

CHEIRO

Creating Expressive Textual Communication and
Anthropomorphic Typography

Francis Lam

B.Sc. Computer Science
The Chinese University of Hong Kong
May 2000

B.A. Design
Hong Kong Polytechnic University
May 2003

Submitted to the
Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of
Master of Science in Media Arts and Sciences at the
Massachusetts Institute of Technology

June 2006

© 2006 Massachusetts Institute of Technology
All rights reserved

Author
Francis Lam
Program in Media Arts and Sciences
May 12, 2006

Certified by
Judith S. Donath
Associate Professor of Media Arts and Sciences
Thesis Supervisor

Accepted by
Andrew Lippman
Chair, Department Committee on Graduate Students
Program in Media Arts and Sciences

CHEIRO

Creating Expressive Textual Communication and
Anthropomorphic Typography

Francis Lam

Submitted to the
Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of
Master of Science in Media Arts and Sciences at the
Massachusetts Institute of Technology

June 2006

Abstract

Despite rapid technological advancement and improvement in network bandwidth, people still like to use text for remote communication. Simplicity, directness and anonymity make textual chat the most popular method of communication in the technology-mediated world.

In face-to-face communication, people can use gestures, facial expressions, eye gaze and other body languages to alter, emphasize or strengthen their spoken words. Unfortunately, current textual platforms do not have a proper channel for these cues and signals. These non-linguistic cues are able to convey social and emotional information accompanied by the spoken words. I argue that a well-designed textual communication system can increase the expressiveness of text-based chat environments, and we need to have a better chat interface to improve our social interactions in the digital world.

Cheiro is an exploration of user-centered gesture-based interfaces that enable expressive textual communication. My approach is to use common input devices, such as mouse or keyboard, as the gestural interface to amplify or change the tone and meaning of the text, and send non-linguistic signals using graphical elements and anthropomorphized kinetic typography. The goal of this thesis is to find an intuitive mapping between the user's gestural input and the graphically enhanced text output, which constructs a novel expressive textual communication platform.

Thesis Supervisor: Judith S. Donath
Title: Associate Professor of Media Arts and Sciences

CHEIRO

Creating Expressive Textual Communication and
Anthropomorphic Typography

Francis Lam

The following people served as
readers for this thesis:

Reader
John Maeda
Associate Professor of Design and Computation
MIT Media Laboratory

Reader
Walter Bender
Senior Research Scientist
MIT Media Laboratory

Acknowledgments

Thank you very much Judith, my respected advisor. You teach me many things - from big ideas to subtle experiences - verbally and non-verbally.

Thank you John. Your words and works change my view on “design”.

Thank you Walter for your patience and guidance on this thesis.

Thank you Aaron, Christine, Fernanda, Orkan, Scott, and all of you from the Sociable Media Group. It is a wonderful experience to work and learn with you all.

Thank you James, Edward and all my friends in the lab. You guys rock!

Thank you mom, I love you more than words can say.

Table of Contents

| | | |
|----------|---|-----------|
| 1 | INTRODUCTION | 10 |
| 1.1 | BACKGROUND | 11 |
| 1.2 | RELATED WORK | 12 |
| 1.3 | WHAT IS CHEIRO? | 13 |
| 1.4 | NEW ACHIEVEMENTS | 15 |
| 2 | THEORETICAL FRAMEWORK | 16 |
| 2.1 | THE PROBLEM | 16 |
| 2.2 | THE INPUT | 18 |
| 2.2.1 | <i>Face-to-Face Communication</i> | 18 |
| 2.2.2 | <i>Non-linguistic Communication Cues</i> | 20 |
| 2.2.3 | <i>Meanings and Mappings</i> | 22 |
| 2.2.4 | <i>Orality and Textuality</i> | 23 |
| 2.3 | THE OUTPUT | 27 |
| 2.3.1 | <i>Motion and Kinetic Typography</i> | 27 |
| 2.3.2 | <i>Why Anthropomorphic?</i> | 28 |
| 2.3.3 | <i>Production Process</i> | 31 |
| 3 | BUILDING CHEIRO | 36 |
| 3.1 | OVERVIEW | 36 |
| 3.2 | A STUDY OF EMOTIVE MOUSE GESTURES | 38 |
| 3.3 | THE NEW EXPRESSIVE TEXTUAL COMMUNICATION PLATFORM | 42 |
| 3.3.1 | <i>Animated Text</i> | 42 |
| 3.3.2 | <i>Non-Linguistic Communication Channel</i> | 47 |
| 3.3.3 | <i>Mouse Gesture Recognition</i> | 48 |
| 3.4 | DEVELOPMENT PROCESS | 50 |
| 3.4.1 | <i>Avatar Chat</i> | 50 |
| 3.4.2 | <i>Synchronous Gestural Chat</i> | 51 |
| 3.4.3 | <i>Other input devices</i> | 52 |
| 3.5 | USER STUDY | 54 |
| 3.5.1 | <i>Online User Study</i> | 54 |
| 3.5.2 | <i>Qualitative Analysis</i> | 54 |
| 4 | FUTURE WORK..... | 58 |
| 4.1 | IMPROVEMENTS..... | 58 |
| 4.2 | FUTURE DIRECTIONS..... | 60 |
| 5 | CONCLUSION | 62 |
| | BIBLIOGRAPHY..... | 64 |

1 Introduction

Despite rapid technological advancement and improvements in network bandwidth, people still like to use text for remote communication. Simplicity, directness and anonymity make textual chat the most popular method of communication in the technology-mediated world. However, current text-based interfaces, such as Instant Messaging (IM) clients and Short Messaging Service (SMS), are not designed to express subtle communication cues and nuances.

In face-to-face communication, people can use gestures, facial expressions, eye gaze, intonation and other body languages to alter, emphasize or strengthen their spoken words. Research has shown that non-linguistic cues contribute to the majority of the information conveyed in a conversation [7]. Moreover, non-verbal information can act as direct signals to convey social meanings to the viewer. For example, you shake your head to indicate refusal or nod to show approval. Unfortunately, current textual platforms do not have a proper channel for these cues and signals.

I argue in this thesis that a well-designed gestural input system can increase the expressiveness of online text chats, and we need to have a better chat interface to improve our communication in the digital world. In face-to-face conversation, hand gestures are able to: (1) convey social signals and information accompanied by the spoken words; and (2) deliver non-verbal communications

[10]. They are ubiquitous and intuitive. People even use gestures while they are speaking on the telephone [41]. In the digital arena, because of the limitation and the convention of the computer input technology (primarily mouse-keyboard combination), text is mainly formulated by hand-based devices. It becomes an opportunity to collect gestural input from these devices to convey nuances in online chat.

Cheiro is an exploration of gesture-based interfaces that enable expressive textual communication and convey non-textual information in the online world. My approach is to use common input devices, such as mouse or keyboard, as the gestural interface to amplify text by using kinetic typography, and enable non-verbal communication by visual elements. In terms of the system input, metaphorically, it is the non-linguistic signals such as gesture and intonation that amplifies our spoken words. In terms of the output, it is like writing a note; you underline, color or use graphical elements to visually annotate the written words. The difference between these two metaphors is the timing of amplifying the meaning of words. Spoken words are accompanied by gesture simultaneously, whereas written words are annotated afterward. Thus, the challenge is to find an appropriate and intuitive mapping between the gestural input and the visual output, so that users can easily learn and understand. In this thesis, I describe the new expressive text-based chat environment *Cheiro*, its development process, the theoretical framework, and the user studies conducted.

1.1 Background

The most common way of making text expressive and more than the words themselves in online chat is to use emoticons or smiley faces [13]. Those simple faces function as punctuations to convey expressive information that amplifies and clarifies the sender's intention. Yet, as we are so sensitive to the face, it may signal unintended information - particularly when many of the current IM interfaces push the limit of the graphical sophistication of their smiley icons (see figure 0).

Kinetic typography or animated text is a common technique to add emotive characteristics into text. The animation of text with changing type, size, and color offers a substantial promise in

conveying the speaker’s emotion, tone of voice and character, as well as directing the viewer’s attention [18]. It is widely and successfully used in film and advertising industries, but rarely in real-time communication interfaces [24]. I hypothesize that kinetic typography can also convey user’s emotion in online chat spaces. Although research has been done in developing tools for animating expressive text [24], my focus is to apply kinetic typography to convey emotive information and subtleties in real-time environments.



Figure 0. Emoticons used in MSN Messenger [27]

1.2 Related Work

Kinetic Typography-Based Interface

Bodine *et al.* [8] attempted to incorporate kinetic typography in an IM client. The user can specify four different types of effects to apply onto the text message using a button-based interface. Text is animated by changing of size, speed or motion when the user hit a specific function key to represent emotion. However, the interface is limited because of its primitive input mechanism. Users might feel unnatural and restricted [17], as they would not be able to convey the nuanced emotional information by a limited number of specific animations and keys.

Affective Computing Interface

Other systems tried to bring expressive qualities of the face-to-face communication to the digital arena by gauging the user's affective states using physiological sensors [23][39]. Because of the involuntary nature of these interfaces, privacy issues have to be concerned. Users may not want to expose their emotions or facial expressions to someone who they do not know. *Conductive Chat* [28] tried to avoid the privacy concerns by translating the galvanic skin response into different forms of text, but the user still cannot deliberately control the text output. The focus of my work is not about guessing or sensing intention for expressions. Rather, I would like to design a user-centered system to allow the user to consciously convey non-textual information through an intuitive and smooth input mechanism, and display the visual output in a legible and meaningful way.

Gestural Interface

Cheese [29] is a web interface that records all the user's mouse movements on a page. It shows that certain gestural behaviors are common among users and are able to increase the effectiveness of the interface design. Fagerberg *et al.* [17] created an affective interaction model for designing gestural input to a mobile service for affective messages. They did an in-depth analysis of shape, effort, and valence to find a set of affective gesture mechanisms. They also developed a gestural input system to convey emotions through a mobile device, where output is expressed as a gradation of colors and patterns. Their work is mainly focused on the input side, which creates a fundamental framework for using gestural input in computer-mediated communication. This thesis work focuses on the visual output, specifically, how the gestural input can be mapped to kinetic typography and graphics, and how they can convey meaningful information.

1.3 What is Cheiro?

Cheiro is an online text-based chat interface that recognizes users' mouse movements as gestural input for expressing emotions and social cues in an online chatting environment. It explores the graphical possibilities of improving textual communication, and attempts to increase expressiveness in online chat environments using common input devices, such as keyboards and mice. The output of the message is appeared as an animated text reminiscent to a human posture,

to express emotional and social cues. I call it *Anthropomorphic Typography*. In next chapter, I describe in detail how it works and what the conceptual meanings of the motions are.

The interface consists of four main parts (see figure 1): the text input area, output stage, chat history, and gesture toolbox. Users type in their message in the upper input area and the animated message is shown in the output stage below. When the next message comes in, the one on the stage is moved to the conversation history and archived in the database. The users can animate the text by a set of mouse gestures, which are simple one-stroke patterns derived from the study of emotive mouse gestures. Users can also change the gesture patterns for different emotions or define their own ones by the mouse gesture customization toolbox.



Figure 1. Cheiro interface. (1) Text Input Area (2) Output Stage (3) Chat History (4) Gesture Toolbox

(<http://www.media.mit.edu/~francis/cheiro2/>)

1.4 New Achievements

The main achievement of *Cheiro* is enabling non-linguistic communications in text-based chat interfaces. These non-linguistic cues are reminiscent to the intonation and the rhythmic patterns in one's spoken language, and the gesture and body movements accompanying the spoken words. They send implicit signals about the speaker's emotion and emphasis. *Cheiro* allows users to convey these cues and emotions by an intuitive gestural input over a text-based interface.

The *Anthropomorphic Typography* shortens the distance between the gestural input and the graphical output. All visual communications cues (e.g., gestures, gaze, etc.) in face-to-face conversations are invisible in online text-based environments. They are mediated by graphical elements, which may induce extra cognitive load for users to interpret them. *Cheiro* tries to reduce the cognitive load in processing these mappings and make them easier to remember. In the following chapters, I discuss the theoretical framework that *Cheiro* is grounded and the process of building the system to create an expressive textual communication platform.

2 Theoretical Framework

In this chapter I describe the theories and the studies for the current implementation of Cheiro. I first introduce the problem and the motivations for building the system. Then I structure the theoretical framework into two parts: the input and the output. The input part focuses on the non-linguistics features of face-to-face communication, and how these features shape the design of the input mechanism of the interface. The output part discusses the graphical manipulations on the text, which help convey subtle information and facilitate social interactions in an online text-based chat environment. In chapter 3, I explain how these theories and studies are applied to the design of Cheiro.

2.1 The Problem

Text is inarguably the most direct means of communication in the digital arena. Millions of people use text-based IM clients to keep in touch with their peers everyday. Millions of dollars are spent everyday for sending text messages over the mobile phone network. People like the directness of text to convey complicated meanings quickly. The anonymity of text also makes it popular in the computer-mediated world for exchanging information. However, human communication is more than exchanging of words, we have many social cues and signals embedded in our face-to-face communications which sometime even convey more information than the verbal content. Current text-based communication systems have many shortcomings because they do not support nuanced expressions. They are difficult to express emotions, social and other non-linguistic cues. The goal of this thesis is to design a textual communication tool

that is catered for nuanced communication and is able to convey subtle non-verbal information in the online space.

Smiley Faces function as extensions to punctuations, which depict facial expressions for conveying emotional cues [43]. The simplicity of these emotional punctuations makes them become very prevalent in online chat environments as an extra non-verbal channel in textual communication. However, they are too primitive to depict complicated or mixed emotions and it is difficult to memorize many of them to be used fluently. Newer IM clients (e.g., Microsoft's MSN messenger) map these symbols to tiny graphical icons (or Emoticons), which can be used with text messages. However, another drawback of these icons is that, if the representation becomes more sophisticated, it will have more connotations. It may lead to misinterpretations of the information conveyed. If the icon is too detailed, it may even prevail over the actual content. For instance, an icon showing a photograph of a laughing man or a cartoon character can be used to convey happiness. However, it also sends out other social cues such as gender, age, or race, which may not be intended by the sender. Because most of the Emoticons are created for general purposes, they inevitably contain some wrong connotations about the sender. These problems in the current text-based chat environments frame the development of the basic objectives of this thesis. They are: (1) to design an input mechanism which can intuitively attach emotive cues while composing the text message, and (2) design rich output that can represent subtle emotions while not incurring too many incorrect connotations.

