of the network at time \( t_0 \), \( \lambda \) is a time constant and \( t \) is the current time or loop iteration. This is an exponential decay function so the neighborhood surrounding the BMU that is affected gradually reduces until only the BMU is changed by the input values.

The neurons within the neighborhood radius in the SOM algorithm 3 above have their weights adjusted according to the following formulae where \( W \) is the neuron weight, \( t \) is the time step, \( \theta \) influences the amount the weight is adjusted by the learning rate, diminishing with the distance from the best matching unit, and \( L \) is the Learning rate, the amount the weight is ultimately adjusted, which is also a diminishing amount of the difference between the neuron’s weights and the input values.

Equation 3: \( W(t + 1) = W(t) + \theta(t)L(t)(V(t) - W(t)) \)

\( \theta(t) \) is a neighborhood function defined by:

Equation 4. where \( \text{dist} \) is the distance of the neuron from the BMU and \( \sigma \) is the width of the neighbourhood function from Equation 2.

\( \theta(t) = \exp \left[ -\frac{\text{dist}^2}{2\sigma^2(t)} \right] \)
Equation 4: \( \theta(t) = \exp \left[ -\frac{\text{dist}^2}{2\sigma^2(t)} \right] \)

where dist is the distance of the neuron from the BMU and \( \sigma \) is the width of the neighborhood function from Equation 2.

For an excellent introduction to the SOM algorithm see the online article by Mat Buckland from which I have based my examples. (69)

**Experiment 1: Creating associative memory of movement performance using Self-Organising Maps**

Self-Organising Maps have been used successfully to learn characteristics of movement in non-human organisms and for simple human gestures as described in chapter 1. I decided to test them with some complex motion-captured movement. I was unsure of what the exact results would be as they use unsupervised learning and it would be up to the network to encapsulate the data within its neural network. There was a possibility that unsupervised learning would lead to nonsensical, or at least physically impossible, movement outcomes for the agent.

Sequences of movement were captured using both Motion Analysis and Optitrack optical motion capture systems to determine if the method was system independent. In essence the motion capture systems could be viewed as the sensory apparatus through which the system engages with its environment. Both systems used multiple cameras to record the positions of reflective markers attached to the body of the dancer. (Figures 44-46)
A standard markerset was used throughout the research in order to maintain consistency with the data output. 34 markers were used for the Optitrack and 40 markers for the Motion Analysis system. The different number of markers for the different systems was due to the current usage conventions. The 34 markerset is a preset template for the optitrack system that is built in to the control software and allows for rapid identification of the dancer. The 40 markerset used with the Motion Analysis system is a relatively simple markerset that has been proven to work for the types of captures used in this research. The reflective markers were attached with Velcro to a black lycra suit. The Optitrack system used 12 e250 cameras with a resolution of 800 x 800 pixels. The Motion Analysis system used 24 Eagle cameras with 4 megapixel resolution.

The motion data were produced to represent both limb position as defined by the 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. I did the first tests in the Matlab environment using the Neural Network Toolbox. Using Matlab proved to be a quick way of determining whether the basic premise of using SOM to learn and classify motion capture data was sound. The initial test SOM 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 neural network. For the Optitrack, each frame contained 34 marker positions 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. For the Motion Analysis system each frame consisted of either 40 markers (120 values) or a proprietary format of hierarchical local rotations comprising 161 values.

Each frame containing all of these values is presented to every neuron in the SOM. The one deemed to have the closest match (Best Matching Unit, BMU) is the winner and the map is adjusted accordingly as described previously. 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 is in the form describing the number of positive classifications or hits each neuron achieved. (Figure 47-48)

Figure 47 shows the number of frames each neuron gathered as a class or group with similar data patterns. The patterns here, being frames of motion capture data, could be considered dynamic postures extracted from the movement sequence. To test the trained network, motion capture 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. (Figure 48) 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 motion capture 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 and provided an initial indication of the validity of the classification technique.

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 possible within this first testing stage.

One of the defining properties of the SOM is its propensity for aligning itself with the input — the neuron weights gradually move to match the inputs. This characteristic is sometimes referred to as topographically-preserving and points to the close (some would say inseparable) (39) relationship between the SOM’s states and their environment.
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.

This first experiment demonstrates a model for the imprinting of environmental influences (movement data in this case) upon the network of neurons. However the SOM is not like a linear recording of the movement, as are the original motion capture data. Potential movement postures are clustered around the network. In order for movement to be synthesised from the neural network, neurons must be stimulated in sequence. The standard SOM has no temporal aspect to it, so if the neural network is to learn appropriate links between neurons, the SOM will need to be extended in some form.

The network no longer contains the original temporal trajectory of the movement within the weights of the artificial neurons. There is the potential to trace a similar trajectory, resulting in a close facsimile of the original movement, but not an exact replica. There is however the added potential for a multitude of new trajectories through the learnt data space. The introduction of new data to the network triggers the neuron with the closest matching weights effectively forming a current state according to the external influence on the neural network. From a potentially creative perspective, this capacity for re-association may hold the key to the software agent’s ability to

Figure 49. SOM Weight Positions. A 2-dimensional representation of the first two vectors of the 79 inputs. The green line is the input data and the grey points are the neurons with connections in red. The grey neurons have moved to closely match the input data.
The fundamental feature of neural network models is that they treat cognition and behaviour as resulting from learning about the environment, thus creating reflections of regularity in the environment in our behaviour. Neural network models are also notable for their treatment of cognition as resulting from the massively parallel operation of simple processes, a characteristic that connectionist modellers point out is also a fundamental feature of the brain.\(^\text{70}\)\(^\text{p.288}\)

This cognitive modelling perspective, though alluding to higher level behaviour, is descriptive of the environmental pressure exerted on the neurons in the formation of their weights and the resultant reflection of environmental regularity in the neurons as shown in (Figure 49).

The SOM proved successful at encapsulating movement into its network. The next step was to incorporate the SOM into a software environment in which an agent-based approach to movement learning and generation could be tested.

**Unity Game Engine**

I had been using the Unity 3D Game Engine for a lot of my visualisation and performance work and was keen to utilise it again for this work. I had already developed plugins for the motion capture streams and had the capability to use the data for animating skeletal characters and other 3D assets in real-time. I also had a stereo projection solution for Unity and wanted to make use of these options in the live performances. Rather than attempting to develop these capabilities from scratch or using another platform, it was much easier to use this existing pipeline for the development of the performances. The main requirement was to incorporate the neural network learning into Unity in order to develop the agent’s capabilities. Before launching into a full

Figure 50. SOM Colour Clustering. A test of the SOM algorithm in Unity. A 100 x 100 map of neurons is initialised with random weights representing Red Green and Blue (RGB) colour values, three vector weights per neuron.
movement based SOM in Unity I ran a few tests to see how it would work in Unity.

**Experiment 2: “Hello World” Colour clustering with SOM**

I implemented a SOM in Unity based on an example provided by Andrew Krillov. (71) The typical first test for a SOM is its use for clustering colours. This gives a good practical and conceptual view of the SOM’s capabilities for ordering and displaying data relationships. I implemented a standard SOM within the Unity game engine. Unless otherwise noted, all the following work was done within Unity. I created a 100 x 100 neuron network and initialised the neuron weights with random colours using Red Green Blue (RGB) values. (Figure 50) The SOM was then left to discover relationships within the network of colour data. Over the course of 100 iterations the SOM was able to cluster the colours into distinct regions of the map. (Figure 51) The artificial neural network successfully attempted to find patterns in the input colour data. The illustrations show the before and after of a gradual process of clustering colours from randomly scattered starting positions. The network has no concept of colour and no language with which to label the elements, however a distinct pattern emerges that is not reliant on language for coherence. As with my later experiments with movement phrases, no language or labelling was involved with their introduction to the network, yet the system was able to decipher coherent patterns within the...
movement streams. For the emerging entity, language is not a requirement for engagement. The SOM has no concept of “blue” yet its output reveals an emerging pattern that suggests awareness of that component. That the SOM was able to cluster the colours into appropriate groups when run inside Unity provided a convincing indication that a response to movement data might be possible.

