

# Semiotic Analysis for Gestural and Emotional Human-Computer Interaction

Roman Danylak

University of Technology, Sydney  
roman@emotional-computing.com

**Abstract.** The discussion that follows describes the design of an interactive artwork *To be or not to be* using semiotics. The goal of *To be or not to be* was to create a user experience that was coherent and continuous, generating gestures and emotions - not uncommon Human Computer Interaction (HCI) objectives. The design problem was to isolate what could be elements of user experience as inputs and outputs in a multimedia interactive system. Essentially there were five parts to the process: the first was an understanding of gesture and its modeling within the framework of generating expressive gestures in theatre – defining media before simulation; the second, was to define the input and output process by which gestural interaction using HCI media might proceed; the third was to create a semiotic matrix of both the theatrical and HCI terms as *equivalences*, creating a system by which the design could follow; the fourth was evolving an experience, in this case an interactive film-game, that generated gestures and associated emotional content; the fifth was a user evaluation and statistical analysis (results summary only). The emphasis presented here is on the preparatory stage of correct process modeling, leading to the effective application of semiotic analysis. Readers are encouraged to access the URL for youtube description of the work.

<http://www.youtube.com/watch?v=jKNvSpXG0Z0>

**Keywords:** theatre, game, gesture, emotion, metaphor, metonymy.

## 1 Introduction

The recent rapid development of extended sensory input and output devices, including tactile and visual aids, has meant that new forms of data emerging from human experience can be processed by a computer. Bolt's work [1], *Put That There* (1980) was the first instance of gesture-based multimodal interaction where deictic (pointing) gestures would move on-screen objects. Since then, the range of human experiences as input data has continued to rapidly increase. A design question then arises: 'could the role of human communication as presented by what is commonly known as *body language*, be integrated into a human-computer

interaction? To rephrase and refocus this ambitious goal, would be to say that some useable, communicating configuration of gesture and emotion in a multi-modal computer interaction might be possible. *To be or not to be* partially realized this objective by generating gestures and emotional response in a multimedia interactive context. The success of the design was owing to the careful modeling of the expressive gesture process based on expressive gesture generation in theatre, followed by semiotic analysis of both the human and machine design elements, enabling the design in a game format.

## 2 The Problem of Gesture: Ambiguity and Context

The notion of language in view of the problem of gesture and emotion persists. A major difficulty with gestures and their communicated emotional content is that meaning – a key feature of language - can be ambiguous. Gestures and interpreted expressiveness is context dependent: a raised open hand could mean ‘stop’; it could mean ‘hello’; or ‘throw me the ball!’; or it may be done to sense the air temperature and the presence of a breeze. Alternatively, what seems to be expressive content may have no meaning at all – “I was only stretching my muscles!”. Most importantly gestures then, do not exist in isolation. Establishing a clear vocabulary where ‘this movement means X’ is very difficult to achieve. As such, an all-encompassing system of design that would harness gestural emotional human computer interaction – a means of simplifying the complexity of user experience and the design process - is as yet elusive as Oviatt and Pantic state respectively [2, 3].

## 3 A Five Step Design Process for *To be or not to be*.

There were five key steps in designing *To be or not to be*. Step 5 will be discussed in summary only. These steps are as follows:

1. understand gesture and its modeling within the framework of generating expressive gestures in theatre, defining the media machine *before* simulation.
2. define the input and output process that gestural interaction by which the HCI might proceed.
3. create a semiotic matrix of both the theatrical and HCI terms as *equivalences*, creating a system by which the design could follow, modifying terms as necessary.
4. evolve a user experience applying semiotic matrix, in this case a game.
5. user evaluation and statistical analysis of objectives.