2.2 The Input

The design of the input mechanism is largely based on face-to-face communication. I discuss the features and functions of face-to-face gesture and expression, which shape the gestural input of *Cheiro*. People are very sensitive and attentive to the non-linguistic cues, which may change the meaning of words spoken radically. The core function of *Cheiro* is to convey these non-linguistic cues in a mediated text-based environment. But then, the parallel between face-to-face communication and *Cheiro* is also very important. There are many functions of gestures or postures that *Cheiro* cannot do. The lack of simultaneity in textual communication is also a problem. I mention the difference in orality and textuality, as well the practicality of an online chat interface. These concerns all contribute to the design of the input system about how people convey their emotions and contexts using *Cheiro*.

2.2.1 Face-to-Face Communication

In face-to-face conversation, the information conveyed by purely linguistic cues is often less than the non-linguistic ones [7]. We often use gesture, intonation and other non-verbal signals to communicate emotional states and social context cues. In telephone conversation, although it is less expressive, we can still use our voice, intonation and pace of speech to convey implicit messages such as the speaker's age, gender, health, attitude and emotion state. The listeners can also use the background noise to determine a range of social contexts. Current textual communication platforms, despite of its popularity, have very few concerns in addressing these issues. The vast majority is still very primitive in conveying these non-verbal cues, let alone nuanced communications that expresses complex emotions. There is a need in online text-based chat space to have more expressive channels for conveying these non-linguistic cues to embellish meaning, emotion, and social context.

My approach is to first look at these non-linguistic cues in our face-to-face conversation and shape the design of the *Cheiro* to provide more expressive richness. Gesture, *which co-occurs with spoken languages* [10], is an important element of the input of *Cheiro*. Cassell [10] points out that gestures serves a communicative function in face-to-face communication, despite the fact that the listener and the speaker may have different interpretation of the gesture from different cultures [14][16]. The design of the system flourishes from the mouse interaction; it is the most

basic “hand gesturing” carrier in the computer-mediated communication. The idea is to create a gestural input mechanism for textual communication, which uses mouse as the starting point.

For most of the time when we are talking, we are gesturing as well [26] - even when we are talking on the telephone [34]. We are very attentive to people’s gestures. They influence the way the accompanying words are interpreted. Gesture can be intentional or spontaneous. The final design of the system is more aligned with the intentional gesture because the unplanned ones already have a lucid meaning in our mind. It will limit the design possibility of the interface because we already have a set of mental representations of these gestures [10].

Many affective chat projects (e.g., [39][28]) deploy physiological sensors to read the spontaneous gestures or other non-linguistic cues in an online communication setting. However, for the visual output the concern is only text but not the hand gesture itself. The mismatch in the mapping of gestures and output of text is considerably large. After all, I do not explicitly transfer the hand gesture to the chat environment, but try to take gesture and speech as a metaphor in making conversation more expressive. My approach is to create a new gestural language reminiscent to hand gesture, conducted by common computer input devices, such as a mouse, to transform subtle non-linguistic cues into expressive animated texts. These *consciously produced* gestures are easier to remember thus more appropriate for applying to a new communicative interface.

In light of these theories, I proposed a gestural input mechanism for Cheiro. It is in the middle of the continuum between involuntary (e.g., affective chat) to very deliberate input (e.g., emoticons). In the actual interface, users can use their mouse (or other input devices which use the same mapping) to control the motion of the animated text by drawing a certain pattern on the screen. It does not resemble gesture, but it takes some basic forms of hand gestures from users to convey something beyond their words. For example, moving your mouse sideways would make an animated expression of disagreement or farewell on the text. It relates to hand gesture in the conceptual level, the user uses text (words) and mouse movement (gesture) together to convey information with both content and context. The actual patterns of these gestures and the motions of the animated text are described in the next chapter. Extended design development and requirements are also covered.

2.2.2 Non-linguistic Communication Cues

In this section, I talk about different kinds of non-linguistic communication cues that are commonly used in online and offline conversations, their importance and how they facilitate social interaction and provide social context cues. I also mention briefly how these analyses lead to the development and the design of *Cheiro*.

In face-to-face communication, we can use gesture, posture, facial expression, appearance, eye gaze and prosodic variation to convey explicit, implicit, intentional and unintended messages. Donath [13] points out that facial expression, eye gaze, and appearance are able to convey one's individual identity, social identity, emotion and physical condition. In the online world, faces are also widely used in chat programs to show identity by means of avatars or profile pictures. Emoticon as mentioned earlier is an example of adding emotional cues to textual communication. Eye gaze is also a very important channel for conveying social information [13][37]. Where the speaker is looking can direct the listener's attention. People are sensitive and adept at perceiving gaze direction. For instance, if you say: "I'm talking to you" to someone who sits next to you without actually looking at him/her. They probably would not be aware that you are referring to them, or they may consider you are being rude. Researches (e.g., [2][20]) show that gaze, gesture and language work together to facilitate social interaction in conversation. They can be used to establish social hierarchy, negotiate turn taking, and indicate understanding, agreement, and attention [9]. Listeners always direct their gaze to the speaker (about 60-70% of the time [13]). Often times, listeners wait for the gaze signal from the speaker before starting to speak. Eye movement can also function to convey emotions. People sometimes determine if someone is telling the truth by looking seriously into their eyes. Gaze is extremely important in face-to-face communication. It is still one of the biggest challenges in the area of videoconferencing.

The intonation and the rhythmic patterns in one's spoken language send implicit messages about the speaker's emotion and emphasis [35]. Prosody works the same way as other non-linguistic cues and it is even more closely coupled with the spoken words, but it has to coexist with them. Intonation can express mixed emotions with contextualized communication situation, and communicate a speaker's intention, personality and state with higher precision than other types of cues. When we say something in exactly the same wordings but with different intonations, the meaning can be totally different. These characteristics have influenced extensively the output design of *Cheiro*. Text is needless to say the most significant visual output of a text-based chat

program. Keeping the emotional and contextual information closely tied with the text is very important to convey mixed cues. As one of the major pitfalls of emoticons is not being able to convey complex emotions in one sentence, *Cheiro* is designed to express multiple meanings in one single line of message. My approach is based on the manipulation of text, which has a certain resemblance with the intonation and speech pair. I describe in depth about *Cheiro*'s visual output and how it is coherent and closely coupled with the text in the next section.

Cassell [10] presents a framework of natural human gestures in face-to-face communication. She argues that gestures can be used to express an immediate meaning (*emblematic gesture*), depict a form (*iconic gesture*), represent a metaphor (*metaphoric gesture*), locate a physical space (*deictic gesture*), and accompany the speech like a small baton (*beat gesture*). Gesture also has many communicative functions, such as negotiating turn taking, settling agreement and indicating understanding. These cues like gaze can work alone to convey simple and direct information. For example, you nod your head to represent you are following the speaker, or you shake your head slightly to convey a different opinion.

People are very sensitive and attentive to all these subtle cues, which may change the meaning of words spoken radically. You would not feel the happiness and joy when your intimate friend says, 'I'm happy' in a flat tone and a long face. Hence, my goal is to design a non-linguistic channel for people to communicate online textually. Many research projects approach the problem in a different angle. They try to read involuntary non-verbal cues by physiological sensors and computation; however, the problem of automating these cues is whether it is conveying what the person behind wishes to convey [13]. If these cues do not really reflect the user's intention, it may make the system appear to be deceptive and lose its function of being more expressive. For *Cheiro*, I take a user-centered approach, which lets users always know what and why the system is working this way. Another characteristic about these cues is that they all work together to convey more complicated or even different meanings. *Cheiro* also takes this concept to develop non-linguistic channels that can interact with each other to emphasize and convey multiple-layered social information online. The actual visual output is discussed in the next chapter.

2.2.3 Meanings and Mappings

Many non-linguistic cues, especially gesture, are culturally biased [16]. People can interpret them in different ways and contexts. A naive example is waving your hand; it can mean “goodbye!” or “no!” depending on the social situation and the language spoken along with it. In face-to-face communication, all these cues are both input and output. We make gestures intentionally or spontaneously to amplify our language, but in the same time, we interpret others’ gestures and gaze to find out their intents and internal states. However, in a textual environment where visual cues are almost invisible (except the text itself), the system should be able to generate visual output, which is intuitive to read and correlate with the gestural input. Figure 2 shows the difference in the input and output of face-to-face conversation and textual communication platform.

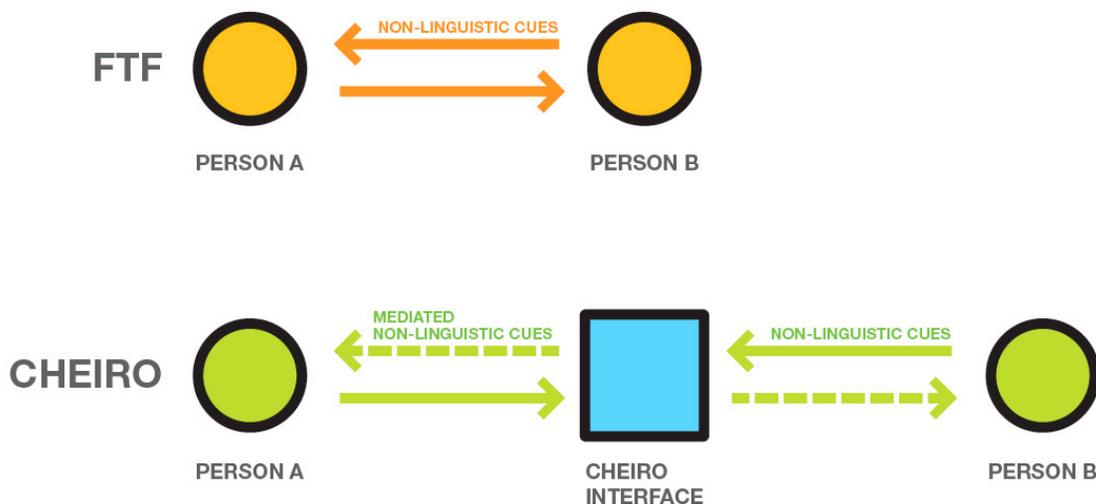


Figure 2. Input and output of non-linguistic cues in FTF and Cheiro

As a consequence, the challenge of establishing non-linguistic channels in text chat is to reduce the extra cognitive load of processing another metaphorical mapping. When I gesture or rest my gaze on you, you can directly see my hand and eye movements. Although our interpretation of these cues may not be the same, we are judging the same visual output. However, in a text-based environment where direct visual cues are not provided, we have to make a mental linkage of what we input (e.g., your gesture) and the output (e.g., the graphics produced on the screen by your gesture). In other words, my design goal for the gestural input is to create an intuitive

transformation from the mouse movement to the visual output. Current chat interfaces with emotional expression capability do not put in much consideration in this aspect. When we type a sequence of characters or click on an icon from a menu to choose an emoticon, there is no correlation between the input action and the output emotion. In *Cheiro*, I aim to create a gestural input mechanism that has a correlation with the visual output. It helps the user convey emotions or other social cues in a more intuitive manner. I argue that by doing this the user can remember the gestures easier and create a more intuitive method for expressing emotional information.

2.2.4 Orality and Textuality

From what I have mentioned above, the input part of *Cheiro* is fundamentally a gestural-based system, where the user can use a mouse to convey useful non-linguistic cues. These cues help express contextual and emotional information and amplify the meaning of words. As stated earlier, I take many references from the face-to-face communication, where non-linguistic cues are very important. However, Ong mentions that *the condition of words in a text is quite different from their condition in spoken discourse* [30]. The lack of simultaneity in textual communication is a major disparity. Delivering cues in the same way as face-to-face communication via a text-based interface is impractical and problematic.

Ong also describes writing as a solipsistic operation; you are alone while you are writing or typing. It is similar in mediated text chat - we are on our own when composing a message. Although the period of composing each message is short, comparing to the spoken discourse, real-time text chat over the computer is still relatively “asynchronous”. This difference should influence the design of the non-linguistic channels, which are synchronous in face-to-face conversation. The design for the gestural input system should also concern about the textuality of the words, and how it constitutes the interaction design of the system. Spoken utterances come concurrently with intonations, gestures and other expressions. However, written words are supplied with extratextual clues for intonations [30] such as punctuations sometime afterwards. Also, these clues can be repeatedly modified until the text is published. These textual characteristics lead to the separation of delivering non-linguistic cues and words in designing the gestural input for *Cheiro*.

Nevertheless, I also implemented a counter example by designing a more “synchronous” gestural input system, which is the first version of *Cheiro* (see figure 3). The intention to discuss it

example is to illustrate the problems in delivering textual information and show how the current version evolves. This system allows users to express emotions while they are delivering the words simultaneously. Instead of pressing the “Enter” key to send out a line of message as a whole, the user has to move the mouse to “push” out each character. The speed and direction of the mouse movement determine the motion of the characters, so that the receiver of the message sees an animated sequence of text moving across the screen from right to left. The concept is to enable non-linguistic cues, facilitated by the mouse movement, to be coupled with individual words or even letters. Although the sender also synthesizes the content and the emotion separately, the viewer perceives the message with emotional encoded data in the same time. By simple mappings for basic emotions, such as the fast and angular motion for anger or the slow and smooth for relaxed, the system works as an utterance baton that conducts the intonation of the message and display it visually and textually.