**Experiment 3: Interactions with associative memory of movement.**

Having established a potential mechanism for acquiring memory fragments of performed movement in Experiment 1, and successfully incorporating SOM into Unity, the next step was embedding this potential for associative memory into a practical performance environment and testing interactions between the synthetic system’s remembered atoms and a live motion data stream. The significance of testing a live motion stream is that it tests the ability of the network solutions in a real-world situation in which unanticipated inputs can arise. Though the model tested in Experiment 1 showed promising results, the training and testing data were both pre-recorded and loaded from stored files and were therefore not as would be encountered within a live performance context. Experiment 2 investigated the artificial neural network’s ability to respond to incoming live streams of movement data based on what it had previously been exposed to.

The Unity game development environment was again chosen for this experiment as I had some modules already prepared for the visualisation of skeletal characters, essentially enabling the data
to animate a synthetic body or avatar, allowing a more intuitive reading of the movement. I had previously used Unity to animate digital characters directly from live motion data streams, and as such it provides a comparison between the characters representing the dancer’s live motion stream and those representing the responses from the synthetic memory. Unity as a software platform is also easily extended using the C# and C++ programming languages, and I consequently wrote modules to connect and receive streams from the motion capture systems, to animate the skeletal systems of virtual characters, to construct and implement Artificial Neural Networks such as the Self Organising Map and to enable rudimentary language processing which will be described later.

I also wrote software applications to capture the live movement streams from both the Optitrack and Motion Analysis systems and save them to a file. This saved data was then passed to a neural network created within Unity. I chose to use joint rotation data within this experiment as it could then be visualised as an animated skeletal character. The network created a map of the input data over 500 learning iterations. This trained network was then saved to disk so that it could be subsequently loaded into other Unity scenarios without the need for retraining. The learning process took between one hour and three days depending on the length of the motion stream, the number of neurons in the network and the number of iterations per learning epoch. (Figures 52-54) show a representation of 3 of the 99 dimensions of one movement sequence. The neural network (red) has become closely aligned with the input space (grey). While not visible in this 3 dimensional representation, close coupling of neuron weights to sensed data continues through to

![Figure 53. SOM training using motion capture data in Unity. The red neuron weights gradually become closer to the input weights.](image)
the other 96 dimensions not visualised here. In the case of the movement data captured with the Motion Analysis system, 161 dimensional data are passed to the network.

The learning phase resulted in a trained network for use in the next phase of the experiment, which was to introduce live streams of movement to the network to find out how it would respond. A typical way to do this is to record multiple takes of the same movement and use some for training and some for testing. The testing captures can still be streamed to the neural network and will appear exactly like a live stream from a dancer. Visualisation of the trained network as shown in (Figure 55) presents results consistent with the modelling performed in Experiment 1. As skeletal based characters are used to visualise the movement data, the data formats chosen represent the global position of the dancer and their joint rotations. These data can animate the character’s skeleton directly and are more appropriate for our current purposes than positional data only.

Once the SOM was trained with the motion data I introduced some pre-recorded motion streams to the network to see how they would respond. The pre-recorded stream has the same properties as a live stream from a dancer and the live and pre-recorded streams are indistinguishable to the neural network. Thus it is convenient to use pre-recorded streams to test the networks responses offline. When live motion data is introduced to the trained network, the neuron with the closest match is stimulated and its weights are used to animate a second character. (Figure 55) When using the pre-recorded training data as live input to the network the neurons which fired in
response often contained weights that were extremely close to the incoming data as seen by the similar resulting movement. This was not always the case, however even when the movements differed between the live data stream and the networks memory, the result was most often still discernible as plausible human movement. The movements were both anatomically reminiscent of human movement as well as have a resemblance to the movement the network was trained with. From an artistic perspective the firing of neurons with weights representing movement different to the incoming movement data is not necessarily a weakness, and may even be preferable as long as the movement is still plausible, since it demonstrates that the system does not simply mimic the incoming data stream. If the synthetic system were seen to only mimic the dancer then the rationale for such a system would be called into question as the original recorded data could more easily be used to directly animate a character. Responses that were new, unexpected and surprising were preferable as long as they seemed to still have some association to the live dancer’s movement stream. The image in figure 55 used the training set as the testing set, and yielded extremely close matches in movement responses. This provided a good test of the SOM’s potential capabilities for responding to the dancer’s movement data as the neurons are stimulated by the dancer’s data.

Initial testing of the trained neural network with live motion data demonstrated potential for a synthetic entity to respond to a live dancer’s motion stream and perform appropriate movements through its avatar’s body. In this experiment the movement data had a relatively limited vocabulary in dance terms, staying close to the original movement. The goal was to see how closely the agent
In this experiment the agent was essentially being asked to follow the dancer using what its neural network has learnt. As the incoming dancer’s movement data sequentially stimulates the agent’s neurons, the agent’s avatar is animated by the weights of those neurons. The success of the agent in performing this task prompted the use of this simple choreographic mechanism in later performances.

One potential problem with the training of the neural network is that of over-fitting whereby the network is able to classify exact matches from the incoming data, but is unable to successfully extrapolate to data that differs slightly from what is learnt. In performance this would result in the agent being unable to reasonably follow the dancer unless her movements were virtually identical to the original performance of the movement the dancer provided to the agent for learning. This would be problematic, since in any dance performance, and particularly one involving unstructured or semi-improvised performance, there is a high likelihood that the dancer’s movement will vary somewhat from performance to performance. The agent should therefore be able to cope with some variance in the dancer’s performance of the shared material. I experimented with the parameters of the training phase to ascertain the optimum values for network size and number of training iterations in order to maximise the network’s capability to cope with variance in the input streams. I found that 500 iterations was usually enough to allow the network to converge on a close encapsulation of the dancer’s movement. Network size worked best as a function of the total...
number of input frames of data. I had success using a network size with the number of neurons being approximately one third the size of the number of input frames of data. For example if the input movement contained 3000 frames of data, a two dimensional network of between 30 and 35 neurons for a total of 900 to 1,225 total neurons provided good results. It required some experimentation and I would tend towards the lower boundary of network size in order to reduce training time. Overall, known movements (those used in the learning phase) tended to trigger very similar movements within the neural network.

Standardised Data Structure

The previous experiment used the native data structures provided by the motion capture systems used. This provided a good basis for testing. However, for the proposed performances it would be better to have a shared, consistent format for both dancer and agent to avoid having to maintain different versions and to allow as much interoperability between the agent and dancer’s data as possible. I chose to standardise the dataset structure between the dancer and agent, and between the different motion capture systems used, which were Optitrack, Motion Analysis and Kinect. I used the same skeletal structure throughout the latter part of the research. This enabled the different systems to interact and for the dancer and agent’s avatars to be interchangeable in rehearsal and performance, allowing them to exchange roles in performance and for their data streams to be directly compared and equally influenced. The Skeleton structure is a simplified skeleton consisting of 23 bones and 19 joints. (Figure 57) This was enough to allow a reasonable amount of articulation for the avatars to express fluid movement while reducing the volume of data.
the agent was required to process. The simplified skeleton also sped up the training phase and
allowed good responsiveness in online testing and during real-time performances.

All the subsequent experiments were undertaken within Unity and the actual data format was
determined by the native format used by Unity to express position and rotation data. Positions are
described in XYZ format in metre scale and rotations are described using Quaternions. The
motion capture systems define positions in millimetres so scaling by 1000 is necessary to match
Unity’s scale. The Optitrack system also uses quaternions for rotations so no change was
necessary apart from stipulating that the data should be right-handed to suit unity’s rotation order.
The Motion Analysis system uses Euler angles to describe rotations, however once these are
applied to the skeleton in Unity, the required quaternions can be extracted. The motion capture
system data streams also contain header information that is not required for training the SOM and
is ignored.

Recording and Pre-processing the data for SOM training
I settled on a standard pipeline for recording and processing the data with which to train the SOM.
This was to ensure consistency in behaviour between experiments and to speed up the process
as once the programs to manage the data and train the SOM were written they were re-useable
for subsequent experiments. The process of preparing the data for the SOM began with recording
the required motion in the studio. The dancer and I discussed the concepts defining the dance
scenario with the agent and then she improvised around the theme until we were happy with the

Figure 58. Live capture and streaming of the dancer’s movement in Cortex.