### 3.1 Step 1 Defining Gestures, Emotion and Media in Theatrical Practice

The literature on gestural interaction displays a difficult variety of definitions, primarily arising out of its ambiguous nature. The modeling adopted was that of theatrical

practice, adopted to reduce the volatility of the problem. Actors perform repeated expressive gestural sequences in the virtual space of the theatre, suggesting it to be an appropriate model. The process by which actors prepare their bodies for performance is often based on yogic exercise as exemplified in the work of Grotowski [4]. In yoga, gestures are defined by stillness; this stillness allows for the examination of emotion associated with the pose. A still form is then what constitutes a gesture; the test for the presence of gesture then is a photographic still. Significantly, intention, communication, reception and potential ambiguity with this definition disappear from the user experience: bodily form equals gesture.

In terms of defining media, creating gestures in theatre is essentially a three-step process, that is, if we assume the creation of the play plans – the script – already to be in existence. The gestural and verbal expression of the play is devised by the director who interprets the text and directs actor actions on stage to create the illusion of live reality through the iteration of rehearsal; this is the case in classical Stanislavskian scripted theatre, which is still quite commonly practiced [5]. Figure 1 diagrammatically shows key steps.



### 1/TEXT:

The play exists as script; dramatic conversations, inscribed in text, to be interpreted by directors and actors. What will the play mean? What shall be its expressions, its gestures?



### 2/REHEARSAL:

Gestural and verbal iterations. Actors and directors work together to form and language shaping the virtual reality of the work.



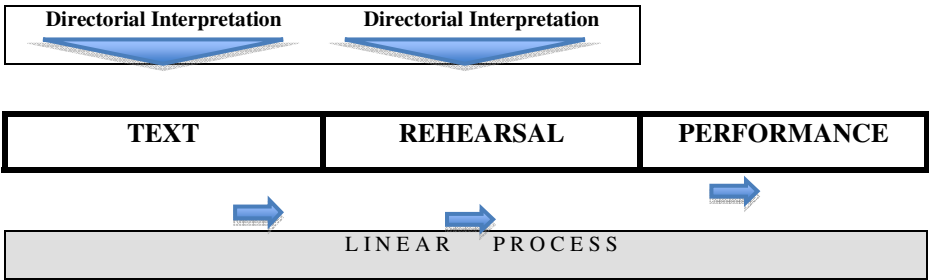
### 3/PERFORMANCE:

Gestural and verbal iterations are now fixed. Repeatable live performance is possible with a set, expressive paradigm of language and gestures.

**Fig. 1.** Diagrammatic representation of the three stages in expressive gesture development for theatre: a multimedia process. Images source: Google

The process of text / rehearsal / performance is then, importantly, a *linear multimedia process*, each phase separate to the other, developed through directorial interpretation of each step. These elements underpinned the design of the gestural interface following semiotic analysis. Directors interpret text with the outcome of repeated performance in mind. The process is further summarized in Figure 2, emphasizing the separate components, highlighting the presence of elements that are fixed, such as text, versus elements that are interpreted and therefore arbitrary.

Therefore, in Step 1, the severe ambiguity of gestural meaning was reduced. What emerged is a modelled process where expressive gestures can be manufactured according to a scripted plan.



**Fig. 2.** The linear multimedia process used in classical theatre to create expressive gestures in a narrative context. Director’s interpretation at script and rehearsal phases influences expressive content. Once performed, interpretation is complete. Scripts (text) are nevertheless static.

### 3.2 Step 2 Defining Multimedia HCI Inputs and Outputs

The range of input and output devices available was quite extensive and growing rapidly. The next question that needed resolution was: what combination of input and output devices would underpin a successful design where gestural and emotional human-computer interaction could proceed? The assumption that this would satisfy is that the experience was to be somehow modeled on the theatrical model. Table 1 indicates the large choice of user sensory experience; Table 2 indicates the wide choice if input devices.