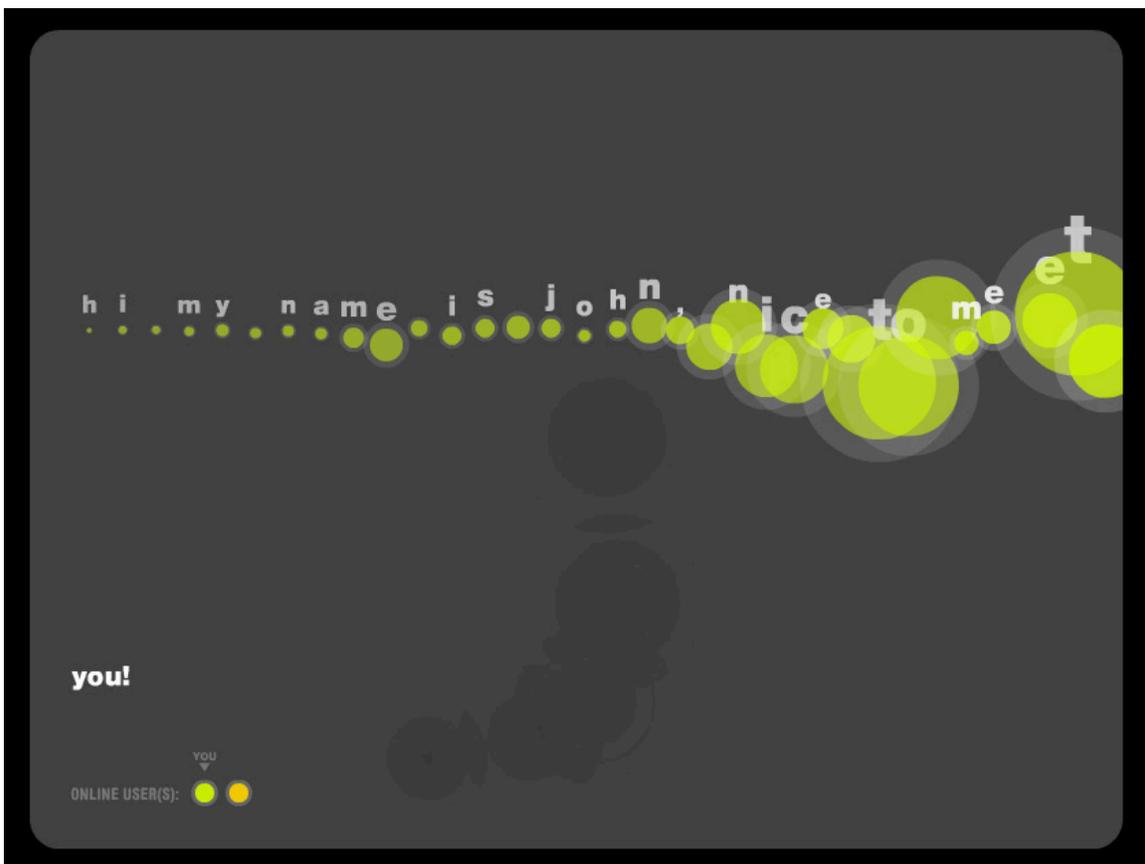


Figure 3. Cheiro Version 1 – Conveying emotion and words synchronously

After some informal user testing on the preliminary version of the system, I find that most of the users are able to quickly understand how the system works without advanced description. They

think that it is intuitive yet the legibility seems to be an issue. Moreover, the users sometimes are not able to convey what they really want to say because the system is too “real-time”. They cannot control or make sure how the characters look like before sending them out. Unlike face-to-face communication where non-verbal cues are mostly both input and output, the system’s gestural input (instead of being the output as well), triggers a set of emotional visual outputs. The mapping and recognition processes need a certain amount of cognition time, which might not be intuitive as an interface for adding emotions to the text.

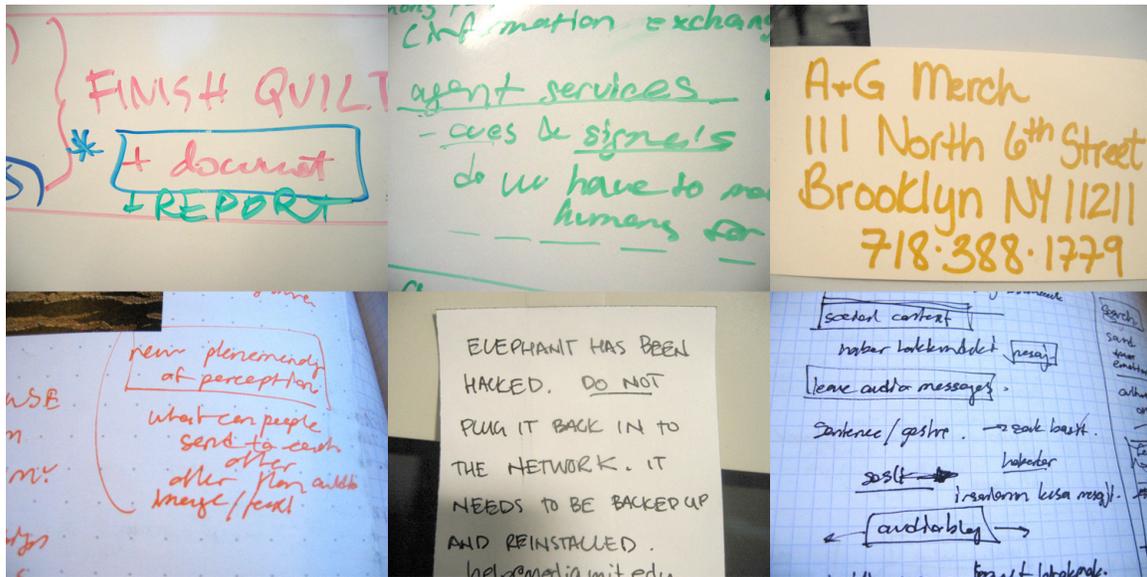


Figure 4. Handwritten notes

After this version, I turn in to a new approach for conveying emotion by gesture, which is concerned more with the textuality of the words. Text is meant to be written and read asynchronously [30]. The input part of this version becomes more analogous to writing than speaking. When people write a note, they normally add in some visual markers or drawings to amplify the meaning after inking the words. We can underline, cross out, circle or even fold the note to represent different connotations added to the semantic meaning (see figure 4). This process is after the formation of words, which I think is actually more appropriate for the chat interface as well. However, the input should not be like a drawing tool that requires fine movements to elicit an emotion, but rather an abstraction that conveys emotion. The final design of the input system takes references from both the qualities of orality and textuality. It reads mouse movements and transforms them into visual cues that convey emotional information

beyond the words. It also takes into account the asynchronous characteristic of writing and guides the design of the input mechanism.

Mouse is primarily a pointing device, however in *Cheiro*, it is an input device for delivering non-linguistic cues and emotions as well. Although mouse is not designed as a device for expressive input, when the interface allows this to happen, I want to study if there is any converging pattern, which can be used as a framework for designing the gesture recognition system. In the next chapter, I will describe a user study for examining how people use mouse to express themselves emotionally.

2.3 The Output

In this chapter, I describe *Cheiro's* output mechanism, aesthetics and meaning. The main focus of this chapter is about the method of manipulating text and making it more expressive to convey nuanced information. Graphic designers use typefaces to represent intent, emotion and personality. The form of type is the visual manifestation of language [12]. In designing the output of an expressive textual communication tool, the typeface of an alphabet is arguably the most important element to study on. Type design is a very subtle activity. Minute changes in form can communicate different intent and meaning in the message. Sharp edges and heavy weights may convey dark thoughts, while graceful and delicate characters evoke lighter feelings [12]. In *Cheiro*, my approach in type design is not static typography but more into the realm of kinetic typography, where does not only the form but also the motion governs the expressiveness of the text. The method I introduce is called *Anthropomorphic Typography*. It departs from the traditional kinetic typography in the use of motion. The conceptual framework and the production process are described in detail in the following sections, whereas the actual applications in *Cheiro* are shown in the next chapter.

2.3.1 Motion and Kinetic Typography

Motion is the strongest visual appeal to attention [3], it implies a change in the conditions of the environment and is more likely to get our attention than other static visual cues. Motion is dynamic and vibrant; it is intuitive and suitable for representing social data such as online chats and discussions. The classic graphical representations like color, shape, position and transparency of an object are limited for depicting the heterogeneous nature of data [21]. Also, the overuse of these representations would exceed the human's perceptual capacity to efficiently get their meanings [4]. It is perceptually rich and catches our attention rapidly. Its pre-attentive and interpretative perceptual properties [3][4] make visualizing a large set of data possible. It enables visualizing an overview of the chat archive at a glance. Social interactions are not static. I believe that the dynamic and expressive nature of motion allows us to visualize some social patterns of online conversations more intuitively.

There are many different ways of using motion [21] [5], including transformations of object (color change, growth, shape change) and movement of the object in 2- or 3- dimensional space.

Motion is good at both grouping and filtering of visual elements in a chaotic view of numerous objects [6]. Objects with trajectories, which have similar shapes and are not directionally distinct, can give a sense that they belong to the same group. Moving objects stand out from the static (or barely moving) ones in the background, which provides an extra dimension of representation in the same visual field. *Cheiro* focuses on the transformation of the form rather than other attributes, as the motion effect should be subtle and not affect the legibility of the text.

Kinetic typography or animated text is a common technique to add emotive characteristics into text. The animation of text with changing type, size, and color offers a substantial promise in conveying the speaker's emotion, tone of voice and character, as well as directing the viewer's attention [18][24]. It is widely and successfully used in film and advertising industries, but rarely in data visualizations. Lee *et al.* [24] identify a set of useful visual and motion features of text, based on principles of traditional animation [36], which can represent affects, intentions and other social meanings by kinetic typography. These features mathematically encode the text movement into different waveforms and paths, and enable dynamic generation of animated text.

Many researchers and designers endeavor to create an extensible system for animating expressive text (e.g., [24][34]). But, my approach is different from theirs in the use of motion. They attempt to animate text using some primitive dimensions such as size, amplitude, speed and motion path. While these attempts try to digitalize and systemize kinetic typography, my approach is on an analog route. *Anthropomorphic Typography* captures human movements and embodies them in text to communicate the subtleties of social interaction. It humanizes the generic-looking text and furnishes it with emotions and the power to convey social context and meaning. It is all about the subtleties that matter.

2.3.2 Why Anthropomorphic?

Textual communication and broadcasting are ubiquitous in both online and offline medium. Many authors use variations of typeface, font color and size, smiley faces and animated icons in their text to convey emotions and contexts. To visually represent these expressive textual data, researchers and designers usually use color, size or motion to depict the differences (e.g., [38]). However, these depictions are sometimes too generic and fail to represent subtle or complex emotions and meanings. A person would not do a particular action the same way in two different emotional states [36]. My approach is to capture the movements directly from a human actor to

create more humanistic typefaces. I argue that my method can be applied to create more expressive textual representations based on the following reasons:

Variations of color, size and motion in text are sometimes limited in expressing multiple or complicated emotions or other attributes. These mappings can be ambiguous and may not be intuitive enough to make the users understand the meaning immediately. For example, bouncing letters are normally used to represent pleasant or excited feeling, yet they can also produce worrying or anxious expression. Anthropomorphic typefaces can also be ambiguous, but they are relatively more precise and can convey a richer vocabulary of emotion.

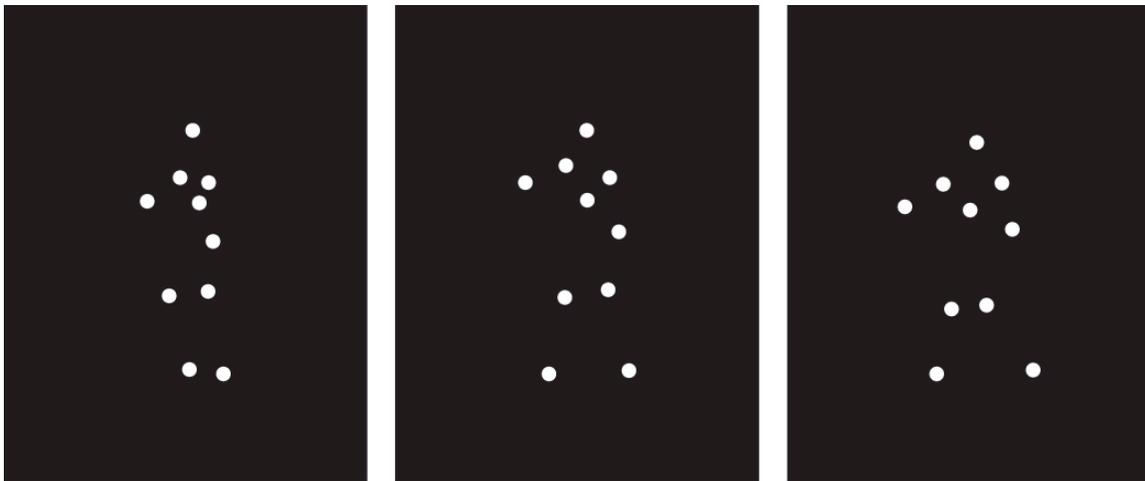


Figure 5. Snapshots of a point-light stimuli animation

Humans are very sensitive to anthropomorphic motions [1]. Strong evidence can be gathered from studies of point-light stimuli or the perception of biological motion. Johansson [22] invented the first point-light stimuli in 1973 (see Figure 5). He used a technique similar to chronophotography, which he attached light bulbs to an actor's ankle, wrist, knee, elbow, hip, and shoulder joints before taking a video of him/her. The actor was dressed in black and later filmed in a dark room with a video camera. When subjects are shown the video with only some sporadic moving light points, all of them can easily recognize that is a walking human figure.

On the other hand, human emotions are not discrete states; we express our feeling according to our personality, social settings and other individual factors. In perceiving emotions, we also cannot regard them as solely as a mental state [15][32]. Emotions can be generated through one's imagination without any physical embodiment or interaction; but they can also be generated and

perceived from body movements [15][17]. For example, in face-to-face conversation, subtle movement like leaning a little bit forward is enough to signal interruption or catch the speaker's attention. In addition, Allison et al. [1] also provide strong neurological evidence that certain kinds of motion (gaze, lips, hand and body movement) are closely related to human's perception of social information. Thus, I hypothesis that by visualizing text anthropomorphically, which bears a resemblance to body movements and postures, is able to convey subtle emotions and social signals. Likewise, the viewer can also have a better grasp of the meaning of those familiar movements at a glance.

