

# The Conductor Interaction Method

DOROTHY RACHOVIDES

University of Surrey

and

JAMES WALKERDINE and PETER PHILLIPS

Lancaster University

---

Computers have increasingly become part of our everyday lives, with many activities either involving their direct use or being supported by one. This has prompted research into developing methods and mechanisms to assist humans in interacting with computers (human-computer interaction, or HCI). A number of HCI techniques have been developed over the years, some of which are quite old but continue to be used, and some more recent and still evolving. Many of these interaction techniques, however, are not natural in their use and typically require the user to learn a new means of interaction. Inconsistencies within these techniques and the restrictions they impose on user creativity can also make such interaction techniques difficult to use, especially for novice users.

This article proposes an alternative interaction method, the conductor interaction method (CIM), which aims to provide a more natural and easier-to-learn interaction technique. This novel interaction method extends existing HCI methods by drawing upon techniques found in human-human interaction. It is argued that the use of a two-phased multimodal interaction mechanism, using gaze for selection and gesture for manipulation, incorporated within a metaphor-based environment, can provide a viable alternative for interacting with a computer (especially for novice users). Both the model and an implementation of the CIM within a system are presented in this article. This system formed the basis of a number of user studies that have been performed to assess the effectiveness of the CIM, the findings of which are discussed in this work.

Categories and Subject Descriptors: H.5.2 [Information Interfaces and Presentation]: User Interfaces—*Input devices and strategies, interaction styles, prototyping*

General Terms: Design, Experimentation, Human Factors

Additional Key Words and Phrases: Human-computer interaction, gaze- and gesture-based interfaces

## ACM Reference Format:

Rachovides, D., Walkerdine, J., and Phillips, P. 2007. The conductor interaction method. *ACM Trans. Multimedia Comput. Comm. Appl.* 3, 4, Article 27 (December 2007), 23 pages. DOI = 10.1145/1314303.1314312 <http://doi.acm.org/10.1145/1314303.1314312>

---

## 1. INTRODUCTION

Humans are increasingly using computers as part of their everyday lives. Computers can be used to facilitate human activities such as work, communication, and entertainment. Most of our activities today involve computers, either directly or in a supportive role.

---

Authors' addresses: D. Rachovides, Digital World Research Centre, University of Surrey, Guildford GU2 7XH, Surrey UK; email: [drachovides@acm.org](mailto:drachovides@acm.org); J. Walkerdine (contact author), P. Phillips, Computing Department, Lancaster University, Bailrigg, Lancaster LA1 4YW UK; email: [j.walkerdine@lancaster.ac.uk](mailto:j.walkerdine@lancaster.ac.uk), [phillips@comp.lancs.ac.uk](mailto:phillips@comp.lancs.ac.uk).

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or direct commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or [permissions@acm.org](mailto:permissions@acm.org).

© 2007 ACM 1551-6857/2007/12-ART27 \$5.00 DOI 10.1145/1314303.1314312 <http://doi.acm.org/10.1145/1314303.1314312>

ACM Transactions on Multimedia Computing, Communications and Applications, Vol. 3, No. 4, Article 27, Publication date: December 2007.

Over the last 30 years much research has been carried out into developing methods and mechanisms to assist humans in interacting with computers. As computer technology has evolved, so have the techniques used to interact with it, some more successful than others. A key example of this has been the general shift away from command line interfaces to more visual interfaces such as window, icon, menu, pointing (WIMP) devices, as a result of the increased resources being made available to interface designers.

As computers are integrated more into our lives and used to perform a larger variety of tasks, humans have to learn how to use these new systems efficiently. In the past there had been cases where the effort required to learn the use of a system had outweighed the benefits of its use [Shneiderman 2005], for example, from command-prompt-based systems (where a multitude of available commands are typically hidden from the user) to complex 3D modeling packages (where the user is typically presented with a plethora of options). The learning overhead involved is one reason why users are often reluctant to try new applications or interaction techniques.

A number of techniques have been developed for interacting with computers [Dix 2004], some of which have been used for many years and continue to be used (e.g., command-line interfaces, menus, and form-fills), while others are more recent and still evolving (e.g., (pseudo-) natural language and question/answer dialogs). Finally, there are the widely-used WIMP interfaces (e.g., as found in Mac OS), point-and-click interfaces (as used in web browsers), and 3D interfaces which are widely used in computer games and virtual reality environments [Benford 1995].

Such techniques certainly can provide benefits for interaction, as demonstrated by their continued use. Nevertheless these techniques do have certain failings, as described next more specifically.

- Existing interaction techniques are not natural and typically require the user to learn a new interaction method.* As most techniques previously mentioned are significantly different from each other and from human-human interaction, users need to devote a significant amount of time to learn to use any given interface. Most of them heavily rely on memorization of commands, which increases the learning curve.
- Existing interaction techniques can be inconsistent.* This is an issue that arises with interfaces that are of the same type, but having small differences. They can be confusing for any user, no matter how experienced.
- Existing interaction techniques can restrict creativity and be oversimplistic.* Existing interaction techniques are task driven and tend to be very structured towards that end. This approach can be very useful in guiding the user through a specific task, but can also be seen as an obstacle in the expression of the user's creativity.