Figure 59. The streamed motion data displayed on a humanoid avatar in Unity.
results. We then recorded the movement to disk, opting for a number of takes of each sequence. For the main studies and the performances we used the Motion Analysis system in the Motion.lab. The Motion.lab studio was much more amenable to performing as it is a larger space than the studio housing the Optitrack system and has a dance floor covering which is more favourable to the dancer. Once the required movement was captured it was cleaned using Cortex, the software provided by Motion Analysis as the main capture and processing software. (Figure 58) Cleaning involves removing any gaps in the marker data and making sure the markers are correctly labelled. The actual capture data is of the 40 markers attached to the dancer’s motion capture suit. A skeleton matching the 3D avatar’s skeleton is constructed within Cortex and centred within the 40 markers. The markers drive the skeleton within cortex and it is this skeleton’s data that is used to animate the 3D avatar within Unity. (Figure 59)

Once the marker data is clean and the skeleton is loaded in Cortex and animated by the marker data, the position and rotation information from the skeleton is streamed into Unity to animate the 3D avatar. I wrote a program in Unity to take the position of the hips and the joint rotations of the avatar and write them to a file. So that the movement wouldn’t be locked to a particular point in space, I used the change in position from the last frame of data rather than the global position of the avatar. This allows the agent to handle movement no matter where it is in the space. This spatial independence was important otherwise the agent would only recognise movement if it were performed in the same place as it was performed in the sequences it learnt from. The incremental

---

**Figure 60.** Rows of data used for SOM training. The first 3 numbers are XYZ positions, each subsequent 4 numbers are quaternion rotations.

**Figure 61.** Motion Analysis HTR data structure.
positions allowed greater generalisation of the movement independent of where in the space it was performed. The positions were written to file as XYZ, metre-base floating-point numbers and the rotations as four part quaternion descriptions, XXYZW. (Figure 60) These are both the default for Unity and meant that the agent could decode the data back onto the avatar without further translation of the data.

The resultant file contained the description of the motion data for the SOM training. Each line contained 79 entries, the first three being the incremental positions of the avatar and the remaining 76 being the four part quaternions for each of the nineteen joints. The length of the file was determined by the length of the original capture take. I used a set frame rate of 50 frames per second for all the experiments. This was important in order to keep the learning and performing of movement independent of the frame rate of different computers as the rendered frame rate in Unity can differ markedly between machines.

Once the movement streams had been written to file, the files were passed to the SOM for training. (Figure 62) I also implemented the SOM training into Unity. The training was slower in Unity than in a purpose written program. However I was not sure what components I might need in Unity for the performances and decided to implement all the modules into Unity so as to have access to them for further development. I was unsure what components would eventually be used and this also allowed components to be built upon in further iterations.

The SOM was trained over 500 iterations with a learning rate of 0.01. I used a two-dimensional...
map throughout the experiments. An input file of 3000 frames took around 5 hours to train. Once trained the SOM was saved to disk so that it could be loaded back into Unity at a later stage for use in testing.

Once the technical aspects of training the SOM along with the software modules were complete, I was able to start thinking more concertedly about the appropriate structures that would govern the relationship between the dancer and agent. Using the SOM as the basis for the agent’s memory and cognitive functioning, I turned to devising structures that would give rise to appropriate choreographic tasks for the dancer and agent.

**Creative Process**

The initial experiments were successful in suggesting that the use of Artificial Neural Networks and in particular Self-Organising Maps, would be a suitable structure as the basis for the performing agent’s learning and performing capabilities. I did have some pre-determined conceptions of the types of interactions that might be suitable as choreographic structures for the dancer and agent to use. However the experiments with SOM also suggested simple choreographic structures that could be used in the creation of the performances. I had some broad conceptions of the agent being able to recognise and respond to the dancer’s movement data. The previous experiment suggested the simple task of one dancer following the other as a

![Diagram](image-url)  
*Figure 64. Process for creative development of the agent learning for rehearsal and performance.*
possibility. This prompted me to start thinking more specifically about choreographic tasks that could inform the development of shared movement for the performances.

I wished to use a developmental process for the dance creation that would be familiar to two dancers working together. (Figure 64) This was in part a response to the readings into scaffolded and distributed cognition, suggesting that there was a strong support system for the agent’s development already in place in the form of the dancer’s honed dance experience. To exploit this I attempted to closely replicate a process that would allow the dancer to proceed on a familiar creative trajectory during the rehearsal process. The human dancer developed movement phrases through structured improvisation based on the kind of relationships that would be used in the performance; being followed by or following the agent, providing starting postures for the agent to begin a movement phrase with, making short movement gestures that the agent might recognise. These movements were passed on to the agent for learning. Once the agent had gone through the learning process, the dancer improvised with the agent to reacquaint herself with the movement and to investigate how the agent had assimilated her original movement and the relationships envisaged when developing the original movement material. We used a familiar rehearsal and performance process in order to embed not only the movement itself but also the relationship between dancer and agent into the development process. This allowed the dancer to follow a familiar creative trajectory, and supported the agent’s capability by providing a close association with the dancer.
Studies

The initial experiments used to develop and validate the usability of the system were undertaken using recorded motion data for training and testing. Because the testing data was streamed to the neural network it appeared just like a live stream from the dancer. The training and testing data were recorded in the same session so the skeleton structure and setup were known to be identical. In the next studies, a live dancer provided the motion data stream as I tested the capabilities of the SOM in a performance situation. The live movement stream from the dancer was used as input to the SOM, which was trained on data from a previous capture session. These ‘live’ sessions tested the plausibility of the choreographic structures employed and whether the SOM could generalise movement from different capture sessions. The main concern was that markers would not be placed in exactly the same position on the suit and that the skeleton would have small variations in how it is aligned to the markers in the new session.

Agent following the dancer from memory

The first live performance study consisted of the agent simply following the dancer by matching the dancer’s live movement to that which it had learnt. (Figures 66 - 68) I recorded both a known movement sequence and some improvised movement of the dancer. This tested the agent in following both set and improvised movement of the dancer. Once the agent’s SOM was trained, the dancer was suited up and her live movement stream presented to the SOM. The dancer’s live stream was first used to animate an avatar identical to the agent’s. The data from the dancer’s

Figure 66. The agent was able to follow the dancer successfully when the dancer worked within the range of movement she had passed to the agent to learn. Anything outside the vocabulary it had learnt caused more unpredictable behaviour.
avatar was presented to the agent’s SOM so that the data format would be identical to the data it was trained on. The incoming movement stream stimulated the neuron with the closest match and the agent performed this movement contained in its memory. The movement vocabulary was confined to what the dancer had shared with the agent. The dancer could perform the material that was presented to the agent and improvise around that shared movement material. The agent could also recombine postures it had learnt, but it did not create new movement postures. This is akin to two dancers sharing a known vocabulary of movement and performing it together. The agent’s movement was not exactly like the dancers as the learning process caused the neuron weights to gradually move towards the input movement rather than directly recording and replicating them. However, the responses were generally close and the movement of the agent was recognisably within the shared vocabulary. When the dancer improvised with the dance material the agent sometimes became less fluid, and there were recognisable differences between the dancer and agent, especially during transitions between movements. However, these uncertainties would tend to be present between human dancers if one was improvising and the other attempting to follow.

Generally the agent did a good job of following the dancer. Especially when the dancer was performing the known sequences, the agent was able to follow quite closely. The movements, even though not exactly like the dancer, were still recognisably form the same movement vocabulary and style. When the dancer was improvising around the shared movement material, the agent was still mostly in keeping with the dancer, however the agent’s movement appeared

![Figure 67. Agent (red avatar) following the dancer (white avatar).](image)
more tentative and jumped into somewhat strange postures at times.

Because the agent had learnt only the movement as captured by the motion capture system, there was no additional information other than the position and joint rotation information. There was also no physics information or data describing external influences. At times this lead to visually strange juxtapositions between the agent and dancer. The dancer might be in a standing position with a particular pose while the agent might be in a similar position with its feet off the floor and at a perilous angle, as if it was in the middle of a jump or off-balance manoeuvre. Because the agent was responding with the movement contained in the neuron stimulated by the dancer’s data, there was no other contextual information for the agent’s movement. The stimulated neuron activated the agent’s avatar without any filtering for the appropriateness of the movement. Considering there was no programmed internal filtering or external rules to guide the agent’s choice of posture, the SOM worked very well in the task of following the dancer. This experiment showed the SOM capable of giving the agent a rudimentary ability to respond appropriately to the dancer’s movement in a live setting.