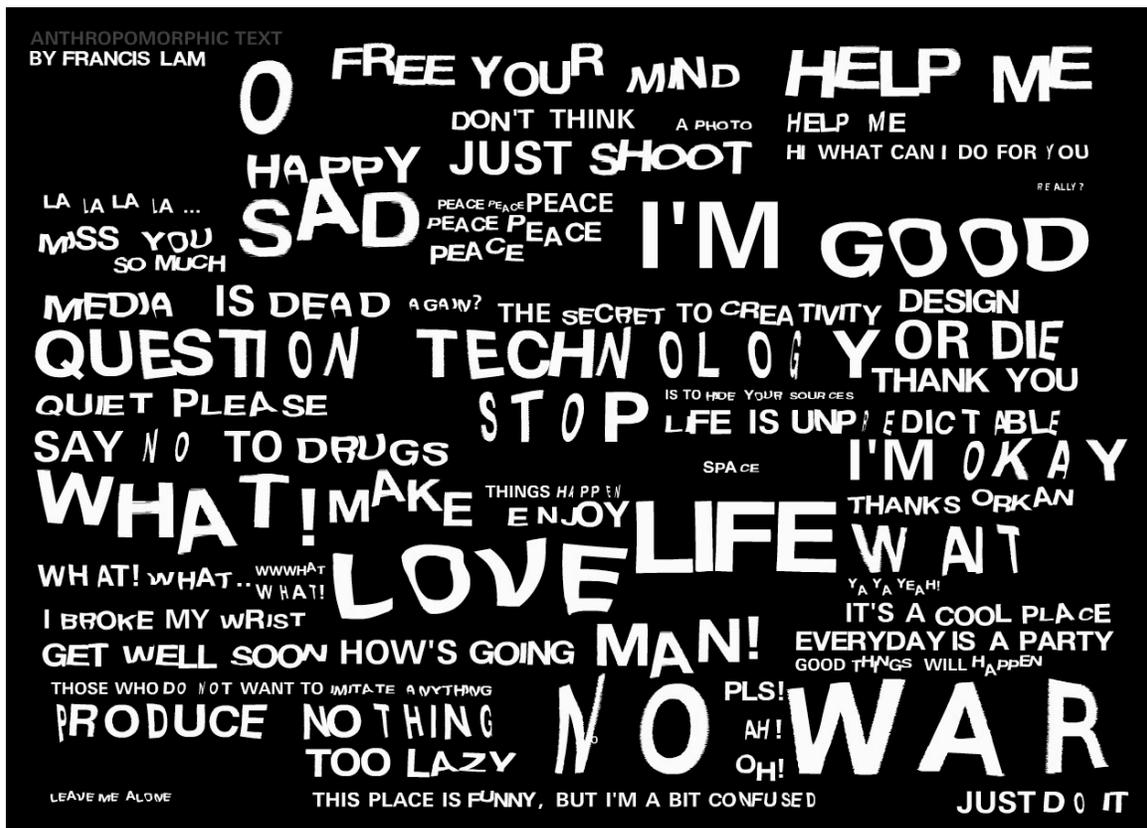


Figure 6. A snapshot of an animated *Anthropomorphic Typographic* landscape

Although many text-based chat users are used to include smiley faces, emoticons, animated graphics or even small photographs into their text messages to convey emotional meanings other than the content itself, I do not think it is the best way. Images usually contain too much information. A smiling face does not only convey happiness but also other unintended connotations, such as age or race. These extra cues may mislead or distract the viewer from getting the right emotion or meaning that is intended to convey. On the other hand, my approach

minimizes the unintended information by incorporating emotions into the text itself. These animated anthropomorphic letters not only display reminiscent body movements that convey emotional meaning, but also create a cleaner, denser and more consistent typographic landscape of dynamic conversations (see Figure 6).

2.3.3 Production Process

The technique I use to create the animated text is similar to rotoscoping, a traditional method for making animated cartoons. Firstly, I make 26 human-sized alphabet cardboards (A to Z) printed in capital **HELVETICA BOLD**, one of the most neutral modern san-serif typefaces. Then, I film with a video camcorder an actor (myself) wearing these cardboards and doing different action movements for each letter. Finally, I trace the recorded video, frame by frame, to recreate the animated text in Macromedia Flash (see Figure 7). The resulting animations are stunning in spite of the monotonous production process. Although the process can definitely be automated, the production process is manual for a good reason. The variation of motion in each letter and the human errors make the animated word more realistic and human-like.

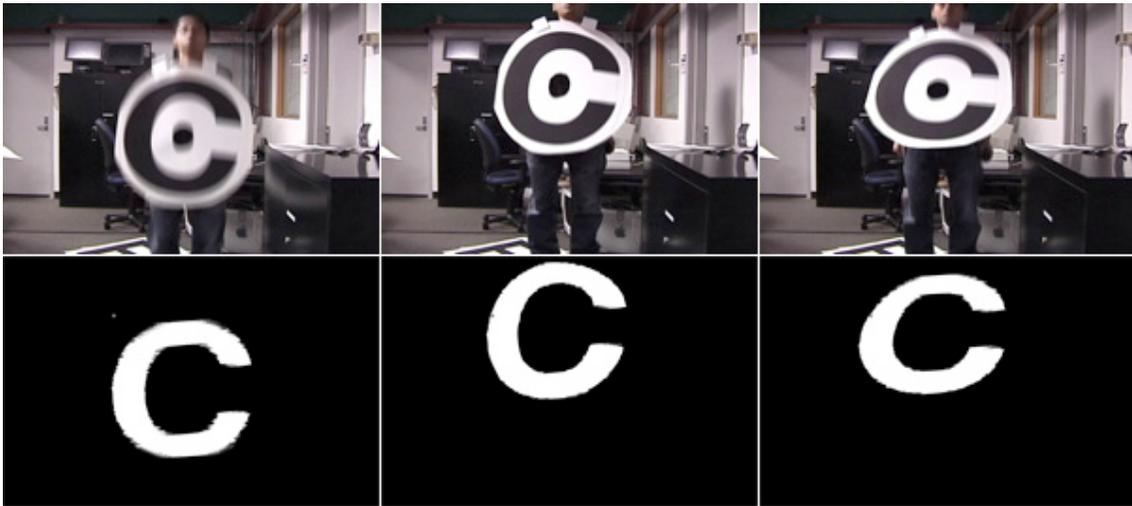


Figure 7. Tracing the hopping “C” from the video

Based on the six emotions identified by Paul Ekman [15] and the most common affects communicated over IM, I handcrafted six sets of anthropomorphic typefaces each consists of twenty-six capital letters. They resemble six different body movements which include:



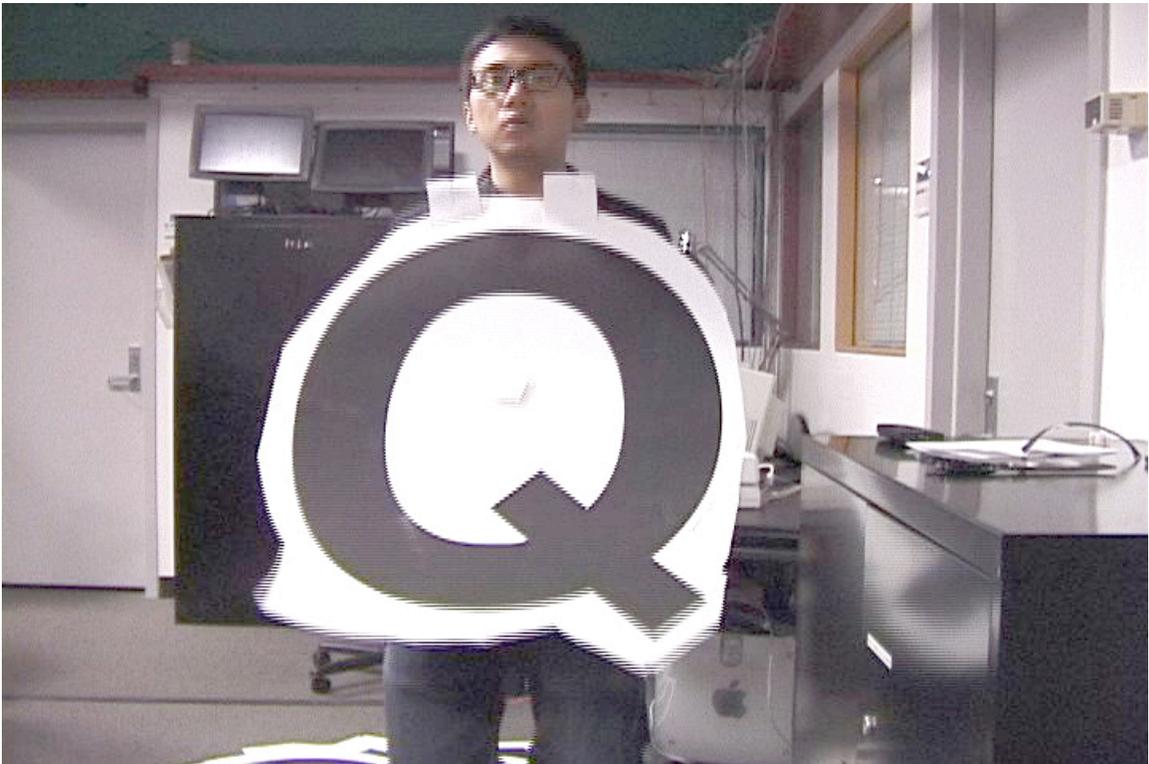
1. Drooping (being shy or depressed)



2. Swinging left and right (expressing negative opinion)



3. Hopping (being delighted)



4. Bobbing (being comfortable or in good mood)



5. Sitting (being relaxed or tired)



6. Swiveling vigorously (being excited or surprised)

These movements can be interpreted in different ways, but people are still able to guess what kind of emotion these letters mean because we are very responsive to human-like movements. For example a drooping letter may make you feel as if it is shy or depressed according to the context of the text. Indeed, despite of the tedious production process, I deliberately trace these frames by hand so that the resulting animations become more humanistic and realistic. I believe that our animated typefaces are appropriate and intuitive also for visualizing large-scale textual data such as email or newsgroup archives too. Some possible examples are:

Visualizing email contents: use of the typefaces to depict the most unique words or phrases in an email archive, so that at a glance, you can see a dynamic landscape of social pattern or mood of the conversation that you had in a certain period of time. The selection of the typeface to use can be determined by a semantic analysis method (e.g., [25]). An affect value is then obtained to map a particular set of animated typeface.

Text as personal portraiture: depicting a person's character from the text he/she wrote in a newsgroup. The anthropomorphic typefaces can lively visualize participants in different roles, ranging from a proliferated poster to a quiet reader, by typefaces resembling appropriate body movements.

To conclude, the main goal of producing anthropomorphic typefaces is to use subtle body movements to convey meaningful social information. The application of these animated letters in *Cheiro* are illustrated in the following chapter.

3 Building Cheiro

In this chapter I describe the actual *Cheiro* interface in detail and explain how it relates to the conceptual framework mentioned in the previous chapter. To begin with, I present an overview of the interface design and describe the general settings of the system. Secondly, I describe a user study conducted to find out how people would use mouse gestures to convey emotions. Thirdly, I explain the features of the novel expressive textual communication platform offered by *Cheiro*. They are animated text, non-linguistic communication, mouse gesture recognition and gesture pattern customization toolbox. Finally, I discuss briefly the development process and other versions of *Cheiro* produced during different stages of this thesis work.

3.1 Overview

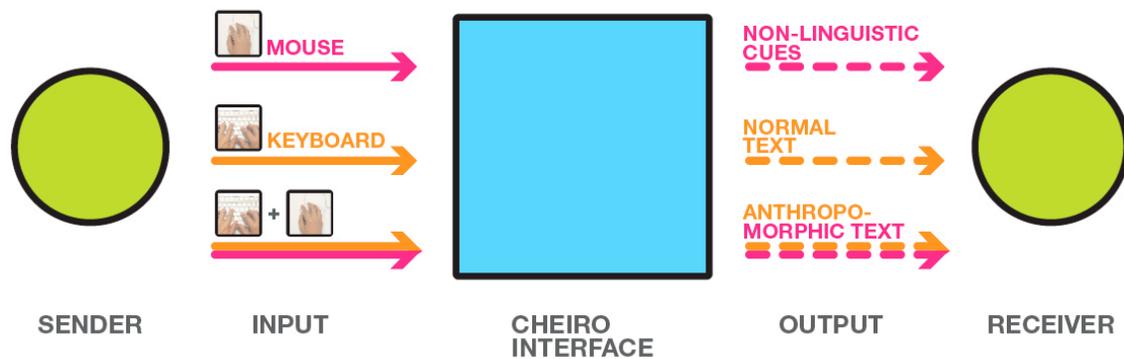


Figure 8. Input and output diagram of Cheiro

Cheiro is a text-based chat interface that recognizes users' mouse movements as the gestural input for expressing emotions and social cues in an online chatting environment. It explores the graphical possibilities of improving textual communication, and attempts to increase expressiveness in online chat environments using common input devices, such as keyboards and mice. It is written in Macromedia Flash MX and supported by Flash Communication Server for maximum of 10 users to chat simultaneously.

Figure 8 shows an input and output diagram of the system. The system takes three different types of input (the solid lines) from the user and generates different mediated outputs (the dashed lines) accordingly. Sole mouse movements are visualized as active non-linguistic cues to show presence, trigger interruption and obtain attention without conveying any verbal cues. These cues are depicted as circular mouse traces which are visible to both the sender and receiver. On the other hand, users can convey verbal communication with other participants by keyboard interaction like normal chat interfaces. Moreover, they can also add emotional information to the text by drawing with the mouse on the screen. These mouse gestures bring the static text to motion and enable it to convey nuanced contexts through anthropomorphized kinetic typography. The viewer sees the message as an animated text reminiscent to a human posture, which expresses emotional and social cues.

The *Cheiro* interface consists of four main parts (see figure 1): the text input area, output stage, conversation history, and gesture toolbox. Users type their message in the upper input area and the animated message is shown in the output stage below. When a new message comes in, the one on the stage is moved to the conversation history and archived in the database. Users can animate the text by a set of mouse gestures, which are simple one-stroke patterns derived from the study of emotive mouse gestures. They can also change the gesture patterns for different emotions or define their own ones by the Gesture Customization Toolbox.