**Table 1.** Sensory channels for potential user experience in HCI modality

| Sensory perception | Sense organ | Modality   |
|--------------------|-------------|------------|
| sense of sight     | Eyes        | Visual     |
| sense of hearing   | Ears        | Auditive   |
| sense of touch     | Skin        | Tactile    |
| sense of smell     | Nose        | Olfactory  |
| sense of taste     | Tongue      | Gustatory  |
| sense of balance   | Cochlea     | Vestibular |

The vast array of multimedia input / output systems was quite daunting. At this stage of development, the range of devices was seen as useful opportunities. The only design distinction that was made is that gestural interaction, that is to *touch*, was a foundational process as evidenced by Sutherland’s invention of pen based interaction in 1962. The action of touching the screen was a gesture that located user presence in the matrices of the computer system. Touching is also an act of signification [6] a key semiotic process.

**Table 2.** Key advances of multimodal input /output devices with gestural application, adapted from Turk [7]

| Attached Technological Device                   | Description   | Date invented | Inventor / significant technical demonstration     |
|---|---|---------------|--|
| pen based                                       | pin pointing of light manipulated on screen   | 1963-5        | Sutherland (1963)                                  |
| tracker - based gloves                          | hands used as sophisticated control device  | 1992          | Sturman (1992)                                     |
| tracker based - data suits/ coda motion         | continuous gesture analysis in combination with eye tracking  | 1994          | Wexelblat (1994)                                   |
| passive and computer vision gesture recognition | tracking of gestures using cameras  | 1870/ 1983    | Muybridge([1887] 1957)<br>Moeslund & Granum (2001) |
| head and face                                   | FACS system measurement of facial movement  | 1978          | Ekman & Friesen (1978)                             |
| hand and arms                                   | graphic extraction of hand and arm positions  | 1995          | Stark & Koehler (1995)                             |
| sign language (hearing impaired)                | visual recognition of coded hand signals  | 1995          | Starner & Pentland (1995)                          |
| body gestures                                   | Pfinder system 2D representation of tracking and gesture recognition using statistical models of colour and shape | 1996          | Wren, Darrell, & Pentland (1996)                   |
| dance gestures                                  | dancer generating music and graphics through body movements using Pfinder system                                  | 1999          | Paradiso & Sparacino(1999)                         |
| laser based tracking                            |   |               | Perrin, Cassinelli & Ishikawa (2004)               |
| floor pad systems                               | registers users entry into an environment and spatial coordinates through pressure sensitive pads                 |               | Hightower & Borriello, (2001)                      |
| magnetic field tracker                          | use of magnetic field to establish spatial coordinates  |               | Hightower & Borriello, (2001)                      |
| Cave Virtual Automatic Environment              | gestural manipulation of biological models  | early 1990's  | Berry, Pavlovic & Huang (1996)                     |

**Fig. 3.** Ivan Sutherland's ground-breaking Sketchpad console, 1962. Sketchpad is operated with a light pen and a command button box (under left hand) [8] - evidence of gestural interaction.

### 3.3 Step 3 Forming a Semiotic Design Matrix

Winnograd [9] has stated that the computer is a *language machine*. Its processing of mathematical symbols, a part of its deep functions derives the term ‘programming languages’; its extensive manipulation of image, text and sound representations in various combinations makes this invention a primary machine of communication. The computer then has many language systems within it. Semiotic analysis, a meta-language process that describes the symbolic and sign processes of potentially any language system, logically, would seem to offer a possible solution to design problems associated with computers - this language machine - an approach first proposed by Andersen [10]. The semiotic system here followed is that of Saussure [11], which is also Andersen’s prescription. In particular, what was useful for the project at hand was Jakobson’s [12] further extension into the notions of metaphor and metonymy. Significant modifications to semiotic analysis were necessary, particularly in view of the non-temporal capacity of semiotics; semiotics does not easily define dynamic, time-based systems, a key characteristic of expressive, performed gestures.