**The agent learns to generate independent movement**

The next task I had for the agent was to be able to generate movement using the SOM. In the previous experiment the agent was able to perform basic recognition of the dancer’s movement. In this trial I was searching for a way for it to generate movement from what it had learnt.
The SOM used by the agent for memory has no inherent temporal information. Movement postures are clustered throughout the neural network without links to describe how they originally connected in time. I could see that by traversing the neural network it would be possible to generate movement by connecting the movements together. By moving from one neuron to another over time the postures would be connected to form a sequence of movement. The question was, how to do this in order to produce movement that looked plausibly human and in keeping with the style of the dancer’s original movement? I didn’t want the agent to generate random associations, but was interested in enabling the agent and dancer to have a cohesive, shared approach to the movement performance and style.

The problem seemed to be one of how to traverse the network in order to generate appropriate movement. When stimulated by the dancer’s live input, the agent was able to respond with a sequence of movements that was both plausible and in keeping with the dancer’s own movement. It therefore seemed that synthesising appropriate movement would also be possible if I could develop a system for traversing the map over time.

**Self-Organising Synaptic Map**

I eventually settled on the idea of using the concept of neural plasticity to help govern the neuronal behaviour of the neural network. The essence of this concept is that in the human brain, synaptic connections between neurons strengthen in response to increases in activation rates. I adopted the simple premise that as subsequent neurons fired, a link would be formed between those
neurons, and that this link would be strengthened according to the number of times the first neuron fired into the second neuron. I assumed this process would create pathways through the neural network that had varying strengths according to the firing frequency between the neurons. I called this variation of SOM a Self-Organising Synaptic Map (SOSM).

I created a second map to contain these synaptic connections. On the last iteration of training, when the SOM is in its final form, the input data is passed through the SOM one last time and the neural connections activated by the sequential Best Matching Neurons are embedded into the second map. Any single neuron will have a finite number of connecting neurons that it could potentially fire into, determined by the connections it has accumulated during the learning phase. (Figure 70) Some neurons might have only one connection to another neuron, some might have four or five. Many neurons have a link to themselves, especially if the neuron represents still movement such as standing on the spot or more static postures. This linking to itself allowed the agent to stay still at times. Most neurons will have at least two connections, and five was the maximum connections depending on the particular movement phrases learnt by the neural network.

In order to generate movement the agent could use the synaptic links in the SOSM to determine the pathway through the postures in the first map. The connections lead to plausible movement as they are developed through the learning process. As a neuron is activated, in turn animating the agent’s avatar, the neuron subsequently chooses a synaptic link to follow in order to stimulate the

Figure 70. Tracing temporal connections between neurons as they fire. This concept I modeled on synaptic plasticity. As the neurons fire a connection is made from one neuron to the next. A second map contains these connections. If a path from one neuron to another is fired more that once the strength or weight of that synaptic connection is strengthened accordingly.
next neuron. As the links in the SOSM are weighted according to the firing frequency determined by the original movement phrases, the movement emitted by this movement through the neural network is changeable according to the choice of which synaptic link is followed. Links with a higher weighting tend to be followed more often, or more easily than the weaker links. There are a number of ways the SOSM can be used to activate the first neural map. Influenced by reading into neural noise and the high degree of variability in neural spiking associated with neural firing, I tried using some Gaussian noise acting on the synaptic map to activate the links. (72-79) The idea was that neuronal noise plays an important role in the behavioural outcomes of biological neural networks and I was interested to test its influence on an artificial neural network. Similar to my use of artificial neural networks as a basis for the agent learning system, I was interested in the neural noise model because of its biological basis. The introduction of noise to the network led to some strange timing in the agent’s performance of the movement it generated. The noise introduced a random element to the choice of neuron firing sequence, which while still creating plausible connected movement, introduced long pauses interspersed with short rapid flicking between movements. I believe the neural noise approach could have application in the right circumstances, and could well do with further investigation. It is possible that the neural network used in this research is too small to benefit from this technique and the effects of the noise are too severe. The human brain has 100 billion neurons compared to the maximum of 10,000 used in this research.

The methods that eventually worked best were to use the weights of the synaptic map as a probability matrix for choosing the next neuron to fire. The stronger links tended to occur more
often and the resultant movement looked both plausible both in the manner in which movements connected and in the timing of the movements. An alternative that also worked well was to use the last known good link rather than the synaptic weighting. The last known good link is defined by the path to the last Best Matching Neuron that the current neuron connected with from one movement frame to the next. This produced movement with a little less variability, yet the movement was still recognisably in keeping with the style and quality of the dancer’s movement the agent had learnt.

**Agent and Dancer moving independently**

The second movement study attempted to have the agent and dancer performing together but generating independent movement. (Figure 71) The dancer was tracked by either the optical motion capture system or a Microsoft Kinect sensor and the movement represented by her avatar. The agent created continuous movement from its neural network, generated from the store of movement it had learnt from the dancer’s original movement. The movement was based on a series of phrases developed by the dancer through iterative improvisation and selection of most preferred movement motifs. The resultant movements were performed by the dancer, recorded, and then presented to the agent’s SOSM for learning and development of the neural and synaptic maps. Using the synaptic map to traverse the neural map produced continuous movement for the agent, which was juxtaposed with the dancer’s movement derived from her live movement data stream.

**Agent generating movement from movement seeds supplied by the dancer**
One of the desirable attributes of the agent was for it to have a continuous internal representation of the dancer’s movement. From its neural network the agent always has a best guess at what the dancer is doing, prompted by the incoming motion stream. In the next study the agent uses this current representation of the dancer as the starting point for generating a movement phrase from what it has learnt. The agent uses the mechanism for generating movement used in the previous experiment, however it does so in response to the dancer’s current movements at particular times. At certain intervals the agent looks at the dancer’s live movement data stream to see what she is doing. This stimulates particular neurons in the agent’s neural network, which in turn creates trajectories through the synaptic map of the SOSM, creating movement responses. The agent will continue on its own movement creation until it again looks at what the dancer is doing and begins on a new movement trajectory in response to what the dancer is performing. (Figure 72) The point at which the agent will look at the dancer for a new seed could be triggered by a few methods; voice prompts from the dancer, a hand-held trigger or keyboard control from an external operator. The movement it generates is still based on the shared choreography it has learnt. It doesn’t attempt to create any stylistically new movement, but attempts to generate movement on the basis of what the dancer is performing. The process is similar to any two dancers performing variations of known material together. The agent can begin performing a sequence when prompted by a voice command, at regular or random intervals or on cue through a keyboard or other input device. The agent will continue to generate material until the next command, when it notes the current movement of the dancer and uses it as the seed for generating the next phrase. The agent performs this task well, though it doesn’t have the subtlety of the dancer, it movement transitions
can look sharp, and it adheres closely to what it has learnt.

**Combining a Self-Organising Synaptic Map (SOSM) and Hidden Markov Model (HMM)**

The ability to have a moment-to-moment representation of the dancer is one thing, but such a reflexive condition is somewhat limited. The agent would have greater scope to respond to the dancer if it was able to recognise phrases of movement not just momentary postures. I applied the combination of the multi-layered synaptic SOSM and a Hidden Markov Model to see if it was able to recognise short movement phrases.

A Hidden Markov Model is a statistical method for modelling sequences of information and is particularly suited to modelling sequences of data for temporal pattern recognition. The data I used with the HMM was the sequence of neurons as they were stimulated by the dancer’s incoming movement data as she performed particular gestures. This created a sequence of events that the HMM could use to generate a model which could be subsequently used to recognise live movement gestures similar to those learnt.

A HMM is a probabilistic model that attempts to approximate a real-world situation based on its observed behaviour. Typically, sequences of observations are fed to the HMM. The HMM then tries to ‘learn’ from these sequences and create a behavioural model to fit the observations. Once it has created a model, if more sequences of observations are introduced to the HMM it will

$$p(x, y) = \prod_{t=1}^{T} p(y_t | y_{t-1}) p(x_t | y_t)$$

Equation 5. HMM where $x$ is the sequence of observations, $y$ is the sequence of states, $p(y_t | y_{t-1})$ is the probability of being in the state $y_t$ given the previous state was $y_{t-1}$, and $p(x_t | y_t)$ is the probability of encountering the observation $x_t$ if we are currently in state $y_t$. 