In the following section, I discuss extensively on how each component of *Cheiro* works together, how they reflect the characteristics of the face-to-face conversation model, and how they answer the problems of the current textual communication.

3.2 A Study of Emotive Mouse Gestures

I conducted an online user study to understand how people would use mouse gesture to convey emotion. A mouse gesture is a way of combining mouse movements and clicks, which the system recognizes as a specific command. The goal of this study is to find if there is any distinctive pattern in these gestures, which help develop a gestural input for an expressive text chat system. Nearly 130 anonymous users participated in the study using an online interface for recording mouse gestures (see figure 16). They are asked to use a simple one-stroke mouse gesture to express their feelings or emotions upon 20 predefined scenarios. They are also requested to describe verbally the emotion or feeling they have conveyed and the reasons of using that mouse gesture.

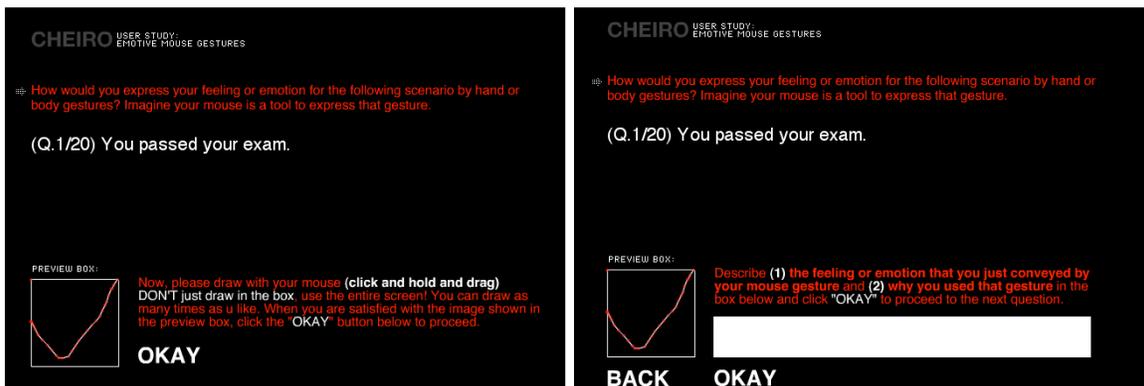


Figure 16. Interface for the emotive mouse gesture study

The scenarios are:

1. *You passed your exam.*
2. *You are chilling out with your best friends in a very cozy lounge.*
3. *Your dog passed away*
4. *It's a beautiful Sunday, but you have been watching TV at home for 5 hours...*
5. *Finally, you have finished your thesis...*
6. *Someone accuses you of being a racist.*
7. *You've got stuck in the same question of a problem set for 4 hours.*
8. *When you don't agree with something.*
9. *You find a snake in your jacket.*
10. *You just won a lottery of \$10,000!*

11. *When you're in love...*
12. *You're going to meet with your most favorite idol shortly.*
13. *Your girlfriend dumps you, your boss fires you, and someone breaks into your home and takes away everything all in one day!*
14. *After working for 24 hours non-stop...*
15. *Somebody steps on your new shoe.*
16. *Someone tells you something you're not interested in.*
17. *When you tease your best friend...*
18. *Your daughter is nominated for the Nobel Prize.*
19. *When you see a baby girl yawning...*
20. *Your friend's dog just died and you want to show him some sympathy.*

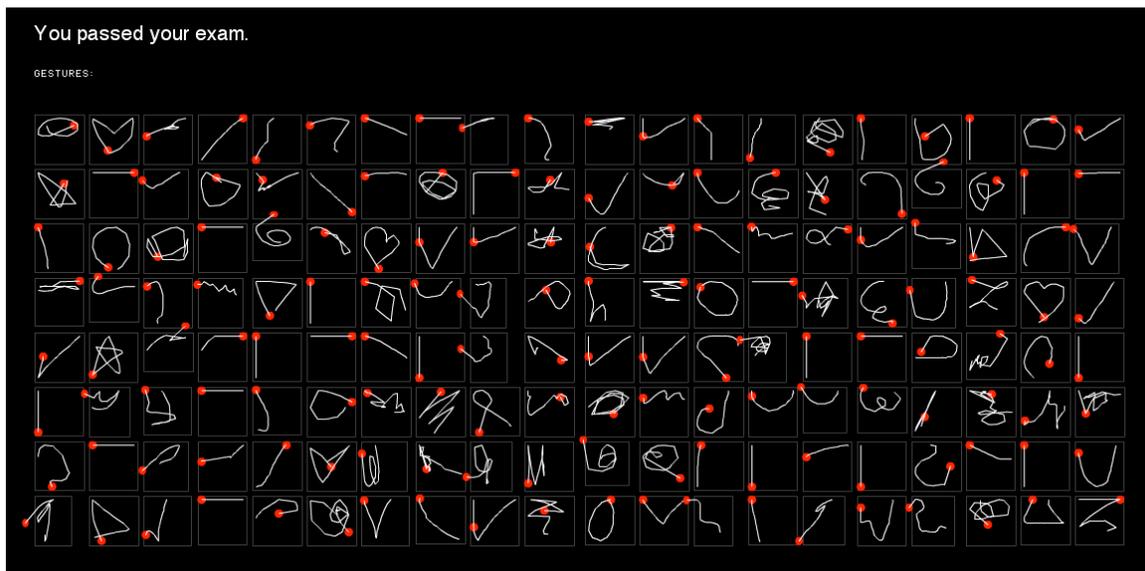


Figure 17. Mouse gestures of conveying the emotional feeling for “You passed your exam”

Apparently each scenario denotes a typical emotion, for example “*your dog passed away*” would probably induce sadness. However, instead of prompting the user with the emotional word “sadness” explicitly, a scenario is given for obtaining the mouse gesture because it is more natural and likely as in real chats or conversations. The user study interface collects participants’ mouse gesture pattern, speed and magnitude for each scenario. The reason for using the gesture and what emotion or mood has been conveyed are also recorded. Figure 17 shows some mouse gesture samples that the study collected for scenario 1.

Mouse gesture is a free-form input. Participants' emotive mouse gestures can be based on different reasons, mappings and metaphors. Sometimes it can be totally random too. Interestingly, the result indicates certain level of convergence of using a mouse gesture in some emotions. The following charts (figures 18-20) show in detail each emotion and its mouse gestures.

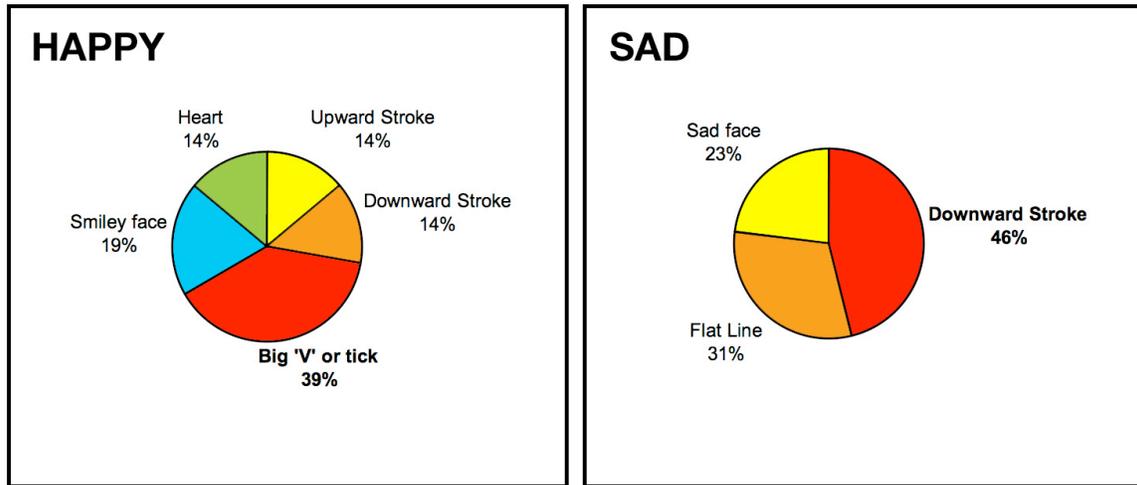


Figure 18. Mouse gestures for conveying “happy” and “sad”

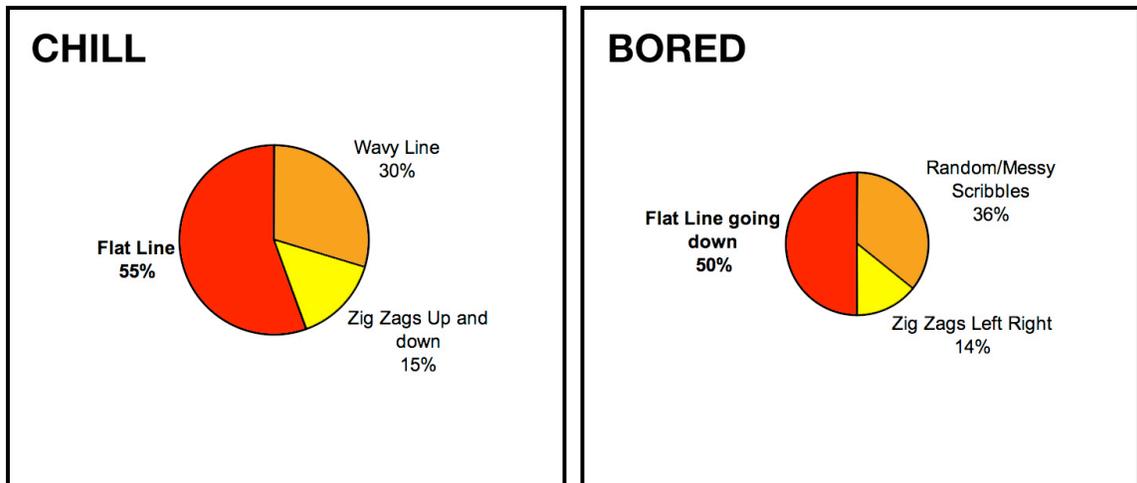


Figure 19. Mouse gestures for conveying “chill” and “bored”

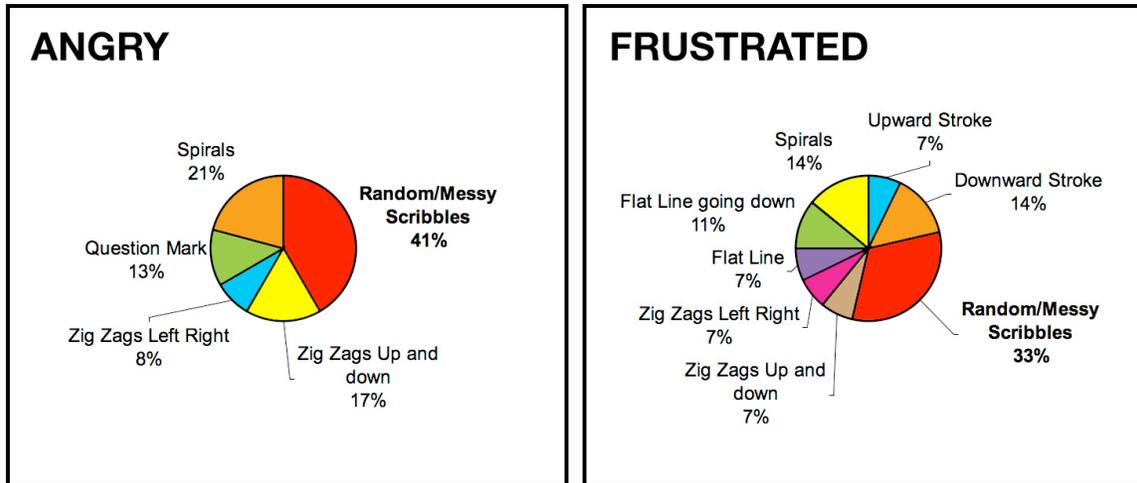


Figure 20. Mouse gestures for conveying “angry” and “frustrated”

The study produces important findings about how people use mouse gesture to convey emotion. For examples, users always use upward and downward strokes to express happiness and sadness respectively; flat lines are often used to represent bored and relaxed feelings; random scribbles are mainly appeared in angry and frustrated scenarios. However, the result cannot be directly used to design the gestural input for *Cheiro* due to the following reasons: (1) many mouse gestures are shared by multiple emotions, which even cannot be distinguished by the speed and magnitude features. (2) The same type of mouse gestures can be expressed differently by varying the speed, size and duration. Therefore, the results need to be compared and generalized in order to form a useful set of emotive gestures. The findings of the study do provide a substantial promise in using mouse gesture to express emotion. They act as a guideline for designing the default gesture set. The mouse gestures actually used in *Cheiro* are illustrated in chapter 3.

3.3 The New Expressive Textual Communication Platform

3.3.1 Animated Text

Text is undoubtedly the most important visual output in any text-based chat environments. In *Cheiro*, text can move like human postures. It can hop, sit or resemble other predefined body movements. The animations are controlled by the user's mouse gestures. Before getting into the animated text, I first give an introduction about how text appears on the interface.



Figure 9. Text areas in *Cheiro*. Left: Text Input Area, Message Stage, Chat History

Figure 9 shows a life cycle of the message “HELLO” appearing and fading out in the interface. When the user types with the keyboard, text is prompted in the grey *Text Input Area* on top of the screen. Then it moves downward to the *Message Stage* after the user presses the “Enter” key. That is the area where the most recent message sent from all users shows up. When a new message comes in, the current message on the *Message Stage* slides gradually further down to the *Chat History*, in where all messages accumulate. Then it keeps coming down and fading out when more messages are received. The flow of the text maintains a consistent progression pattern from the top of the screen to the bottom. It gives a coherent visual field for the users to get familiar with chat interface.

Users can choose to employ plain text or anthropomorphic text to convey their messages. As referring to the handwritten note example (see figure. 4), plain text can be animated by using the mouse to draw a particular set of patterns for different types of motion. The detailed information about the gesture recognition and the customization of patterns are explained in later sections. Now, I demonstrate some examples of using *Cheiro* in response to some conversational problems in normal online textual communication. The figures used to illustrate these examples are only stills from a series of animations captured from the actual interface (see figure 1).

