MODULATIONS OF VOXEL SURFACES THROUGH EMOTIONAL EXPRESSIONS TO GENERATE A FEEDBACK LOOP BETWEEN PRIVATE MOOD AND PUBLIC IMAGE

M. HANK HAEUSLER
University of Technology, Sydney, Australia
matthias.haeusler@uts.edu.au

Abstract. My proposal is an investigation into the perceptual boundaries between human and architectural expression. It asks how architecture can creatively adopt human expression by using the emotions 'displayed' on the 'surface face' as a generator for displaying a surface on a voxel façade to achieve a cross-connecting perceptual change with modulations through emotion (Massumi, 2006). Through voxel facades the public with their expressed emotions will be included in the decision process of defining space, by expressing our innermost feelings through an architectural medium. Thus emotions of the individual have a platform and can be conveyed indirectly to the public, and in turn open up discussions about the state of the community through the state of the façade. An alliance of media and place in an urban context can be achieved and created, with the participation of its inhabitants, along with a new perception of how media and architecture can together shape and inform spatial relations for a feedback loop between private mood and public image.

Keywords. Voxel façade; simulation; human-environment interaction; dynamic space.

1. Introduction

With the advent of 3D displays, also known as voxel façade technology, new design paradigms for designing dynamic space with voxels become necessary. Recent developments of these LED-based static 3D display system such as NOVA (Schubiger, 2008) or the Spatial Dynamic Media Systems (Haeusler, 2007) allow the commercial purchase of voxel systems and thus call for design of a media content to use their full potential.

At present most concepts for media content and most interfaces are still based on forms of participation in a two-dimensional mode of operation. Media contents previously used for display on a 2D screen are fed into the system or participation with the voxel façade is effected via a 2D screen. These observations raise the question of what kind of media content and what kind of participation would allow a voxel façade to demonstrate its full potential.

This abstract offers media content suitable for a voxel façade generated by a new kind of public participation in urban environments, with movement and human expression being compared to conventional architectural 'expression'. This modulation of voxel surfaces through emotional expressions will then be evaluated to investigate whether a feedback loop between private mood and public image can be achieved.

2. Voxel Façade

To generate this translation of human into architectural expression uses a voxel façade system. In a voxel façade system, data is represented as 3D or 4D form by using a matrix of voxels, a volume element representing a value on a regular grid in 3D space. Each voxel is represented as a sphere by the use of a LED-based static 3D display system. The light of each of these LEDs produces an intangible surface made of light points This intangible surface can be achieved by placing a number of RGB LEDs next to each other, where the distance of the centre of each LED is the same as to the other neighbouring LED in all three axes to create the 3D-grid or a zone in which media content is displayed. Through this arrangement an equal resolution can be achieved in all three directions: images can be displayed in an X, Y, and Z plane and, more importantly, as a 3D-object. In this way, a 3D-object, surface, form or image defined by light points can be realised within the zone.

Thus a voxel facade is able to display a time dimension within a spatial construct. This time dimension can be achieved by ever-changing data that defines the movement within the zone set by the voxel arrangement. Each voxel is represented as a sphere by the use of a LED-based static 3D display system such as NOVA (Figure 1) or the Spatial Dynamic Media Systems (Figure 2).

A voxel façade is able to display a time dimension within a spatial construct. This time dimension can be achieved by ever-changing data that defines the movement within the zone set by the voxel arrangement. The data concept allows the participation of the public by collecting media content generated through public expression. Through this data injection the public can be included in the process of defining space.

Figure 1. NOVA system, Zurich Switzerland *Ó* Horao

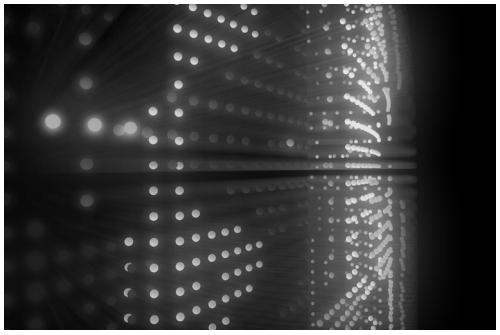

Figure 2. Spatial Dynamic Media System

One possible form of data injection could be through expressing our innermost feelings which are then represented and articulated through an architectural medium. Thus emotions of the individual have a platform and can be conveyed indirectly to the public, and in turn open up discussions about the state of the society through the state of the façade. An alliance of media and place in an urban context can be achieved and created, with the participation of its inhabitants, along with a new perception of how media and architecture can together shape and inform spatial relations for a feedback loop between private mood and public image.

