

## **9th Workshop in Speckled Computing**

School of Informatics, University of Edinburgh, 24-25/11/2010

### **Video tracking of human activity and voice recognition co-combined in augmented interactive performance environments.**

Simon Biggs

Edinburgh College of Art, November 2010

In this talk I will present documentation of a recent art work titled "Bodytext" and of works that have led up to this work which illustrate the technologies and concepts involved.

Bodytext is a performance and installation art work that involves speech, movement and the body. A dancer's movement and speech are re-mediated within an augmented environment employing real-time motion tracking, voice recognition, interpretative language systems, projection and granular audio synthesis. The acquired speech, a description of an imagined dance, is re-written through projected digital display and sound synthesis, the performer causing texts to interact and recombine with one another through their subsequent compositional arrangement. What is written is affected by the dance whilst the emerging recombinant descriptions determine what is danced. The work questions and seeks insight into the relations between kinaesthetic experience, memory, agency and language. Bodytext was produced by myself, Sue Hawksley (texts and dance) and Garth Paine (audio) whilst the artists were in residence at the Bundanon Trust, New South Wales, and the VIPRe Lab at the University of Western Sydney.

The technical realisation of Bodytext was achieved through several iterative steps, building on both off-the-shelf technical solutions, custom software developed previously by myself and other dedicated developers and new software developed specifically for the project.

I have been working with video tracking systems applied in unencumbered interactive environments since 1984, when I developed a system whilst artist

in residence at the Commonwealth Scientific and Industrial Research Organisation's National Measurement Laboratory in Sydney, Australia.

<http://www.littlepig.org.uk/right/CSIRO/index.htm>

Since then I have developed or contributed to the development of several iterations and generations of similar software systems. The video tracking software used in Bodytext was initially written by Josh Nimoy and is known as Myron (a reference to Myron Kruger, a pioneer of unencumbered interactive systems). The Myron code is open source. In 2004 Josh Nimoy and I worked together on developing a new version of Myron designed to allow not only video tracking but live video image processing. This software was initially applied in the art work "Habeas Corpus", completed in 2006. Myron was also used as the video tracking software for "Bodytext".

<http://www.littlepig.org.uk/installations/corpus/index.htm>

Since 1994 I have also been working with generative grammar systems as a means to create interactive text engines as poetic experiences. The software for these systems has been developed by myself, following the principles of Chomskian formal grammars. An example, from 1996, is Great Wall of China.

<http://www.littlepig.org.uk/wall/thewall.htm>

More recently, I have been developing interpretative grammar engines. These systems differ from generative systems in that the software is required to interpret live text, which might be typed by a user or received as live speech. This software is designed to understand the grammatical structure of a sentence or text and to then allow the modification of a text whilst it retains its grammatical structure. Whilst the grammar system for this work was written by myself the speech recognition technology used in this work is the off-the-shelf software, Dragon's Macspeech Dictate. A work that employs this technology is "Utter".

<http://www.littlepig.org.uk/installations/utter/index.htm>

In the instance of "Utter" the modification of a specific text is achieved by two or more texts, that might intersect one another, reading one another and

parsing each other's grammatical structure. If the texts determine they are able to exchange semantic elements whilst retaining their grammatical structure they do so. The outcome of this process is that the individual textual fragments re-write one another, with new texts emerging from the process. "Utter" also uses a basic video tracking system, based on Myron, to observe the hand gestures of the speaker. This information is used to determine the orientation and movement of the textual fragments.

In "Bodytext" the video tracking, speech recognition and interpretative grammar systems outlined are used in relation to a live dancer and speaker who is the sole source of both text and movement information in the work. The piece is further enhanced with the addition of live generative granular audio synthesis by the Australian acousmatic composer Garth Paine. The sound for "Bodytext" was generated entirely from the dancer's spoken language within the work. The dancer also had control of the work through reserved command words. This allowed the dancer to shift the modes of the work, for example from a state of recording or interacting to recording and interacting.

<http://www.littlepig.org.uk/installations/bodytext/index.htm>

Individual textual fragments were programmed to be responsive to the dancers movements. The interaction profile used was not a simple one to one system, such as repulsion or attraction, but predicated on a more complex model where numerous aspects of the input data were resolved to determine the behaviour of the text fragment. This included dynamic data about the activity of the dancer's limbs, their orientation, speed, acceleration and direction of movement, their proximity and the frequency of other text fragments in a region. Individual fragments acquired energy from interactive events and used this to execute their movements. Usually they would not expend all the energy and thus stored a proportion, depending on how they were interacted with. The dynamism of each textual element was thus a function of the interaction and their current state. The texts were also programmed to "re-write" one another, as already described, when they passed through one another, this capability also partly conditioned by the current state of the textual elements.

As the performance in Bodytext unfolded the levels of noise in the system increased. This was largely due to errors that crept into the system through the less than perfect interactive systems employed in the work. The speech recognition system, in particular, was challenged by the volume levels of the generative audio system, which introduced increasing noise and errors into the speech acquisition and text generation processes. The result was that the work becomes less predictable and more chaotic as the performance proceeded. In the last phase of the work the texts ceased exchanging their semantic elements and instead randomised the letter order in each word of a sentence as the sentences passed through one another. This ensured the sentences retained their structural coherence whilst losing their semantic reference and sense. The texts, at this point, also began to spin out of control, unleashing the energy that had been accumulated in them during the performance. The entropic dynamic of the work was complete.