What is noteworthy is Kant’s attempt to develop a system where all art forms could be represented in a schema, calling it a *System of the Arts* [13]. He begins with the premise that art is *essentially expressive*, just as language is, and for this reason language is a means by which the expressiveness of art can be understood (see Table 3). The key concepts used were *word*, *gesture*, and *tone*. He stated, that there are equivalent characteristics of both the execution and their function in human sensory processes. This presents something of a prototype to the process evolved. The qualities he attributed are categorized as follows:

**Table 3.** Kant’s *System of the Arts* [14]

|                |               |           |
|----------------|---------------|-----------|
| <b>Word</b>    | Articulation  | Thought   |
| <b>Gesture</b> | Gesticulation | Intuition |
| <b>Tone</b>    | Modulation    | Sensation |

As mentioned, the key semiotic tools used in this study were Jakobson’s notions of metonymy and metaphor, to which was also added the term, utterance (see Table 4); the first two terms are characteristic of Jakobson’s work on semiotics, metaphors being that which represents something else or other – from the Greek *metapherein* meaning ‘to transfer’, whilst metonymies represent the same - also from the Greek meaning ‘to change name’. The third element added is utterance [15], that is, the speech act itself. This is the word as sound not as meaning. The significance of utterance is its temporal quality and that is the emanation of the word: it is ephemeral, fleeting. The ‘act’ itself of speaking is significant; something that happens once and once, only emphasizing the purely performative quality of language as voice. This element being ephemeral cannot be programmed whereas the metonymic, sequential

**Table 4.** Key Characteristics of Semiotic Terms

| Semiotic Terms   | Definition                   |
|------------------|------------------------------|
| <b>Metaphor</b>  | Representation of other      |
| <b>Metonymy</b>  | Representation of same       |
| <b>Utterance</b> | Ephemeral, not unlike speech |

elements can. These terms and their key characteristics are represented in Table 4. Hence the expressive nature of the gesture is made equivalent to the expression of the word. To these three key terms abstractions are added: form, sequence, wave and world, to aid in understanding their characteristics and context - see Table 5.

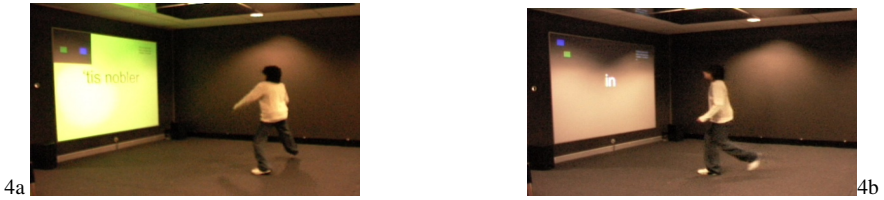
Further, Table 5 shows comparisons of a number of the design elements for analysis and includes the semiotics for the final interactive film game. Most importantly, the theatrical model has within it elements that may support the HCI design in the game format. The interactive film game is then but a refiguring of the theatrical process. Andersen’s [16] ‘acting machine’ is an analysis of a ship’s multimedia system built by Andersen, highlighting the difference between two dimensional screen systems to the real action – or gestures – of users. The metonymic elements are particularly useful as they can be designed into the interaction of the system. Metaphors can be included but require interpretation. Whilst anything in the ‘utterance’ category is ephemeral and cannot be programmed or designed interactively.

**Table 5.** Semiotic Matrix of Human and Computer Elements

| Abstraction | Semiotic          | HCI     | Yoga      | Theatrical Model | Game Model   | Acting Machine |
|-------------|-------------------|---------|-----------|------------------|--------------|----------------|
| Form        | <b>Metaphor</b>   | Image   | Gesture   | Gesture          | Film content | Tool           |
| Wave        | <b>Utterance</b>  | Sound   | Emotion   | Emotion          | Emotions     | Automaton      |
| Sequence    | <b>Metonymy</b>   | Touch   | Narrative | Text             | Text/ Touch  | Media          |
| World       | <b>Linguistic</b> | Virtual | Somatic   | Stage            | Ludic        | Engineering    |