Chapter 2: Experiments
produce outputs that have some similarity to the past sequences. The model assumes there are a number of states that the sequence will pass through, and that these states will be unique to the particular sequence. We do not have to tell the HMM what the particular states are, only how many should be needed to model the observations. This is the Hidden part of the Hidden Markov Model—we do not tell the system what the states are, and we do not necessarily even know what the states are. The model works out what is suitable during the learning process. The Markov part of Hidden Markov Model holds that the current observation in the sequence will only be dependant on the observation immediately before it and not on any other observations. Thus a Markov Model attempts to generate the probabilities of moving from one state to another with potential future states only dependent on the current state. The probability of a sequence of observations occurring in relation to a sequence of states is described by the following equation:

\[
p(x, y) = \prod_{t=1}^{T} p(y_t | y_{t-1}) p(x_t | y_t)
\]

is the sequence of observations, \( y \) is the sequence of states, \( p(y_t | y_{t-1}) \) is the probability of being in the state \( y_t \) given the previous state was \( y_{t-1} \) and \( p(x_t | y_t) \) is the probability of encountering the observation \( x_t \) if we are currently in state \( y_t \).

For further description of Hidden Markov Model see Cesar de Souza’s article on HMM on which I have based my implementation. (80) The HMM generates two probability matrices, one for the probability of moving from one state to another, \( p(y_t | y_{t-1}) \), the Transition Probabilities, and another probability matrix for the probability of encountering particular observations when in a
given state, \( p(x_t|y_2) \), the Observation Probabilities. This can be summarised as:

Hidden Markov Model Summary: \( \lambda = (\pi, A, B) \)

where the model \( \lambda \) is defined by the total number of states \( n \), the matrix of Transition Probabilities \( A \), the matrix of Observation Probabilities \( B \), and the initial state probabilities describing the probability of starting in each of the different states in the model.

To create a sequence classifier using HMM, a HMM can be produced for each class of sequence and when we pass a new sequence of observations is passed to the trained HMMs they will each produce their own probabilities or likelihoods that the observations arose from the states within their model. The model with the highest likelihood would become the preferred choice among the given models and the class it represented would be the best candidate of those available with which to classify our new observation sequence.

Using SOSM and HMM for full-body gesture recognition

I wanted the agent to be able to recognise short discreet movement phrases when performed by the dancer, and be able to respond in an appropriate manner. The HMM looked like a likely candidate as it has been shown to work in the areas of speech, handwriting and hand gesture recognition. However its use for full-body gesture recognition would be a difficult challenge due to the relatively large volume of data contained in the motion capture stream.

Instead of using the HMM on its own to attempt gesture classification, I considered using HMM in
conjunction with the SOSM, as the combination seemed to me to be a viable option. (Figure 73)

To test whether the HMM would work for classifying short full-body movement phrases I worked with the dancer on developing eight short movement phrases or “gestures”. The eight gesture phrases were recorded using the motion capture system. I recorded 12 takes of each gesture, 8 for training the HMM and four for testing. The short phrases were approximately 2 – 3 seconds in length. (Figure 74, 75) The takes were cleaned and then two of the takes were played back through the SOSM pre-processing application in Unity, to create the files for the SOSM to train on. (Figure 76) These new recordings of the motion as applied to the dancer’s avatar were then passed to the SOSM for training. The data used to train and test the HMM consisted of the sequence of neurons (BMU) as they were stimulated by the incoming motion data. Each neuron was identified by an integer and the HMM operated on this integer sequence. I used a two second window at 50 frames per second resulting in sequences of 100 integers representing the path through the SOSM as triggered by the dancer’s movement gestures. Once the SOSM was trained eight of the takes were presented to the SOSM and the sequence of BMUs were recorded and passed to the HMM for learning. Thus there were eight sequences of observations for each of the eight movement gestures. A HMM was created for each of the gestures which can be considered as distinct classes. At the end of the HMM training I had eight HMMs, one for each class of gesture.

The test recordings of the sequences were then passed to the collection of HMMs, which...
attempted to classify the movement “gestures” according to the model developed on the training data. Each HMM class was presented with a test recording of the neuron sequence emitted from the SOSM and produced a likelihood that the sequence could have been produced by that HMM. The HMM with the highest likelihood was deemed the most probable classification and the sequence was assigned the class belonging to that HMM.

Using pre-recorded test data, the HMM was able to accurately classify the test movement sequences with 100 per cent accuracy, with the likelihoods of the movements belonging to their correct class being very high. This was an unexpectedly high result and proved the combination of SOSM and HMM to be extremely viable as a gesture recognition combination. The next task was to test the combination with live input in real-time from the dancer.

**Agent recognizing short dance “gestures”**

Given the good results obtained in the testing of SOSM and HMM for gesture recognition I proceeded to test with a live dancer. In this study the dancer performed the short movement sequences and the agent attempted to correctly recognise the phrases and reply with synthesised text associated with a name we had labelled the movement with. The agent was able to correctly recognise the movements and respond accordingly. (Figure 77) The success rate was 100% using test recordings and around 90% in live testing, although I did not run specific tests for recognition rates for the HMM as the main problem was not the method employed using SOSM / HMM, but that of segmentation, finding where the movement started and ended. In offline trials the system
was presented with sequences representing the core of the movements and no segmentation routines were required. In performance the movements must be picked out of the live stream accurately enough for the agent to correctly classify. As we were using a two second window, there was not a lot of margin for error. While the core movements themselves were distinct enough for the agent to classify, the transitional preparatory and ending movements contained gestural similarities, creating ambiguities which if found in the window made classification difficult for the agent.

With manual segmentation by a dancer or operator to tell the agent when the movement has finished, the classification rate was very high. Accurate automatic segmentation of the phrases is one the current areas of investigation and I hope the agent will soon be able to autonomously detect and recognise the dancer’s movement phrases more accurately. For now the agent relies on the dancer indicating when she has finished the phrase and the agent then acts to attempt classification.

The HMM was trained with labels for each movement sequence, adding a supervised flavour to this part of the learning. These labels were descriptive words associated with the dance “gesture” and the agent responded with synthesised spoken statements related to the label descriptions.

Overall the live tests were very positive with around 90% accuracy for gesture recognition of very complex full-body movements. When the agent was wrong in classifying the gesture, it was not necessarily problematic from a performative perspective as the occasional errors added a level of fallibility that could be construed as a “human” touch. Imperfect classifications also provided
possibilities for the dancer to add new types of responses in return, and this also added to the sense that the performance was a genuine exchange rather than reliant on pre-determined programming. I was confident that the gesture recognition could be developed into an interesting performance that would provide multiple challenges for the dancer and agent in their interactive co-creation of the performance.
Figure 77. Test results from the Hidden Markov Model Gesture Classifier. Using the test recordings of the dancer’s movement “gestures”, the HMM was able to correctly classify the movements with 100% accuracy. Four of each of the eight full-body dance phrases were presented to the HMM which was correct on all occasions.
Performance

The experiments carried out into using SOSM and HMM to develop the agent’s performance capabilities proved very positive. They showed a lot of potential in both testing and live tests for the agent to interact meaningfully with the dancer. They also pointed towards some potential choreographic structures that would draw on the inherent neural awareness the agent developed during the learning phase. I proceeded to adapt the relationship between the agent and dancer as a series of performances based on the studies described in the previous chapter. The performances took place in the Motion.lab at Deakin University, where the majority of the research took place. The performances took advantage of the Motion.lab’s 24 camera Motion Analysis motion capture system and the passive stereoscopic projection system. The performer, her motion-captured avatar and the agent’s avatar were juxtaposed in the space. The avatars of the dancer and agent were projected in stereo, providing a sense of spatial perspective and overlap between the three ‘performers’, and a visual context in which the agent’s avatar seemed to have an enhanced choreographic presence. The agent’s movements were similar to those of the performer and her motion capture driven avatar, but not exact, giving the agent’s avatar a sense of independence. The subtle differences in the agent’s movement compared with the other two figures made it apparent that this was not simply a motion capture driven visualization of the performer’s movement, but a form of movement response.

I then visualised the relationship between the agent and dancer in various ways. I attempted to find
visualization environments that were somewhat indicative of the current relationship between the
performers, however the final outcomes were determined by artistic choices and I chose visual and
aural qualities that would have been engaging between any two dancers.