The interest of this paper is in using facial expression as a vehicle to show emotions and use these emotions to generate a voxel façade media content and further to use this example to demonstrate what form of interface could be used to alter voxel facades meaningfully.

3. Media Content

William F. Allman (1994) argues in his book about evolutionary psychology *The stone age present, how evolution has shaped modern life–from sex, violence,*

and language to emotions, morals, and communities that humans are more competent in memorizing and reading animated objects such as faces than in remembering unanimated objects or numbers. The ability to read and understand animated objects and read them in the same way through all cultures can be seen in the ability to universally interpret the movements of the human face such as smiling, frowning, or shock.

Communication between humans can be held verbally but at the same time non-verbally by miming and gestures. Non-verbal communication is generally easier to understand, will be similar in most cultures and will be read by all cultures in the same way. The movement of a certain body part will thus be understood as a movement with a certain intention to express emotion. This non-verbal understanding has found its artistic expression in the performance art of a mime artist. Modern mime is a branch of theatre in which the performer usually uses no voice but instead performs using spatial and corporeal movement, full-body physical expressions, body language and gesture, often with little or no theatrical props. It is often, but not always, done in white face and the movements and expressions are heightened for greater effect.
As Allman (1994) states,

> "Emotional cues are so important to human survival that a 'universal grammar' has evolved in human facial expression. The human facial expressions that spring from feelings of grief, sadness, anger, disgust, surprise, fear, and happiness are universal among all human societies."

Allman (1994) argues as well that:

> "For [the] most people, the facial muscles involved in shaping the face when they are experiencing emotions are not under conscious control. Only 10 percent of us, for instance, can voluntarily pull the corners of the mouth down to make the protypically human 'sad' face. The rest of us can make this face only while also moving the muscle near the chin, which is a giveaway for a phoney expression. Likewise, only 15 percent of people can voluntarily raise their eyebrows at the center of their forehead to duplicate the forlorn look of grief and distress."

To allow humans an easier recognition of these facial expressions a study of the nature of beauty in human faces was undertaken. The conclusion Allman (1994) took from this study was that:

> "… the human brain has been sculpted by evolution to regard as most attractive those faces that are, ironically, most average. This preference may have evolved in our ancestors because, as representatives of the community at large, these average faces would be easier to read for subtle expressions of happiness or concern that serve as important social clues."

Communication through mime has therefore been an important part in our daily life. But what is miming in a spatial sense?

When moving the face to express something, the surface of the face changes and therefore creates a changing surface. The surface face is alterable through parameters such as emotions that will be shown by different arrangements of the face so even if the faces look different the emotion or meaning will always be the same. These expressions can be digitalized and reproduced in a façade that allows the display of 3D images, such as a voxel façade. As mentioned above, each part of the face is responsible for expressing a certain kind of emotion–Allman (1994) refers to the mouth muscle being responsible for expressing sadness, while the chin muscle stays still.

These face movements are the input data used for the proposed feedback loop between private mood and public image. Each expression on each part of the human face will be then located in a field of the voxel façade, ideally in approximately the same location of its original location on the face. Therefore a voxel façade fed by these input data based on emotion is able to express emotions itself and regardless of linguistic diversities and cultural differences, the voxel façade or face–and here I want to highlight the linguistic similarity between these two words–could function as a mediator between private mood and public image.

This will be done by miming the same expressions as humans, but through the artificial nature of the voxel façade it stands between humans and puts the focus not onto the human but on its emotion. It will strengthen community relations and respect the linguistic diversity and the cultural knowledge of different languages by not having the need to use them, through communicating with non-verbal methods.

4. Interface

Having established an argument of the kind of media content allowing modulations of voxel surfaces through emotional expressions I briefly want to raise the question of interface to generate data produced through emotional expression. This question can be divided into two main fields – interface hardware and software. Both aspects are briefly discussed, but due to the paper's interest in the theoretical aspects of generating media content no detailed solutions are offered.

A possible interface connected to a voxel façade would contain as hardware a 3D body scanner, e.g. the Vitrus ahead–3D head scanner©. To quote the manufacturer,

> "Three dimensional scanning is as easy as can be: a person stands in the center

> of the measuring portal and the laser light sources (which are not harmful for the eyes) are switched on. […] A few seconds later, the necessary data is collected. A so-called '3D point cloud with colour overlay' composed of several million 3D measuring points has been generated. A virtual duplicate of the person has been created with maximum precision. Due to its high resolution the 3D head scanner VITUS ahead is able to scan even delicate details of the face. The products generated with the scanned 3D data show an amazing similarity to the real object. The non-contact measurement is carried out by 8 triangulation cameras with a resolution of approx. 1 mm within less than 10 seconds. […] The data can be used for animations as well as for medical purposes. (www.vitus.de/english/: Dec 2006)

According to the manufacturer, the first format of the so-called "3D point cloud" is an ASCII format that can be translated into different 3D modelling formats such as *.STL, *.OBJ, *.VRML. These are common file extensions for 3D model programs such as Rhino, 3DMax and others. The above mentioned programs are all able to work with scripts and are therefore able to take the data from the "3D point cloud", point by point with the exact coordinates for each point as a X, Y, Z value.