**3.4 Step 4 Design of User Experience: Interactive Film Game *To be or not to be***

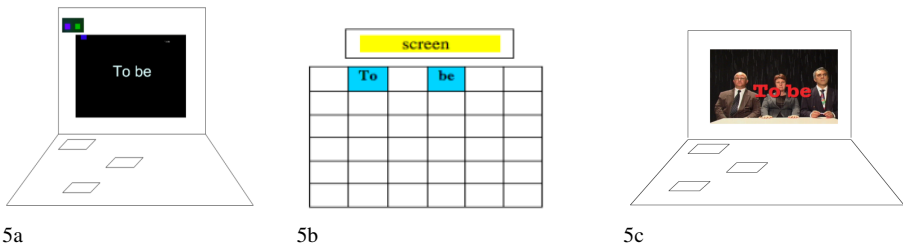
Users walked upon an array of a thirty-six square interactive floorpad measuring 3m x 3m connected to a Mac G5 computer which also controlled an adjacent 2m x 3m back



**Fig. 4.** a, b. Interactor in two different gesture positions. The user is activating programmed floorpad patterns corresponding to the onscreen position map, activating film sequences to assemble a narrative. The emotional response to the assembled narrative as interpreted by the film director is essential. See <http://www.youtube.com/watch?v=jKNvSpXG0Z0>

projected multimedia screen with sound which the user faces when interacting with the floorpad (Figs. 4 a, b).

When interactors entered the room the work was in a steady default state displaying a title and brief description. As the interactors walked onto the floorpad this triggered the game cycle. The second screen gives a brief description of the game rules and the functions of the on-screen gesture map: green squares show a user where they are actively walking; blue squares are an indication of where users should walk to; red squares indicate that the user has hit a target, being in the right place at the right time on the programmed floorpad grid. The goal of the game – to solve all twenty-three gesture / word puzzles for full film play back – see Figures. 5 a, b, c.



**Fig. 5.** a,b,c. 5a shows the initial screen in gesture puzzle mode; 5b is a map of the floorpad programmed with positions to be found to activate the screen, the first of the twenty three such puzzles. 5c shows the system in playback of one film sequence of the Hamlet film.

Figs. 5(a) shows Step 1 of *To be or not to be*. The foreground shows some representative floor pad squares, thirty-six in total number, in a regular square configuration. The first two syllabic phrases of the *Hamlet* text appear in the middle of the screen illuminated by computer video back projection ‘To be’. Interactors can see where they are standing on the active grid from the floor pad map in the upper left hand corner (green active squares). A pre-programmed blue flashing square also appears which the user must locate in the space corresponding to their position on the floorpad to activate the game goal of film playback. Having reached the target square, shown by a red square appearing on the map, a section of the Hamlet film plays. 5(b) is a map of the first two gestures to be solved on the floor pad appearing as blue square screen targets to be located on the floor pad, turning to red on activation. These



have been pre-programmed along with loaded reward film sequences. Fig. 5c shows the first film sequence playback. Each sequence lasts 4-7 seconds and is evolved from the Shakespeare text that has been used in the gesture floor pad puzzle 'To be'. Multimedia text appears on the film surface reinforcing the narrative experience. Essentially, the interactor, by walking about the floor pad, assembles the Hamlet film. Once all twenty-three puzzles are solved the film plays back in its entirety as a single work, communicating specific emotional messages formed by the director's interpretation.

### 3.5 Step 5 Evaluation

Evaluation of the system was carried out with a number of random participants. The gestural interaction hit rate with the floorpads was recorded in the MAC G5 and showed consistent interactor learning of gestures. Verbal protocol analysis of the interactor response to the film content showed consistent emotional reaction to the film when compared to the directorial intention of the film message for the audience. Hence the interpreted film meaning was successfully communicated *emotionally* to users whilst generating a predictable *gestural response* [17]. The equation formed is: gesture sequence 'x' generated an associated emotional response to content 'y'.