**Grevillea Crystalis Incarnadine**

Grevillea *Crystalis Incarnadine* began with the agent using its ability to follow the dancer using
movement it had previously learnt. I re-recorded the dancer performing both a one-minute set
phrase of material and some improvisations based on the material. The agent was given this dance
material to learn and generated a SOSM to encapsulate the data. There were 4,421 frames of data
equating to 85 seconds of material. Each frame of data contained 79 vectors, three for position and
76 for rotation of the 19 joints. The neural network size was 50 x 50 for a total of 2500 neurons.
Training occurred over 500 iterations with a learning rate of 0.3 and an initial learning radius of 3.
The dancer’s live motion capture data was streamed from the Cortex software into the Unity
application using the UDP protocol. Cortex has a built-in function to automatically attempt to fill in
missing markers in the case of occlusions which helped maintain a stable data stream for the
agent. At times enough markers would become unidentified that the skeleton representation would
lose parts of limbs and the agent would have incomplete data on which to operate. When this
happened the agent would still attempt to find the closest match from what it had learnt so these
lapses in data were not problematic in a performative sense.

The dancer performed both set dance phrases and semi-improvised phrases based on the

![Figure 79. Grevillea *Crystalis Incarnadine*. The closeness of the agent in being able to follow the dancer is visualised as a crystal flower. The differences in their respective joint positions are represented by the length of a petal for each joint.](image)
movement previously shared with the agent. The dancer’s data was used to directly animate the dancer’s humanoid avatar and the position and joint rotations of this avatar were used as the data stream for the agent to act upon. The agent’s neural network would be stimulated by this data and the neuron with the highest activation would animate the agent’s avatar with its internal weights. There are no programmed instructions for the agent to follow the dancer’s movements. The learning process and the latent potential contained in the agent’s neural network naturally lead it to be influenced by the incoming dancer’s movement and to follow as closely as possible.

As in the experimental version of the following structure, the agent was best able to smoothly follow the dancer when she was performing the known material in a known sequence. (Figure 78) When the dancer improvised the agent would look less sure and it movements would tend to deviate more from the dancer’s. This is not unexpected as when improvising the transitions between known movements are unrehearsed and the agent must adapt as best it can. This would be the same with another live dancer trying to follow. Known material would be more easily followed, but improvised material would require more concentration to follow and the second dancer would likely be more hesitant in their attempts to keep up with the dancer.

Overall the agent did a very reasonable job of following the movements of the dancer and I was pleased with the relationship implied by such a simple choreographic structure. The structure suggested a potential visualisation of the relationship between the dancer and agent. Because the agent and dancer’s avatars shared the same skeletal structure a direct comparison between them
was possible. I began to imagine how I could visualise the closeness of the agent’s attempts to closely follow the dancer and began looking at a number of potential visualisations.

Ultimately I settled on a fairly straightforward analysis of the closeness of the agent’s skeleton to the dancer’s skeleton. This was probably the most simple and obvious mechanism available. However, it seemed quite appropriate to the performative context.

I imagined the joints of the avatars as the petals of a crystal flower. (Figures 79-82) The petals of the flower each represented a body part and at first grew or retracted according to the distance between the agent and dancer’s limbs. In this phase the petals were most recognisable as emanating from the changes in agent and dancer data. However I explored allowing the flower to continue to evolve along more artistic lines and I allowed the aesthetic form of the flower to take over, to some extent, from the functional visualisation. The petals began to twist as well as elongate in response to the differences between the dancer and agent’s avatars. The petals began to elongate until they were long spikes at which stage they began to curve and wave in response to the difference in rotations of each of the joints of both avatars. The colour of the petals also changed from hues of red through blue to green.

Steph continued to improvise around the movement shared with the dancer. When her movements more closely approached movements with which the agent was familiar, the flower contracted and uncurled. Movements the agent could not approximate closely caused the flower to extend and
flagellate. The movement became quite wild and writhing at times with the flower resembling a flailing crustacean. The performance ended with the dancer holding a rounded, enclosed shape with the flower forming an extended shape that also held its shape and revolved slowly. The dancer’s performance began to visualise a kind of meta-commentary on the relationship between dancer and movement, adding a reflexive layer to the initially straightforward representation of the dancer-agent relationship.

**Recognition**

Recognition (Figure 5) saw the agent and dancer generating simultaneous but different continuous streams of movement that were used by a morphing creature, a “blob”, to animate itself. (Figures 83, 84) The dancer used the live motion capture data to animate her avatar and the positions of the avatars hands, feet, hips and head in turn were used by the morphing creature to create its changing shape. The dancer improvised with the project image of the creature and I captured this movement using the motion capture system. This data was then processed and passed to the agent’s SOSM to learn. Once the learning phase was complete the agent was able to generate new movement from its SOSM using the technique developed of having the synaptic layer control the flow of movement postures to create movement for the agent’s avatar.

In the performance the dancer was tracked by a Kinect sensor and the projected morphing eye creature used the dancer’s movement to alter its form when the dancer was present. When the dancer left the stage area, the morphing eye used the agent’s movement instead. The agent and
dancer worked together to continuously provide movement to the morphing eye avatar, constructed using meta-blobs. The dancer alternated dancing in front of the projected creature with her back to the audience, and leaving the stage so the agent would take over providing movement to the creature.

Recognition was also presented as an interactive installation which uses a performing software agent to effect change in the virtual environment when there are no humans present. Recognition was exhibited at Cube 37, a glass fronted gallery in Frankston, Australia. The use of learned human movement allowed the software agent to quickly acquire the capability to stand in for a human when needed. The projected forms also used pictures of the dancer’s iris as a texture for its body, giving the installation a uniquely organic signature. (Figure 85) Borrowing appropriate material from a human to quickly generate capacity for the agent was one of the features of Recognition.

The installation and performance versions were essentially the same structurally, the only difference is the familiarity of the dancer with the system and the movement choices available to the dancer due to her flexibility and experience. Recognition focused on the agent’s ability to generate movement from what is had learnt and perform this in parallel with a human participant in order to maintain a constant flow of movement data for the installation.

The Cube 37 gallery had a glass front onto which imagery was rear projected, allowing it to be viewed by passing pedestrians and people in cars. A Kinect sensor behind the glass screen tracked
people’s movement on the street in front of the gallery. The Kinect data was used to extract a skeleton representation of the pedestrian’s movement, or at times, the dancer’s movement, and this data was used to animate an avatar which was in the background and unseen by the participants. The joint positions of the unseen avatar were used by the visible morphing eye to change its shape accordingly. There was also another unseen avatar representing the performing agent. When no humans were present the visible eye took its movement data from the agent’s avatar instead. Thus the avatar representing the live human and the avatar representing the performing agent worked together to continuously provide movement data to the morphing eye. (Figure 86)

When a person entered the area in front of the gallery they faced a giant morphing shape that looked out at them from the gallery window. It changed according to their movement, fluidly transforming like liquid or molten metal. When it solidified into a single shape it appeared like a giant eye, casting its gaze over the pedestrian. Depending on the movement of the pedestrian, it could break into smaller parts or components, leading to extremely varied morphology. It invited improvised participation, either from passers-by or at specific times, from the dancer who provided the agent’s movement. It was meant to be playful and engage both participants and onlookers who could appreciate the myriad forms brought about by the participant’s movements. (Figures 87, 88)

When there were no human participants in front of the gallery, the great eye changes in colour and texture from that borrowed from my own iris, indicating that the agent I created was interacting with the avatar, to that of the dancer, Steph, indicating that she was the one engaging with the system.
The movement behind its morphing forms also changed to the movement of the software agent, which had been learnt from Steph’s improvisations. Thus, people in passing cars and pedestrians across the street were able to witness the continued dance of the morphing eye. The main avatar’s body was produced using a marching cubes algorithm and tables courtesy of Paul Bourke. This algorithm allowed the main avatar to constantly reform according to the movement data from the agent or live human.