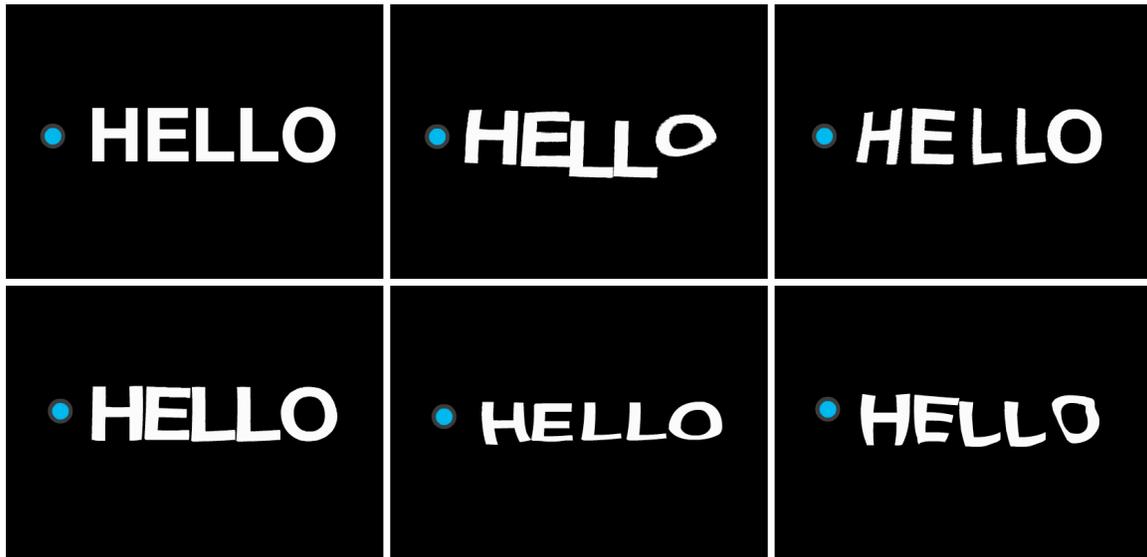


Figure 10. Snapshots from six animated anthropomorphic text examples.
Top left (clockwise): Neutral, happy, disagreeable, relaxed, cozy, sad

Figure 10 illustrates an example of the same message conveying different emotions by the anthropomorphic typefaces. As the still images cannot capture the full range of emotional effect, I highly recommend the readers to check the online demo at <http://www.media.mit.edu/~francis/cheiro2/> for the actual effect. The viewer not only conceives a neutral greeting message but also perceives some kind of emotional information from the sender. A collapsed, sad-looking “Hello” and a hopping “Hello” certainly create two distinguishable impressions about the sender. The animations are subtle but very humanistic. The viewers can easily form a mental representation from these moving letters to some social contexts or affective feelings, just like the point light stimuli example (see figure 5). Sometimes, these impressions can even go beyond words and create a totally different meaning (as explained in the next example). One may argue that the same result can be achieved by attaching a smiley face at the end of the text. However, by closely coupling the affect and the message typographically, not only can the sender convey more cues, the legibility of the text also increases.

As the network bandwidth increases, many text-based IM clients (e.g., [19], [27], [42]) provide a full range of emotional modifiers from tiny emoticons to enormous animations across the entire chat space. But, very sadly, these technological advancements may ruin the simplicity of text. One of the major visual design issues I concern about text-based interface is to maintain a

consistent and coherent typographic layout, yet in the same time, to provide a secondary channel for communicating non-linguistic cues. I hope *Cheiro* and *Anthropomorphic Typography* would light up this issue by designing a proper textual communication platform.

Go Beyond Words



Figure 11. Snapshots from two animated anthropomorphic text examples. Top: Cozy, sad

Figure 11 shows two examples of the non-linguistic cues outweighing the meaning of words. As mentioned in the previous chapter, we convey a significant amount of information non-verbally in face-to-face conversations. These non-verbal cues often correlate with the spoken language. However, sometimes, they don't. How can we establish this kind of communication in an online chat environment? The example on the top of figure 11 shows a message "It's a long day..." and it is animated in such a way that it looks very like sitting down on a couch. The motion suggests that the sender of the message is probably very tired and wants to lie down and take a rest. The animation works together with the word to convey an implicit meaning by a vivid visual representation, which is approximate to someone lying down on a couch.

The second example illustrates an interesting mismatch between the linguistic and non-linguistic cues. Verbally, it is a positive message "I'm fine", but the animation applied on the text is negatively depressed. From the user study of *Cheiro*, most participants' immediate response is, "Are you really fine?" Some others even say, "Why do you look so sad?" When similar situation happens in real life, we often suspect the spoken words, and tend to interpret the meaning and figure out the "truth" from the non-verbal cues. People always use sarcasm or convey

contradictory statements in different social contexts for different purposes. *Cheiro* aims to provide a channel for this kind of contradictory and submissive conversations.

Contextual Text

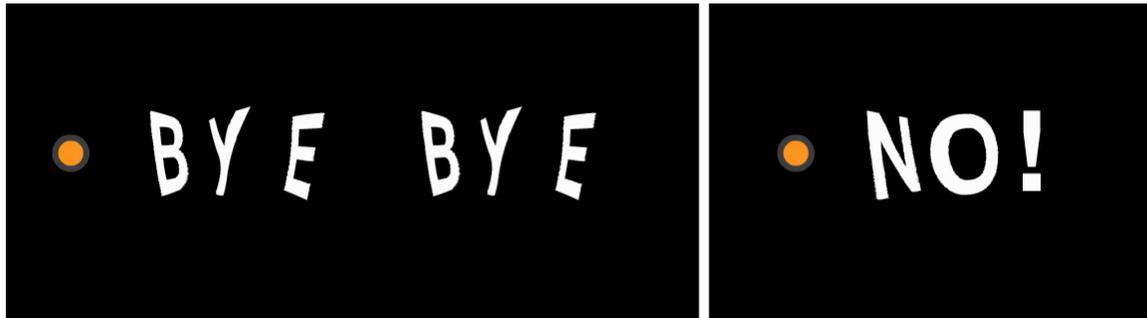


Figure 12. Snapshots from two animated anthropomorphic text examples. Left: Farewell, disagreeable

In figure 12, there are two different messages with the same animation effect. It exemplifies that the same motion can be interpreted differently when the verbal context changes. Imagine if you are in the middle of a discussion, you wave your hand after someone has proposed a ridiculous idea. Even without saying anything your hand gesture signals an immediate disagreement or a negative feedback to what that person has just said. However, in another situation, the same gesture can be used to wave somebody goodbye. The main point here is: although the animation (or the gesture) is ambiguous in some circumstances, the social and verbal contexts take part in shaping the meaning. Because visual cues are invisible in online text-based environments, new mappings (relating a motion to an emotion) need to take a long time to establish in human-computer interfaces. Anthropomorphic motions make these mappings familiar to the user, and hence provide a more intuitive interface for conveying emotions and social cues.

Mixed Emotions

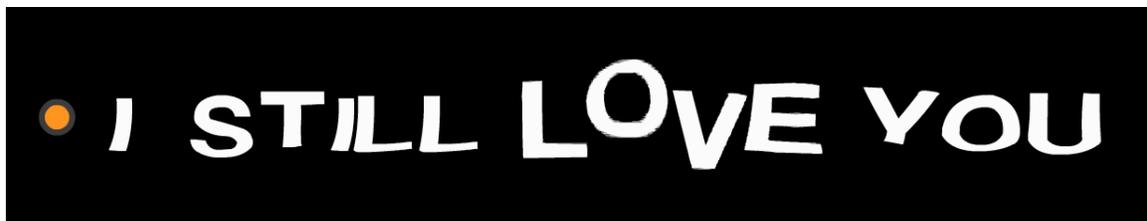


Figure 13. A snapshot from an animated anthropomorphic text example with mixed emotions

Cheiro also supports mixed emotions in one single line of message. Although the input for this to work is not very intuitive because the user requires extra click on each word to modify its animation, messages with mixed typefaces do convey complex emotions even within a short sentence like the example (see figure 13). This typographic approach maintains the consistency of the visual field and the legibility of the text, which may not be easily achieved by graphical means such as emoticons.

Visualizing the chat history

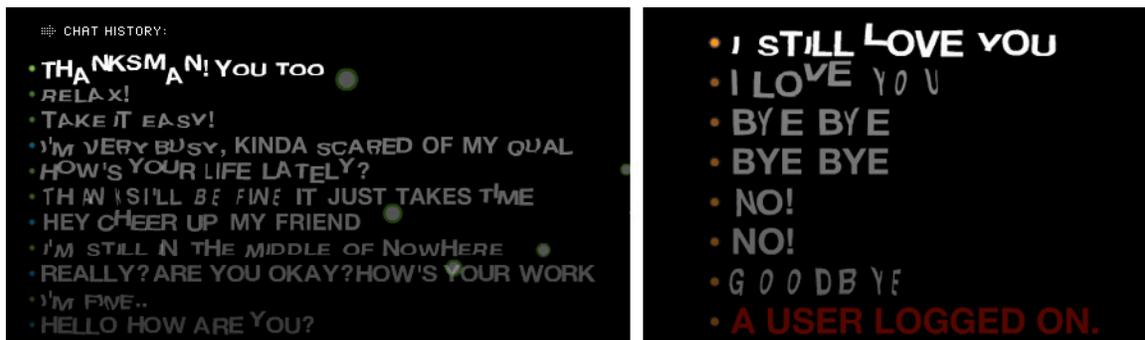


Figure 14. Chat history

Cheiro keeps a chat history that aggregates the messages from all participants in the lower half of the screen. It is not only a static archive of words, but also a visualization of all conversations engaged at the chat interface. This big piece of textual landscape (see figure 14) is moving and interacting among lines subtly. It is dynamic and expressive – denoting not only meanings in the semantic level, but also the emotional and contextual level. The opacity of each line of text depicts its recentness; older messages are darker and will dissolve into the background. When you look at it just at a glance, the motion of the text can abstractedly give you a sense of mood and atmosphere of the discussion – it may be delightful, vigorously or causal. It implies another possible application of the anthropomorphic text, which is visualizing a large piece of textual data and revealing its social pattern.

The above examples illustrate the actual application of *Anthropomorphic Typography* in the textual communication domain and how it supports non-linguistic cues to convey emotion and social context. In the next section, I discuss some examples of using graphical elements in *Cheiro* to depict non-linguistic information.

3.3.2 Non-Linguistic Communication Channel

In this section, I concentrate on the non-linguistic cues that are standalone, and do not accompany the language. As discussed in the previous chapter, *Cheiro* enables the exchange of non-linguistic cues over a text-based chat interface; these interactions are mainly facilitated by mouse movements (see figure 8). When a user logs on to the system, his/her mouse movements are transmitted over the network and visible to all other participants. By making these mouse traces noticeable, users are able to convey simple information or social signals non-verbally. The central idea is to use the mouse as a tool to visualize non-linguistic cues, which are normally invisible in textual communication.



Figure 15. Mouse traces

Figure 15 shows a view of two participants (represented by green and blue circles) chatting at *Cheiro*. When they move their mouse across the chat interface, translucent colored circles are drawn depicting the motion path on all participants' screens. The size of the circle is determined by the speed of the mouse. As these mouse traces are not intended for showing specific shapes and forms, they are not rendered very precisely (the mouse-tracking period is approximately 0.5 second). And they are also quickly diminished (by reducing size and opacity) to clear up the screen. These traces are more likely to be viewed as some ambient patterns representing the atmosphere of the chat room. When more people are in there, more colorful background patterns are generated to depict the online crowdedness. If more people are expressing their emotions through mouse movements, larger patterns will be generated as to show a more social and emotionally expressive chat space.

Not only can the subtle graphics show the activity level of the conversation, they can also be consciously used to signal some social cues non-verbally. Showing presence is one of these signals. When you talk to someone in person, his/her physical presence is visible and unquestionable. However, in an online text-based chat room, you are not sure if the one you are talking to is “there” unless they deliberately state that textually. In *Cheiro*, there is another channel to signal user’s attendance, which is by simply moving the mouse. Slightly moving the mouse can be used as a cue to prove a person’s presence and indicate his/her attentiveness. It can be a light and effortless gesture to say, “yes, I’m listening”, which is like nodding your head. It can also be used to trigger stronger conversational cues such as an interruption or act as a *conscious gesture* [10]. Depending on the context of the conversation, moving your mouse vigorously and continuously can attract people’s attention, request for interruption, or simply response to what has just been said. Sometimes, it can also be used to characterize anger, surprise and excitement followed by verbal explanations. Mouse movements, as rendered as translucent circles, are abstract, subtle and ambiguous but when this secondary channel opens up, I believe new forms of graphical communication behaviors will emerge accompanying with and without the text.

3.3.3 Mouse Gesture Recognition

Cheiro deploys a simple algorithm similar to Palm graffiti recognition [31] to identify patterns in a 2D plane. It is implemented in Macromedia Flash, to recognize the mouse gesture for animating text messages. The algorithm uses the directions of each point on a preset pattern and the input pattern to find a match mouse gesture. Each predefined gesture pattern is represented by a 30-point array; the input mouse gesture is compared with the vector of each point in the predefined pattern. The closest match above a certain threshold value is chosen to be the recognized gesture.

By choosing an appropriate mapping for the mouse gesture and animation, the user can feel more intuitive to add motions to the text and they can be more expressive in conveying emotions and social cues. As mentioned in the last chapter, I conducted a user study on how people would express their emotions by mouse gestures. The result shows some coherent patterns in conveying several emotions, such as a downward stroke for sadness or a flat line for tiredness. *Cheiro* follows these findings and uses the patterns as a default palette for animating the text (*Anthropomorphic Typography*). The following table (figure 16) illustrates the mappings between the mouse gestures and text animations. As the image can only show a still frame of an animated

text, readers are recommended to go to <http://www.media.mit.edu/~francis/cheiro2/> for a live demo of the system.