The second interest in a possible realization of an amalgam of emotional and architectural expression is in software allowing emotion recognition based on the 3D face scanning. The question of software and its ability to capture human expressions and emotions will be discussed when positioning the paper in the field of interest.

5. Positioning of Paper in the Field of Interest

Research conducted by others provides evidence of the existence of software that can capture emotions from, in one case, a 2D image and in the other in real time from viewing a human face. On 17 December 2005 the NewScientist.com news service published the article *Software decodes Mona Lisa's enigmatic smile* that discussed the possibility of recognizing human emotions just by looking at photographs or videos:

> "It's official: Leonardo da Vinci's Mona Lisa was 83 per cent happy, 9 per cent disgusted, 6 per cent fearful and 2 per cent angry.
> Nicu Sebe at the University of Amsterdam in the Netherlands tested emotion-recognition software on the famous enigmatic smile. His algorithm, developed with researchers at the Beckman Institute at the University of Illinois, Urbana-Champaign, examines key facial features such as the curvature of the lips and crinkles around the eyes, then scores each face with respect to six basic emotions. Sebe drew on a database of young female faces to derive an average 'neutral' expression, which the software used as a standard to compare the painting against.

> Software capable of recognizing human emotions just by looking at photographs or videos could lead to PCs that adjust their response depending on the user's mood, as well as smarter surveillance systems." (www.newscientist.com: Dec 2005)

This software could be used as a component when writing a code for software that could fulfil the required task of recognizing the emotion of a 3D-scanned face. However, since the article describes the capturing of emotions from a two-dimensional image therefore another research example should be discussed.

The Bristol Robotics Laboratory at University of the West of England researches about humanoid robotics and social interaction with one focus on what they call Natural Facial Expressions: Towards Empathy in Humanoids (www.brl.uwe.ac.uk/projects/empathy/empathy.html: Dec 2008).

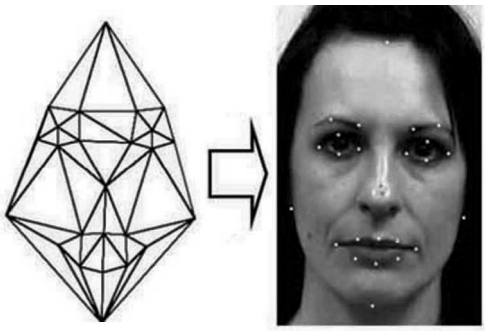

Figure 3. Facial recognition *Ó* Bristol Robotics Laboratory

Their research interest advances the modelling and generation of realistic, dynamic facial behaviour in humanoid robots. The interesting aspects for this paper are that the researchers:

> "… investigate ways of extracting facial motion information from video footage and translation of such results to control commands for application to robotic faces." (www.brl.uwe.ac.uk/projects/empathy/empathy.html: Dec 2008).

They go on to explain that:

> "Facial motion data is obtained by tracking a set of facial feature landmarks in video sequences or live video. We use Active Appearance Models (AAM) (Cootes et al., 1998; Stegmann, 2003), which are generative, statistical models of face texture and shape. Fitting AAMs to video frames is an iterative process, searching for a set of parameters, which yield minimum error between AAM-instance and an input image. The shape of the AAM delivers a low dimensional descriptor of the facial expressions, further used in our experiments. (www.brl.uwe.ac.uk/projects/empathy/empathy.html: Dec 2008).

Both research labs have an interest in using facial expressions and emotions which either:

> "… lead to PCs that adjust their response depending on the user's mood, as well as smarter surveillance systems. (www.newscientist.com: Dec 2005)

In the case of the research conducted by Nicu Sebe at the University of Amsterdam in the Netherlands or in the case of Bristol Robotics Laboratory at University of the West of England to:

> "… investigate ways of extracting facial motion information from video footage and translation of such results to control commands for application to robotic faces." (www.brl.uwe.ac.uk/projects/empathy/empathy.html: Dec 2008).

Neither research project uses these data to allow a reactive or even interactive spatial set up for a human environment interaction. Only by combining voxel façade technology and emotion recognition on faces can a feedback loop between private mood and public image generated by a new kind of public participation in urban environments be achieved.