The soundtrack for the performance version made use of echocardiogram recordings of the heart. This was prompted by the idea of borrowing human data to augment the agent’s capabilities. This issue is central to the concept of the agent-dancer performance paradigm, since the agent is, in some senses, ‘borrowing’ the dancer’s movement data, and hence, in some sense, her embodied experience. Machine learning and data-mining techniques are being increasingly applied to personal human data, and this process is often motivated by the desire to customise services to the person. This is a benevolent interpretation of these activities. Another interpretation is that it can be constituted as a form of identity theft if enough information associated with a person is ‘borrowed’ that the system can stand in as a proxy for the human. In a sense, Recognition is a form of consensual identity theft, since the agent has learnt to move from Steph, has borrowed its visual appearance from Steph’s identifying features (iris scan) and, in the context of the performance and installation, is able to stand in as proxy for her when she is not present. Thus Recognition presents an enactment of identity theft that invites interpretation as to the nature and implications of this theft, which could be construed as either positive or negative, depending on the context.
Instrumental

"Genetic activities and neural mechanisms themselves possess remarkable plasticity, awaiting sociocultural contexts to exert reciprocal influences on them and to be the "coauthors" of mind and behaviour." - Li (31)

One of the most engaging aspects of watching skilled improvisers perform together is to see them cueing each other, co-creating a complex web of possibilities from small hints and fragments of their shared or individual pasts. Whereas *Grevillea* used the momentary recognition capabilities of the SOSM and *Recognition* used the synapse-prompted generation of movement by the SOSM, *Instrumental* combined these processes to allow the agent to both recognise what the dancer is doing at particular times in the performance and to use this recognised movement as the starting point from which to launch into a new thread of movement. (Figure 89)

The dancer was tracked by the optical motion capture system and her movement data streamed to the application written in Unity. The SOSM used in *Instrumental* was larger than the other performance pieces as the movement vocabulary is more extensive. The network was 120 x 120 neurons for a total of 14,400 neurons. The number of frames of data was 29,995. The other parameters of the SOSM training remained the same.

The dancer improvised around the shared vocabulary of movement and the agent’s SOSM would be stimulated by what the dancer was doing at certain times. This could be prompted by a voice command, a Wii button command from the dancer or from a keyboard command from an outside source.

Figure 89. *Instrumental*. The agent creates movement phrases based on the current movement of the dancer.

Figure 90. *Instrumental*. The pathway of the agent is drawn as a trail of crystal that becomes a musical instrument.
operator. The command essentially prompts the agent to look at the dancer’s data stream, which stimulates the neuron with the closest match. From that neuron the synaptic map takes over and different neurons are stimulated in turn to effect animation of the agent’s avatar.

The Agent’s pathway was later drawn as a trail of crystal beams, which became more dense and convoluted as the performance unfolded. (Figure 90) The dancer is prompting the agent’s movement, which in turn creates the pathway of the crystals. When the dancer stopped, the trail became a musical instrument. Crystal spheres begin to fall from above and their collisions with the pathway of beams creates a soundscape of bells. (Figure 91)

Instrumental is the most complex relationship forged from the learning process by the agent’s neural network alone. The learning process enables the agent’s capabilities and the inter-dependencies that allow the agent to respond to the dancer in a co-creative manner without any explicit programming to do so. The effect was akin to a three-way performance by Steph, her avatar and the agent, since the three ‘bodies’, one live and two projected in stereo, seemed visually and physically co-located in the space. The use of stereoscopic projection for the two projected ‘bodies’ strengthened this connection since the projected ‘bodies’ also had volume, and a defined place in the space other than simply a two-dimensional locus on a flat screen.
Chapter 3: Performance

Verbose Mode

Verbose mode was conceived of as a conversation between the dancer and agent in a more literal sense. It was heavily influenced by the work of Sutton, Barnier and Harris into transactive and distributed memory.

Below is a transcript from an interview with a long-term couple from a study conducted by Harris et al. (44) It is an illustration of a dyad with a long history of shared, complimentary processes and experiences.

F: And we went to two shows, can you remember what they were called?
M: We did. One was a musical, or were they both? I don’t ... no ... one ...
F: John Hanson was in it.
M: Desert Song.
F: Desert Song, that’s it, I couldn’t remember what it was called, but yes, I knew John Hanson was in it.
M: Yes.
(44) p. 292.

While their individual contributions to memory are fragmented, the couple is able to jointly reach a resolution through cross-cueing and co-construction. Harris et al found that examples of distributed memory structures were evidenced in such long-term couples as well as long-term co-workers. There may be a case for the inclusion of long term collaborative artists (dancers) within this class of transactions, or even short-term collaborators who nevertheless share transactive structures common in their field of endeavours. There is a particular pattern contained therein whereby one recollection doesn’t trigger the same recollection within the dyadic partner, but rather alternative recollections within the neighbouring regions of the event. Studies such as these may lead to valid
processes by which a transactive performance system may be developed.

This study in co-construction of a shared past event is analogous to the co-construction of performance based on shared movement experiences from rehearsal or past performances. In rehearsal the dancers share movements, which become the materials for cross-cueing in order to construct a present re-synthesis in performance. In a semi or fully improvised performance the performers can cue ongoing creation of the performance. This process can be understood as a form of the scaffolded reciprocal “co-authoring” of the present described by Li. The source material is relatively unchanged yet the agent and dancer create a new version of events each performance.

The rhythm of shared remembering, the conversational to and fro is quite peculiar. Concepts don’t need to be finished - they can be tentative, questioning, fragmentary. Yet they build inexorably to a form of resolution made richer by the interaction. The conversational manner is very like improvising a satisfactory outcome to resolve the remembered story. I was very drawn to the hither and thither of the shared remembering and tried to capture the rhythm in the constructive dialogue between the agent’s verbal commentary and the dancer’s movement promptings.

Verbose Mode (Figure 7) used the gesture recognition capabilities of the agent to create a duet between the moving dancer and a responsive, speaking agent. The combination of Self-Organising Synaptic Map and Hidden Markov Model allowed the agent to recognise the short phrases performed by the dancer. Each phrase was given a name to identify it. The names were descriptive.
or chosen to help remember and individuate them. The names were also used as labels for the recognition training of the HMM, so that if a match was made the agent would have a word to associate with the gesture. Lists of phrases, gathered from internet searches, comprised the agent’s verbal commentary. When the agent thought it recognised a movement gesture with enough certainty it would choose one of the texts associated with the gesture to speak.

The agent and dancer both had similar particle based avatars. The dancer’s avatar was responsive to her movement, the agent’s to its voice. I was interested in representing the avatars in a manner by which they could be seen as individuals, yet they would also be seen as a single environment. The flow of the avatars should follow the meandering, intertwined nature of the conversations. I decided to use a physics based particle simulation as it gave the qualities of complex interweaving that I was after. I had tried a few other forms of representation. However the particles gave the best rendition of the flow and energy passing between the performers. Rather than try to visualise the dancer’s entire body, I created a line of force from the dancer’s pelvis to her head, with the particles generally emitting from the pelvis and swarming towards the head. I created an attractive force that would act as a gravity well centred on the dancer’s head and the particle responded in the manner of objects caught up in the gravity well. The particles from the two avatars could also respond to each other as they collided, adding to the at times appearance of a single entity. When the dancer or agent were still or ceased speaking the particle avatars would coalesce into separate individual forms.

Figure 94. Verbose Mode. There is a conversational quality to the sequence of events, a to-and-fro as in a verbally sparring couple or good friends camaraderie.
I used some frequency analysis on the agent’s voice to change the positions of the particle emitter and the gravitational attractor, creating the movement of the agents “body”. Different bands of frequency operated on the spatial coordinates of the emitter and attractor.

The dancer performed gestures that the agent had previously learnt, and the agent injected comments into the performance based on what it believed to be the gesture performed. The performance had a to and fro quality to it, a conversation in verbal and non-verbal language. The verbal text is sometimes like commentary, sometimes critique of the dancer’s performance. Together it builds up a story of the associations remembered between movement and language. The dancer could stop to listen to the agent or continue performing gestures.

In a typical performance, the dancer performed 21 gestures of which the agent correctly identified 19 and incorrectly 2 gestures. The recognition rate was generally around 90 per cent. However, it did not particularly matter to the relationship between the dancer and agent if the agent was incorrect. The dance derived some amusement from the agent’s correct and incorrect choices.
Discussion

In this project I developed a performing agent capable of utilizing a familiar creative workflow with a dancer to engage in a collaborative performance making process. From conception, through the rehearsal process and on to performance, the relationship between the dancer and agent was considered as a means of supporting the agent’s capabilities and learning. This allowed the dancer to follow a performance making trajectory that was familiar to her and enabled the relationship with the agent to be viewed in similarly familiar terms. The framework of distributed cognition provided a conceptual structure whereby we could envisage the agent and dancer as a single supportive system rather than developing the agent with self-contained capabilities. Allowing the dancer to support the agent throughout the process enabled the achievement of a significant outcome using relatively simple processes. The agent could share the dancer’s structuring abilities to augment its own.