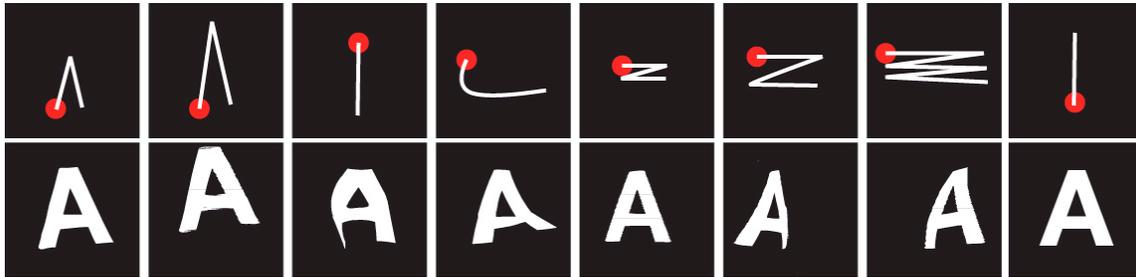


Figure 16. Mouse gestures (red dots denote the starting positions)

As I can only build a limited number of animated typefaces within the timeframe of this project, the latest version of the system can recognize eight different mouse gestures that are shown above. These stroke patterns result from the findings of the user study of emotive mouse gestures, human postures and hand gestures. They are not just one-to-one mappings. Similar text movements yield for the same type of gesture. For example, both the bobbing and hopping motions involve up and down movements, so they are all triggered by an arrow stroke with different magnitudes. Same concept applies to the zigzag stroke, a small zigzag makes the letter shake, a larger one makes it swivel like shaking one's head, and finally the largest zigzag stroke turns the letter swing vigorously like someone is dancing. By grouping similar animations with same kind of gestures, users can remember the mappings more easily. I hope the default mouse gestures can be intuitive to the user. The intention is to make a stronger linkage between the input and the output, and then users can pick up the gestural language more quickly.

In addition, *Cheiro* has a mouse gesture customization toolbox for users who want to define their own mouse gestures. Similar to hand gesture, which is very culturally and personally dependent [10], users can use different ways to express their feelings by using a normal mouse. From the study of emotive mouse gesture, over ten different strokes are identified just to convey the happiness among the participants. They all have different interpretations and reasons for using their gesture. Although some convergent patterns can be spotted for certain basic emotions, not everybody feel the mapping intuitive and natural, and sometimes may even be radically different from the default pattern. For instance, many participants use a downward stroke to express happiness mainly because they want to make an exclamation mark. The main function of the default gesture set is to provide an example of how the system works. Although it is ideal if we

can build a unifying emotive mouse gesture set for the majority, the customization toolbox is also important for making *Cheiro* more usable and intuitive to use.

3.4 Development Process

Before reaching the current approach for designing a new expressive textual communication system, a couple of other versions are designed and built in the process [33]. They are also very important components that constitute to many parts of the work done in this thesis. In the following paragraphs I present three major experimentations that I have gone through. They are an avatar chat program that conveys emotion by body postures, a synchronous gestural chat interface, and explorations of other input devices.

3.4.1 Avatar Chat

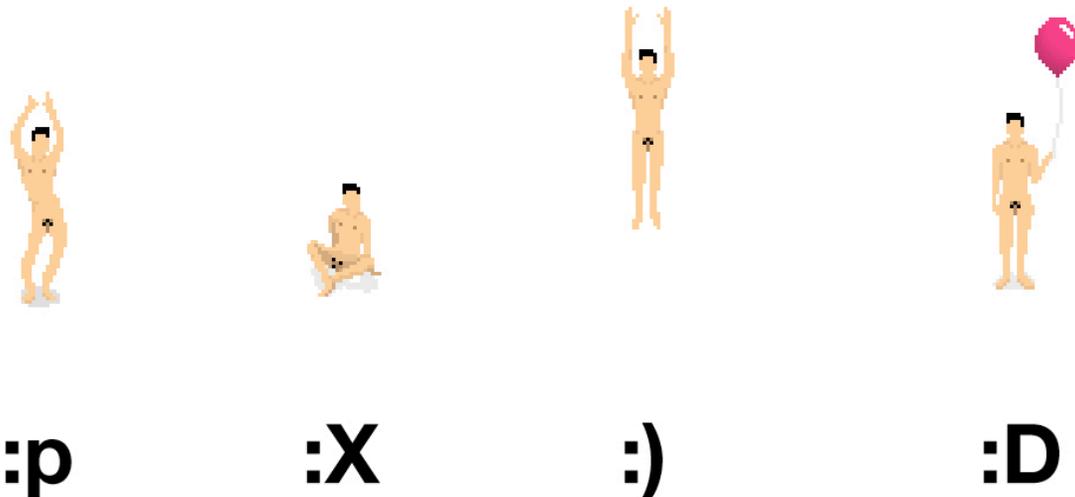


Figure 17. Avatar postures and their corresponding smiley faces

The concept of using body posture and movement to convey emotional information plays a crucial role in the development of the *Anthropomorphic Typography*. Yet, before I turn it into text, I build a mockup interface for examining the expressiveness of using these movements as in a chat environment. In this chat program, participants are represented by a human avatar (see figure 17), which can perform different postures according to the smiley faces appeared in the text message. This mockup system is similar to *Body Chat* [10] that uses an avatar's body

movement to convey subtle social cues and nuanced messages. However, as my goal is to develop an expressive textual communication channel, the separated text and avatar produces a broken link between the verbal and non-verbal cues. When a user sends out a message with a smiley face, the immediate visual output is the text message with a moving avatar in a separate view. It hinders the smooth comprehension of the text, which may cause legibility issues. Although an avatar-based approach is probably not a suitable solution for expressive text-based chat interface, it demonstrates the function of body movement as a communication cue for conveying emotion, and inspires the development of *Anthropomorphic Typography* – where text and human posture come together.

3.4.2 Synchronous Gestural Chat

The first version of *Cheiro*, after decided to use kinetic typography to convey emotion, is a “real-time” gestural input system. What I mean by real-time is that the gesture input and the synthesis of visual output are almost happened in the same time. Analogous to hand gesture, this system works as a baton to control the emotional rhythm of the text message. Figure 3 shows two people chatting in this system. To send a message, one needs to push out the letters by moving the mouse continuously (instead of just pressing the “Enter” key). The user’s gestural movement is translated into the moving letters according to the speed and direction of the cursor. Not only is the text affected by the gestural input, the circle underneath each letter is also animated in the same way. In addition, when a user is not typing anything, they can also move their mouse to generate semi-translucent circles to represent their presence, or signal interruption by rigorous movement.

Besides using mouse movements to directly change the graphical form and motion of the text, I have tried some indirect ways too. One of them is using an invisible force field to control the path of the moving text (see figure 19). In this experiment, users cannot use their mouse to alter the visual form or path of the message straightforwardly. They actually change the invisible force field when they move their mouse across the screen. As the motion of the text message is guided by the underlying force field, it moves differently when the field is changed. The visual effects of this experiment are beautiful, but users seem to have trouble using the interface - mainly because they expect immediate visual feedback and the lag between the mouse and text movements causes confusion. The interaction seems to be too subtle. This version focuses on the design aspects of using text as the main element in a computer-mediated communication. The input section of



Figure 20. Animating the text by hand gestures

A gaming glove controller is used to implement a hand-based gestural input system. In this version, users can use hand gestures to choose how the text be animated after the text entry using keyboard. This is an interesting experiment because users can feel like they are directly manipulating the text without using any device. However, it is quite uncomfortable to always wear the glove and type in the same time.

Through the exploration of using a variety of input devices, I realize that different textual communication platforms require different devices. For example, mouse is suitable for chat-room-like interfaces. It provides a high-resolution free-from gestural input, and supports a more complex gestural vocabulary in a dedicated chat environment. Keyboard is appropriate for simple and quick chat systems, where users do not need to change different devices for input text and emotional cues. Finally, a hand-based gestural system can be useful in short messaging services (SMS). Because mobile devices can already be used to detect gesture patterns (e.g., [40]), it is intuitive and easy to add in some emotional and social cues by just waving the phone before sending out the message. Although *Cheiro* is mainly developed for mouse use, it can easily adjusted to other input devices to support different kinds of textual communications.

3.5 User Study

The user study of *Cheiro* consists of two parts: (1) a mass online user study – for observing the general impression on the usability of the system, and (2) a qualitative analysis – for verifying the core functions of the system and collecting more detailed user feedback.

3.5.1 Online User Study

The system was put online and open to public since 4/10/2006 (<http://www.media.mit.edu/~francis/cheiro2/>). During a period of three weeks, 98 anonymous persons came to the site and sent out approximately 1100 messages excluding spams and system notifications. Unfortunately, very few conversational dialogues are traced mainly because most of them came to the site in different times. In most of the time there was only one user online, thus they use the system more like a public bulletin board than a chat room.

More than half of the total messages (64%) are animated, and they are triggered by about two-fifth of the participants. This is a quite impressive result considering no explicit instructions of using the gestural input is shown on the interface. The goal of this study is to observe users' behaviors upon a new textual communication environment without deliberate explanation of the new technology. Many participants, who discover the gestural input that causes the text to move, show positive feedback through their messages (e.g., “cool!”, “very interesting”). Participants who use the gestural input spend more time and leave more messages than those who do not. Some of them even asked their peers to join the chat. This user study shows that the participants are interested in using the anthropomorphic text, and mouse gesture is fairly easy to use as an input for a text-based chat system.

3.5.2 Qualitative Analysis

I conducted a survey asking questions about the function and the usability of *Cheiro* with 10 university students. They are 7 males and 3 females from 19 to 29-year-old. Firstly, they are given a brief introduction of the system and how each part works. Then they are called to chat with the interviewer over *Cheiro* for around 15 minutes. The interviewer asks them a list of in-depth questions after that chat, which I will present in the following paragraphs with some salient responses and analyses.

“How do you feel about Cheiro's appearance in general? Can you operate it with ease and efficiency?”

The interface in general

7 participants (out of 10) think the interface is easy to use and they can animate the text by mouse gesture with no difficulty. 9 of them like the interface design of *Cheiro*. Some of them point out that the text in the input area is too big; they cannot enter a long message in one line. Few participants suggest to have more messages in the chat history and clearer indication of who the sender is.

“Did you find using mouse gestures to add emotions to texts intuitive? Did the mouse gestures correlate well with the motions that you were trying to add to the text? How well did the motions added to the text correlate to the cues that you were trying to express?”

The gestural input

All participants find the mouse gestures quite intuitive and they can correlate well with the motions that they are trying to add to the text, especially the hopping motion with the arrow stroke (see figure 16). Nevertheless, three of them mention that the sitting down motion (made by a flat line) and the drooping motion (made by downward stroke) are less distinguishable, but they are satisfied with the gesture customization toolbox, which allows them to change the default gestures. In addition, one participant suggests that the animation can be triggered by the smiley punctuations automatically, because many people are so used to adding them at the end of a message. However, I think it would be better to keep it separated because users may just want to add smiley faces in their message but not animate it. Also, the gestural input design would be more consistent in this way.

“Did you find yourself more expressive when using Cheiro to communicate? Did you find yourself more easily understood

when using Cheiro to communicate? Did you find yourself interpreting others better?

“Did you find the animated characters useful in communicating with others online? Were they helpful in accurately expressing your feelings during the conversation? Did they help you convey ideas better? Do they accurately represent the non-linguistic cues that you were trying to convey?”

The expressiveness and the animated text

Most of the users say that the expressiveness of the text depends on the context. They find it expressive when it is used in the right place and the right time. All of the users are generally agree that the anthropomorphic text can convey what they want to convey, but some of them think sometimes it is hard to interpret other people’s intentions when they look at the animated text. Four participants state that the anthropomorphic text is most useful to amplify the meaning of the words. For example, some of them make the message “It’s great!” be hopping to exaggerate the positive mood. Two says that it helps highlight the stress of a message, just like intonation, by animating only one word in a sentence.

To test the ability of the system to convey non-linguistic cues beyond words, the interviewer starts off the chat with all 10 participants by saying “Hello” with a depressed anthropomorphic typeface. 7 of them reply with questions like “Are you okay?” or “What’s wrong?”. They can realize that the interviewer wants to convey a sad feeling without explicit wordings. The others do not aware of the effect of the sad-looking “hello” during the chat, but they all agree that it gives a sad feeling during the interview afterwards. This shows a promising result about the correctness of perceiving emotional cues from the animated text.

“How important is it that you feel the presence or attention of the others during a conversation (e.g., eye contact)? Did the moving color circles help in giving you a sense of presence during the conversation? Did you find them useful when trying to give others a sense of your presence? Did they help you interpret

the feedback from the other users? Did they provide information regarding other user's preferences and ideas?"

The non-linguistic channel

I deliberately leave out the part about the mouse traces during the briefing session. Thus, users can be tested to see if they would use it spontaneously as a non-linguistic channel. Probably because of the short chatting time, only 3 participants actually realize that their mouse traces coexist in the interviewer's screen. They move the traces randomly all over the screen and follow the interviewer's traces as a game. Other participants think that the colored circles are just visual effects and not particularly depicting anything. After explaining the function of these circles in the interview, most of them agree that it can be used to show presence or direct attention. One participant mentions that the mouse traces can also be used as a pointing device to highlight some messages from the chat history. Five of them consider the color of the circle is not enough to differentiate who the sender is. They say that it would be worse if more participants chat in the same time.

In conclusion, the majority thinks the gestural input is intuitive and easy to use. They are also able to understand quite correctly the meaning of the animated text, especially in the right context. The user study also shows that the animated text is able to convey extratextual meanings. Some parts of the interface can be designed better, for example, the user indication of the mouse traces and the chat history.

4 Future Work

In this chapter, I outline the future directions I envision for the system. The chapter is comprised of two parts: improvements and future directions. In the first section, I describe some immediate improvements that will make *Cheiro* become a robust and complete chat interface. These suggestions are mainly subject to the current input and output framework. As reflected in the user study, participants think that the gestural input is easy to use and intuitive to convey emotional information. Most of the difficulties are found in the interpretation of the visual output (text and graphics), which will then become the main focus of this section. The second section is about the future vision and applications of making expressive textual information in general. I introduce plausible applications in the realm of textual communication by using the fundamental concept of *Cheiro*.