6. Conclusion

The question of what kind of media content and what kind of participation would allow a voxel façade to demonstrate its full potential through a new kind of public participation in urban environments, with movement and human expression being compared to conventional architectural 'expression' have been answered by the paper. What conclusions could be drawn from this hypothesis when asking if a feedback loop between private mood and public image through a voxel façade can be achieved?

Massumi's (2006) cross-connecting perceptual change through emotions comes to mind again. Massumi refers there to Lars Spuybroek's/NOX and Q.S. Serafijn's *D-tower* (www.d-toren.nl: Oct 2007) a tower come as a sculpture in Doetinchem, a town in the Netherlands that is hooked up to a website. The residents of Doetinchem complete online questionnaires about their mood depending on the results the tower changes colour. Massumi explains:

> "Affect has been given visual expression. The predominant affective quality of people's interactions becomes visible. This can undoubtedly reflect back on the interactions taking place in the town by making something that was private and imperceptible public and perceptible. A kind of feedback loop has been created between private mood and public image that has never existed in quite this way before."

He continues:

> "There are any numbers of ways for creating this kind of eventful cross-connection, between different perceptual modes, different phases of perception formation and between perception and affect. What has been accomplished thus far is just the tip of the iceberg."

The research presented in this paper works towards exposing more of 'the tip of the iceberg'. The D-Tower includes the perception of colour by the human eye, the origin of colour in materials, colour theory in art and colour psychology. With colour psychology as a field of study devoted to analyzing the effect of colour on human behaviour and feelings, it is shown that these effects differ as various cultures see colours differently. So colour alone, as seen in the D-Tower by NOX Lars Spueybroeck and Q, bears the risk of being misunderstood in a cross-cultural community. In addition, a determined form designed by an architect still does not allow culturally diverse communities to express themselves in an architectural gesture. The D-tower was designed by an architect from one cultural group, which could be used by people from different cultural groups, so even though the colour and the appearance express the emotions of the participants the architecture itself does not.

A voxel surface that is defined by the captured similar emotions–the participation of the beholder in a cross-cultural form of expression–ensures that one or more cultures are no longer excluded. It allows the beholder to reflect themselves in the building, leading to a negation of the design idea of the façade. The author or architect vanishes behind his/her work and the beholder becomes the person in charge of designing the "temporal events-within form", an expression coined by Stephen Perrella (2001) when he talks about topological space as:

> "…influenced by the inherent temporalities of animation software, augmented reality, computer-aided manufacture and informatics in general."

A façade that is a "temporal event-within form" can allow the beholder as a participant to be engaged in the design of the temporary appearance of the spatial relation they have with their environment, i.e. it allows creation through participation. This participation could be stimulated through a spatial representation of emotions via a voxel façade, because, to quote Massumi (2006) once again:

> "Information delivery is the least interesting thing they do. It is only interesting to the extent that it feeds the creative emergence of new forms of lived experience."

Acknowledgements

I would like to acknowledge my former PhD supervisor Professor of Innovation Mark Burry SIAL / RMIT University, Melbourne for his help during my research which has influenced this paper.

References

Allman, W. F.: 1994, *The stone age present, how evolution has shaped modern life–from sex, violence, and language to emotions, morals, and communities*, New York: Simon & Schuster, 1994, p. 94.

Cootes, T., Edwards, G. and Taylor, C.: 1998, Active Appearance Models, in *Proceedings European Conference on Computer Vision 1998*, vol. 2, pp. 484-498.

Perrella, S.: 2001, Hypersurface Theory: Architecture×Culture, in C. Di Giuseppa (ed.), *Architecture and Science*, Wiley-Academy, Chichester, pp. 139-147

Massumi, B.: 2006, Transforming Digital Architecture from Virtual to Neuro, An interview with Brian Massumi by Thomas Markussen & Thomas Birch, published online at: http:// www.intelligentagent.com/archive/Vol5_No2_massumi_

Markussen+birch.htm, (accessed November 2006).

Haeusler, M. H.: 2007, Spatial Dynamic Media Systems, Predicting the Future, in *eCAADe Conference Proceedings*, Frankfurt am Main (Germany), pp. 69-75.

Schubiger, S.: 2008, Large Screen Interaction in Public space: TowerTalk and NOVA, in: Adams, R., Gibson, S. and Müller Arisona, S. (eds.), *Transdiciplinary Digital Art: Sound, Vision and the New Screen*, Springer, Berlin Hamburg.

Stegmann, M. B.: 2003, The AAM-API: An Open Source Active Appearance Model, Implementation, in *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2003)*, pp. 951-952.