One of the most exciting outcomes of this research has been what Anderson descriptively terms emergence. (21) Anderson posits emergence as a means of explaining how the evolutionary history of an agent can be included in an explanation of the foundations for complex behaviour. Evolutionary history in this regard refers to the historical development of the agent through interactions between the agent and the environment over generational lifespans. As an example of emergence, Anderson offers the work of Steels in artificial life as the roots for artificial intelligence. (84) Steels describes a simple robot with two programmed behaviours: the first instruction is to take a zigzag path toward any light source, the second is to turn before moving again if it comes into contact with any obstacles. The robot must replenish itself by moving to recharging stations when the light at the recharging station turns on. There is no explicit programming to tell the robot to recharge...
itself. However this emergent behaviour occurs naturally when the robot is placed into the environment. The programmed
behaviours provide the opportunity for the emergent behaviour yet do not explicitly control the behaviours. Rather the
behaviours emerge through the dynamic interaction between the robot and its environment. While this example seemingly
illustrates emergence over a single lifespan or iteration, complex behaviours could emerge as the result of multiple
iterations with behaviours evolving through successive trials. Indeed this example of behavioural emergence in robots has
its own evolutionary history in the preceding experiments that gave rise to this particular case. Previous iterative
experiments create an evolving history to underpin development of the robot’s behaviour.

The performing agent I created similarly displayed emergent behaviour in its ability to recognise movement the dancer was
performing and to then be able to respond with movement based on what it had learnt from the dancer. However unlike
Steels example, the performing agent was not programmed with any initial simple behaviours, and all behaviour was
attributable to the learning process with no directed behaviour. The use of an artificial neural network to encapsulate the
interactions between the agent and its environment (in this case the dancer’s movement data) proved very successful in
enabling the agent’s performing behaviour to emerge. This was most evident in the performance piece Instrumental where
the dancer’s movement data stimulated the agent’s neural network, which created a response in the closest matching
neuron, which in turn caused reverberations in the synaptic map. The result was the emergence of recognition behaviour of
the dancer’s current movements and creation behaviour in the agent’s response with appropriately themed movement
synthesis.

The enhancement of the Self-Organising Map with a Synaptic layer, which I have termed Self-Organising Synaptic Map, is
an innovation that made the creation of stylistically compatible movement by the agent possible from the trained SOM. The
SOSM allows temporal potentials without a purely linear determination. As the synaptic layer is also developed through learning, it has a strong relationship with the dancer’s movement, yet because of the multiple, weighted pathways developed between the neurons there is still scope for variability in the movement created by the agent.

The choices that led to these emergent potentials were derived from the framework used from the beginning of the research project, namely, the concepts of embodied and distributed cognition. Embodied cognition focused my attention towards devices that would allow the agent to learn from the sensory input provided by the dancer’s movement data. The emphasis was not on developing a model of the performance environment or the dancer relationship, through programmed symbolic representation as in traditional approaches to cognition and programming, but rather to allow the interactive engagement of the agent with the dancer to become the key component of its environment, thereby allowing behaviours to develop. Distributed cognition, which presumes embodiment, provided a framework for imagining the co-creative relationship between the dancer and agent during creative development and performance and hinted at the potential for allowing the agent to naturally embed its learning and potential in the dancer’s evolved intelligence. Adopting this approach allowed the dancer to guide the agent where appropriate and also allowed the dancer in turn to respond to the agent’s reactions. While the dancer was the most experienced partner in the performance relationship, the creative energy was not all one way. The dancer was able to feed her creative choices from the agent’s performance as well.

This research also pointed towards a new approach for me to the development of software. Rather than attempting to build a comprehensive internal model to represent the system and its inputs and outputs, a model similar in characteristics to a traditional, computational view of cognition, the software centres on learning from the interactions between the agent and environment through the sensory engagement with the dancer through her data. This follows methodologies proposed by
roboticists and AI researchers Brooks and Steels into the application of Artificial Life to Artificial Intelligence. (85) This change in paradigm has proven extremely liberating - the author no longer needs to control the behaviour of the agent, but rather provide the circumstances whereby the agent can learn from interactive experience allowing behaviours to emerge.

The use of SOSM in combination with HMM proved effective in providing the agent with the means to learn and express movement based on a human dancer’s movement. The functional abilities of the synaptically-enhanced SOSM and HMM combination to enable a viable, responsive and engaging performing partner were in clear evidence. SOSM / HMM proved successful in allowing the agent to recognise short but complex full-body dance movements performed by the dancer and to respond with appropriate commentary. While this aspect of the performance had a supervised flavour in the format of the text responses, I am working on learning mechanisms that could give the agent more autonomy over its verbal responses. The use of the SOSM as a type of dimensional reduction filter for the HMM proved effective and reliable. The recognition rate was very high, even in a live performance where accurate segmentation of the gestures was quite challenging. Using the emissions of the SOSM for the input to the HMM allowed the agent to respond in real-time to the dancer’s movement and facilitated a smooth conversation between the agent and dancer as they drew on past, rehearsed events to co-create a new version of events for each performance.

One of the great challenges and indeed requirements of the embodied approach to performance creation was the use of both known and improvised movement material. This required the agent to be able to operate in a non-linear manner in order to remain truly responsive to the dancer’s movement creation. The use of set sequences of movement would have made the agent’s job easier as it could be told what state the dancer was in and given direction for suitable responses. The semi-improvised movement required a non-linear interactive relationship between dancer and agent. The agent could not
be sure what the dancer was going to do at any moment and so a structure enabling it to generalise and operate with this uncertainty was necessary. The situated cognition-inspired neural learning provided that structure and successfully so.
Conclusion

The performing agent in this investigation proved quite capable of autonomous participation in an interactive live performance with a human dancer. I was able to successfully develop the circumstances whereby the agent and dancer could work together to co-create a substantial live performance. This is just a first step, and there are many possibilities and further developments that could be undertaken to further enhance the performative capabilities of the agent. There is also the scope for using the learning methods and applying them to other domains.

Spotting gestures in continuous movement

While the agent was able to learn to successfully recognise short movement phrases or gestures performed by the dancer, more work is needed in the area of segmentation of the movement to enable the agent to better identify the beginning and end of the dancer’s movements. Accurate segmentation is a non-trivial task given the complexity of a contemporary dancer’s movement, which has a wide dynamic range encompassing stillness and high velocity movement. Future work could investigate a range of segmentation methods, as well as investigate using the relationship between the dancer and agent to aid movement recognition. There may be shared choreographic, behavioural or learning solutions that will enhance the agent’s abilities in this area. This is potentially one of the major benefits of applying a distributed cognition framework to the relationship between dancer and agent. Solutions need not be found within the agent alone, but can be developed using the shared relationship within the process of performance making.

As an alternative to segmentation techniques I am also investigating using rejection thresholds in the HMM to determine
when a match has been made. (55) This rejection threshold method creates a new model from the HMM classes and use it to either approve or reject the input sequence pattern as a gesture. If it is approved then it is passed to the HMM classifier to see which class has the highest likelihood as usual. I have had some success with this threshold model. However there are issues such as multiple triggering that appear in performance mode. Dispensing with segmentation altogether and using the HMM alone to perform the recognition is potentially a more elegant solution.

**Movement generation variability**

There is scope to investigate the agent’s ability to create more subtle variations on the movements it has learnt. I could have applied known animation solutions such as motion blending, kinematics and body part segmentation. However, I prefer to stay as much as possible within a learning paradigm. I am also investigating variation of movement within a shared, learned style rather than variations in style itself, as a cohesive shared movement style leads to a different experience than mixing multiple styles in the same performance.

**Learning paradigm for robot movement**

I am also interested in learning models as the basis for robots learning to move from humans. One of the main challenges here is that I am not using a physics model for the agent, but a physical robot is subject to external forces. For future projects, I am investigating how a suitable filter might be generated from the learning process that would accommodate different robot morphologies.
The agent interaction as a source for choreographic invention

Just as the agent has, in a manner, ‘learnt all it knows’ from the dancer through her giving over movement and performance guidance, this notion of learning from shared experience can move the other way. The movement generated by the agent and the particular manner in which it is visualised has given rise to new structures for movement invention, which can be used for choreographic inspiration. The particular manner of moving when interacting with the “blobs” in Recognition has encouraged the dancer to move in particular ways, which can become the basis for further movement experimentation. The dancer is taking this new knowledge back into her personal practice and using it to generate new dance performance. Just as human dancers can stimulate each other with new potential movements and approaches, the agent and the particular kinetic qualities of its visualisations can also be a source for new invention that can be effected without further recourse to technology.
References

References


References