4.1 Improvements

One of the immediate focus areas will be the automation of the animated text creation. You can probably guess that the current creation process is very tedious from my facial expressions in the making of *Anthropomorphic Typography*. The users also want to have a large set of vocabularies to convey more emotions. My plan is to incorporate motion capture or computer vision techniques to dynamically generate more animated typefaces. With a computational creation process, the system will be able to include lowercase letters, punctuations, numbers and symbols, or even support other alphabetical languages. Making the process digital does not mean that all the typefaces are generated straight from the computer. I still want to maintain the humanistic and

inaccurate features of human posture and movement. It implies that I will develop a system that can dynamically capture someone's motion and apply it into the text - rather than building a program to simulate human movements.

The current system only supports eight different basic emotions. Many users find that it is not enough to represent what they feel during their online conversations. Richer input and output vocabularies are needed for the system to convey more non-linguistic cues. Moreover, these cues can be divided into two groups: emotional and contextual cues. The first group contains typefaces resembling postures with an emotional implication, such as a hopping letter for excited and happy feelings. The second group contains postures, which are contextual and functional. For example, they can be motions for greetings or doing a particular work. Users can use these contextual typefaces to show social contexts non-verbally. The group of a typeface should be distinguishable to prevent misunderstandings. For instance, a "Hello", which shows a lie-down-and-then-sleep motion, can mean you are tired as an emotional cue or represent you are actually sleeping as a contextual cue. One way to identify which group it belongs to is by color.

A larger set of output vocabularies implies that more mouse gestures are needed. In designing new mouse gestures, it is important to keep them simple and easy to remember. One way to achieve that is by correlating the mouse gesture with the output animation. Drawing an upward stroke is more likely to be remembered than a flat line for making the word jump up. If there are enough samples in the gesture space, users can learn the gestural input mechanism faster. They can arbitrarily make a mouse gesture and the system is able to return a motion, which is similar to the input. They can just explore different kinds of animated typefaces by drawing out random patterns even though they do not know much about the mappings.

In short, the immediate improvements and additions I envision for *Cheiro* are the development of a computational process for making the *Anthropomorphic Typography*, and building a larger set of input and output vocabularies, which can enable richer emotional expressions.

4.2 Future Directions

In this section I situate the work done in *Cheiro* within a broader research domain. Three scenarios are described to illustrate how textual communication can be made more expressive in different application domains and their benefits.

Personalized Anthropomorphic Typography

People have subtle and prominent differences in the ways of expressing themselves. They use different languages, gestures, body movements to convey messages and communication nuances. If the typeface of text messages can also reflect their real body movements, it will be more expressive and able to show the character of the person. Having the production process of *Anthropomorphic Typography* computational, personalized animated text can be generated at the user side. With simple motion capture device such as a webcam, users will be able to create very personal kinetic typefaces, which not only convey emotion but also personality.

Expressive Short Messaging Service (SMS)

Mobile phone is now part of our life. Mobile devices send out millions of text messages everyday all over the world. For example in China, SMS is extremely economical (approx. US\$0.0065 per message). It is not surprising that SMS has become a major communication platform in some cities. However, complicated text entry is still a significant issue that hinders the expressiveness of text messages. Most mobile phones do not have a built-in keyboard. Even if they have, the keys are too tiny for many adults. Many people try very hard to compact their words by using different abbreviations and struggle with punctuations and symbols, let alone using emoticons or adding emotional and social cues.

That leads to a huge opportunity in developing a mobile version of *Cheiro*, which can make SMS more textually expressive. There are already mobile phones with built-in accelerometer (e.g., Samsung SCH-S310) or camera [40], which can recognize hand movements. Enabling the gestural input mechanism in *Cheiro* be conducted by the phone itself, the user can make hand gesture to animate the text by simply moving the phone. Fagerberg et al. study how people can convey emotions by gesture using mobile phones. They implemented a system [17], which allows

users to send emotional SMS by different body postures. The recipient sees a text message with colored background representing different moods. *Mobile Cheiro* can be along the same line as their system but with much stronger expressive power in the visual output, because the *Anthropomorphic Typography* can convey more complex social cues and emotional meanings than just color. Imagine when you enter a short message: “I’ll be there in a minute”, you shake your phone up and down vigorously and the text becomes hopping up and down too. The recipient would probably get some sort of feeling that you are in a rush and making your way to be there soon. The text not only conveys subtle contextual information, it is also more coherent with the gesture input. In short, this *Cheiro* application will make SMS be more expressive emotionally and contextually with an intuitive and easy gestural input method.

Expressive Subtitle

Textual communication is not only limited in chat interfaces. For example, television or movie subtitle is also a form of textual communication, which is one-way and broadcasted to a large group of people. I foresee the possibility of using *Anthropomorphic Typography* in making more expressive subtitles especially for deaf people. Although they may be able to guess the context of the scene and the emotional state of the speaker by visual cues, it is still difficult to perceive subtle and invisible non-linguistic cues such as intonation or irony. The cognitive load to read the subtitle and watch the scene simultaneously may also lower the ability to perceive these cues. Often times, actors or actresses use contradicting facial expressions and voices to convey sarcastic meanings. I hope the approach of *Cheiro* will be able to make subtitles more emotional and dramatic.

I have presented three possible future applications that *Cheiro* can be further developed on. Many more domains involving text, such as speech-to-text applications or interactive text billboards, can take on this approach. By designing for different contexts and mediums, text can be more than just words.

5 Conclusion

“It is impossible to speak a word orally without any intonation. In a text punctuation can signal tone minimally: a question mark or a comma, for example, generally calls for the voice to be raised a bit. Literate tradition, adopted and adapted by skilled critics, can also supply some extratextual clues for intonations, but not complete ones.” – Walter J. Ong [30]

I have presented a gestural-based chat system, *Cheiro*, for communicating expressive textual information over computer-mediated environments. It is an attempt to bring in some intonational quality to the text. Its main function is to provide a channel for conveying non-linguistic cues, which amplify the meaning of words through visual elements. I have also introduced a new method of making expressive kinetic typography called *Anthropomorphic Typography*. It resembles human movements to convey social cues and emotions.

The user study of *Cheiro* shows that a gestural-based interface (using mouse) is viable for making expressive textual communication. Most of the participants think the mouse-based gestural input in *Cheiro* is intuitive and easy to use. The system provides a series of design explorations in the area of kinetic typography, graphics, and user-centered design for conveying emotions. I have also conducted an emotive mouse gesture study, which provides important design implications for gesture-based interfaces, especially for conveying emotions by mouse movements. These results are also used in *Cheiro* as a set of default mouse gestures.

I have discussed the theoretical framework in light of the current implementation of *Cheiro*. I compare the input of the system with the exchange of non-linguistic cues in face-to-face communication. Sending these cues in real-life conversation is spontaneous and direct. It provides an expressive channel to convey meaningful information. In *Cheiro*, however, these cues are carried out by an input device and visualized through animated text and graphics. The mappings of the gestural movements to the visual representations are mostly subjective. Can the sender efficiently correlate the visual output with the gestural input? Can the receiver correctly interpret what the output means? In *Cheiro* I have addressed these questions.

The visual output of the system is a crucial part of this thesis. The challenge is to create an expressive platform (in terms of graphics and kinetic typography), which shows the sender's real intention that can be understood easily by the receiver. The formation of a set of lucid visual vocabularies for gestural input is also challenging. The flow of the conversation should be taken into account too. In addition, building a real-time communication interface with a considerable amount of textual and visual information is a complex design problem. *Anthropomorphic Typography* and mouse traces are the methods used in *Cheiro*. Body movements and postures shape the animation of the text, they make the text convey subtle social cues, which can be easily interpreted by people. The participants of the user study are able to understand the meaning of the animated text correctly, especially in the right context. It shows that the animated text is able to convey extratextual meanings [30].

Finally, I have described the improvements and future directions of this thesis work. *Cheiro* is a project that explores the possibilities of expressive typography. I have also showed that it is able to improve the non-linguistic communication in an online chat environment based on the user study. After all, text should be more playful, expressive, emotional and beyond the meaning of words.

Bibliography

1. Allison, T., Puce, A., McCarthy G. Social perception from visual cues: role of the STS region. TICS, 2000.
2. Argyle, M. and Cook, M. Gaze and Mutual Gaze. Cambridge University Press, Cambridge, 1976
3. Arnheim, R. Art and Visual Perception: A Psychology of the Creative Eye. University of California Press, CA, 1974
4. Bartram, L. Perceptual and interpretative properties of motion for information visualization. Technical Report CMPT-TR-1997-15, School of Computing Science, Simon Fraser University, 1997.
5. Bartram, L. Perceptual and Interpretative Properties of Motion for Information Visualization. In Proceedings of the 1997 workshop on new paradigms in information visualization and manipulation, 1997
6. Bartram, L., Ware, C., Calvert, T. Filtering and Integrating Visual Information with Motion. In Proceedings on Information Visualization, 2001
7. Birdwhistell, R. In Kinetics and Context: Essays on Body Motion Communication, University of Pennsylvania Press, 1970
8. Bodine, K., and Pignol, M. Kinetic Typography-Based Instant Messaging. In Proceedings of CHI03, 2003
9. Bruce, V. and Young, A. In the eye of the beholder: The science of face perception. Oxford University Press, UK, 1998
10. Cassell, J., A Framework for Gesture Generation and Interpretation. In Computer Vision in Human Machine Interaction, R. Cipolla and A. Pentlan, eds. Cambridge University Press, New York, 1998
11. Chalfonte, B.L., Fish, R.S. and Kraut, R.E. Expressive richness: A comparison of speech and text as media for revision. Proceedings of CHI '91, 1991
12. Cheng, K. Designing Type. Yale University Press, New Haven, 2005
13. Donath, J. Mediated Faces. In M. Beynon, C.L. Nehaniv, K. Dautenhahn (Eds.). Cognitive Technology: Instruments of Mind Proceedings of the 4th International Conference, CI 2001, Warwick, UK, 2001

14. Efron, D. *Gesture and environment*. New York: King's Crown Press, 1941
15. Ekman, P. *Emotion in the Human Faces*, Pergamon Press Inc., New York, 1972
16. Ekman, P. and Friesen, W. The repertoire of nonverbal behavioral categories – origins, usage, and coding. *Semiotica*, 1969
17. Fagerberg, P., Ståhl, A., and Höök, K. Designing gestures for affective input: an analysis of shape, effort and valence. In *Proceedings of Mobile and Ubiquitous Multimedia, MUM'2003*, 2003
18. Ford, S., Forlizzi, J., and Ishizaki, S. Kinetic Typography: Issues in Time-Based Presentation of Text. In *Proceedings of CHI97 Conference Extended Abstracts*, 1997
19. ICQ, <http://www.icq.com>
20. Isaacs, E. and Tang, J. What Video Can and Can't Do for Collaboration: A Case Study. In *Multimedia Systems*, 1994
21. Jesse, R. Motion enhanced visualization in support of Information fusion. In *Proceedings of CISST'2001*. 492-497. CSREA Press, 2001
22. Johansson, G. Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 1973
23. Kaliouby, R. E. and Robinson, P. FAIM: Integrating Automated Facial Affect Analysis in Instant Messaging. In *Proceedings of the 9th International Conference on Intelligent User Interface*, 2004
24. Lee, J., Forlizzi, J., and Hudson, S. The Kinetic Typography Engine: An Extensible System for Animating Expressive Text. In *Proceedings of UIST02*, 2002
25. Liu, H. and Singh, P. ConceptNet: A Practical Commonsense Reasoning Toolkit. *BT Technology Journal* 22(4). Kluwer Academic Publishers, 2004
26. McNeill, D. *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press, 1992
27. Microsoft MSN Messenger, <http://messenger.msn.com>
28. Morris DiMicco, J., Lakshmipathy, V., and Fiore, A. T. Conductive Chat: Instant Messaging With a Skin Conductivity Channel. In *Poster Presentation, Conference on Computer Supported Cooperative Work (CSCW '02)*. New Orleans, LA, 2002
29. Mueller, F. and Lockerd, A. Cheese: tracking mouse movement activity on websites, a tool for user modeling. In *CHI '01 extended abstracts on Human factors in computing systems*, 2001
30. Ong, W.J. *Orality and Literacy The Technologizing of the Word*. Methuen & Co. Ltd, London, UK, 1988
31. Palm graffiti, <http://www.palm.com/us/products/input/>
32. Picard, R. *Affective Computing*, MIT Press, Cambridge, MA, 1997
33. Preliminary experiments of Cheiro. <http://www.media.mit.edu/~francis/projects/cheiro/>, 2005
34. Rimé, B. The elimination of visible behavior from social interactions: Effects of verbal, nonverbal and interpersonal variables *European Journal of Social Psychology*, 1982

35. Rosenberger, T., MacNeil, R. Prosodic Font: Translating Speech into Graphics. Computer-Human Interaction Conference of the ACM, Pittsburgh, PA, 1999
36. Lasseter, J., Principles of traditional animation applied to 3D computer animation. ACM SIGGRAPH Computer Graphics, v.21 n.4, p.35-44, 1987
37. Vertegaal, R. The GAZE Groupware System: mediating Joint Attention in multiparty Communication and Collaboration. In Proceedings of CHI '99, Pittsburgh, PA, 1999
38. Viégas, F. Visualization Email Content: Portraying Relationships from Conversational Histories. In Proceedings of CHI2006, 2006
39. Wang, H., Prendinger, H. and Igarashi, T. Communicating Emotions in Online Chat Using Physiological Sensors and Animated Text. In Proceedings of CHI 2004, 2004
40. Wang, Q.Y., Canny, J. F. TinyMotion: Camera Phone Based Interaction Methods. In Proceedings of CHI2006, 2006
41. Williams, E. In Experimental comparisons of face-to-face and mediated communication: A review. Psychological Bulletin, 84, 963-976, 1977
42. Yahoo! Messenger, <http://messenger.yahoo.com>
43. Yuasa, M. Saito, K. and Mukawa, Naoki. Emoticons Convey Emotions without Cognition of Faces: An fMRI Study. In Proceedings of CHI2006, 2006