## 4 Conclusion

As Andersen has stated, the division between the human experience and the machine process is formidable [18]. However, the careful modeling of gestures from the theatrical process set a real and workable foundation for the entire task. Defining the gesture as a 'photographic still' removed the inherent ambiguity of gestural meaning establishing an equation between gesture and computer gestural input. Furthermore, semiotics enabled the design of *To be or not to be* by offering equivalences between the human and experience and the machine design. Semiotic analysis is effective if the correct cognitive modeling is developed with adjustments to the semiotic categories to accommodate temporality. Lastly, it was the game design that refigured the theatrical process into an interactive film game, where interactors could build a narrative using walking gestures and also respond emotionally to the director interpreted narrative content. The design enabled a continuous and coherent gestural and emotional user interaction, evidenced in evaluation studies.

## References

1. Bolt, R.A.: Put That There: Siggraph Computer Graphics. In: Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques ACM, vol. 14(3), pp. 262-270 (1980)
2. Oviatt, S.: Multimodal Interfaces. In: Jacko, J.A., Sears, A. (eds.) Human-Computer Interaction Handbook. Lawrence Erlbaum & Associates, Mahwah (2003)
3. Pantic, M., Sebu, N., Cohn, J.F., Huang, T.: Affective Multimodal Human-Computer Interaction. In: Conference Proceedings, Multimedia 2005, pp. 669-676. ACM, Singapore (2005)

4. Grotowski, J.: *Towards a Poor Theatre*. Methuen, London (1969)
5. Stanislavsky, C.: *Building a Character*. Methuen and Co. Great Britain (1936, 1981)
6. Danylak, R., Edmonds, E.: Touch as the Act of Signification: naming as a Key design Concept for Gesturally Intuitive Interactive Space. In: *Engage, Interaction, Art and Audience Experience*. Creativity and Cognition Studios Press (2006)
7. Turk, M.: Gesture Recognition. In: Stanney, K. (ed.) *Handbook of Virtual Environment Tech.*, p. 8 (2002)
8. Muller-Prove, M.: *User Experience and Interface Design* (2008), [http://www.Mprove.De/Diplom/Text/3.1.2\\_Sketchpad.html](http://www.Mprove.De/Diplom/Text/3.1.2_Sketchpad.html) (viewed March 22, 2008)
9. Winnograd, T.: *Scientific American*. In: Rheingold, H. (ed.) *Virtual Reality*, p. 215, Touchstone (1991)
10. Andersen, P.B.: *A Theory of Computer Semiotics: Semiotic Approaches to Construction and Assessment of Computer Systems*. Cambridge University Press, UK (1990)
11. Saussure, F.: *Course in General Linguistics*. McGraw-Hill, New York (1986)
12. Jakobson, R.: Two Aspects of Language and Two Types of Aphasic Disturbances. In: Pomorska, K., Rudy, S. (eds.) *Language and Literature*, Harvard University Press, Cambridge (1956)
13. Kant, E.: *The Critique of Judgement*. Oxford University Press, UK (1790, 1848, 1978); Meredith, J.C.(trans.)
14. Slater, H.: The System of the Arts. *The Journal of Aesthetics and Art Criticism* 51(4), 611–617 (1993)
15. Greimas, A.J., Courtes, J.: *Semiotics and Language: an Analytical Dictionary*. Indiana Univ. (1979)
16. Andersen, P.B.: Acting Machines. In: Liestol, G., Morrison, A., Rasmussen, T. (eds.) *Digital Media Revisited*, pp. 183–213. MIT Press, Cambridge (2004)
17. Danylak, R.: *A Semiotic Analysis of Gesture and Emotion in Artistic Human Computer Interaction*. Doctoral Thesis. University of Technology, Sydney (2008)
18. Andersen, P.B.: What Semiotics Can and Cannot Do for HCI. *Knowledge-Based Systems* 14(8), 419–424 (2001)