This article describes the conductor interaction method (CIM), an alternative that draws upon existing human-human interaction techniques in order to address some of the shortcomings exhibited by current human-computer interaction (HCI) approaches. This method exploits the gaze- and gesture-based interaction with which users can perform activities within a novel interaction environment. A particular aim for the CIM is for it to provide a more natural interface than existing interaction methods, while having a low learning overhead. It is envisioned that this method will be particularly beneficial for novice computer users, who will be able to utilize interaction techniques those with which they are more familiar (i.e., gaze and gestures), rather than having to learn totally new ones.

The proposed CIM is also supported by recent advances and reduced costs in the technology that can be used for capturing human-human interactions. The development of such devices, whose cost was prohibitive in the past, has made it feasible to consider the adoption of human-human interaction methods into HCI. As a result, it is likely that there will be a growing need for research into new techniques and interfaces to support this.

The CIM has been realized through the development of a presentation conductor system: a tool that can be used to develop multimedia presentations. As well as providing a means to demonstrate the interaction method, the system was also used as the basis for a comprehensive evaluation. The article proceeds by providing a discussion of gaze/gesture-based interaction and related work, before moving on to describe the CIM. The presentation conductor is then presented and employed to demonstrate the method in use. Finally, the evaluation that has been performed is discussed, with key findings and potential areas of future work highlighted.

## 2. GAZE- AND GESTURE-BASED INTERACTIONS

Eye gaze can be regarded as the instantaneous point-of-regard in the visual environment [Karn 2002] and is both a communication channel and a body-language signal for the recipient. Eye contact, when two interacting people look at each other in the area of the face [Argyle 1996], is used extensively in conversation. Through these channels we can identify issues such as whether our audience is listening, bored, interested, confused, etc. By directing our eye gaze and establishing eye contact with another party we can direct the communication towards that person or give a cue for that person to start speaking. For example, a teacher may indicate a readiness to take questions from students by making eye contact with them [Kendon 1992].

Bauml and Bauml [1997], in their *Dictionary of Worldwide Gestures*, define a gesture as “a posture or movement of the body or any of its members, that is to be understood to be meaningful.” Gestures themselves may or may not accompany speech, but are as ephemeral as the spoken word. They are either a visible accompaniment to, or a substitute for, speech or action. Gestures, particularly hand gestures, play an integral part in our everyday human-human communication. Humans use gestures both consciously and subconsciously and they are used either on their own or to support speech. Gestures can be the main communication channel or a back channel.

### 2.1 Using Gaze as Part of HCI

In HCI, the main benefit of gaze is to determine the focus of the user’s attention, or, more specifically, to determine what the user is looking at on the screen. The measurement of a user’s attention can be used, for example, to help evaluate an interface, or to provide another input channel with which to control an interface.

One area in which research has focused is in the development of systems that use gaze to support communication, for example, to determine who a user is listening or speaking to within a conversational system [Vertegaal 2001, 1999]. Although gaze may seem to be most suitable for supporting communication, it can also be used as a form of simple interaction, with numerous systems being developed that have shown that object selection, pointer movement, and “clicking” can be satisfactorily achieved [Instance 1994; Farid 2002a, 2002b; Gips 1996; Salvucci 1999; Silbert 2000; Zhai 1999].

However, gaze is not ideal for all situations, with more complex tasks, such as text selection and manipulation, being especially problematic [Instance 1994]. Other issues that can affect gaze-based systems include the Midas Touch problem [Jacob 1991] and equipment calibration in conjunction with user fatigue [Farid 2002a]. The Midas Touch problem reflects the fact that because a user’s gaze is rarely fixed on one space for a long period of time, parts of an interface could be “clicked” on even though this is not desired. For example, a change of gaze position could result in the wrong letter being selected on a graphical keyboard that is displayed on a screen.

### 2.2 Using Gestures as Part of HCI

Within HCI, gestures can provide an alternative interaction technique that allows users to perform tasks in a way that appears more natural, and closer to the way they carry out similar tasks in their

everyday life. For example, a personal digital assistant (PDA) user may use a stylus to write in a similar way to how he/she would write using a pen. The difference, however, is that gestures performed with the stylus may not necessarily represent the characters in the alphabet (e.g., the letter “A” could be represented by the gesture “Λ”). The effectiveness of the use of gestures in HCI depends on the use of an appropriate technique (e.g., a mouse is appropriate for interacting in a WIMP environment, but when coming to freehand drawing it is not as suitable as a light pen). The use of gestures within HCI has largely been focused on two main types: 2D (as commonly seen in WIMP interfaces, PDAs, computer-aided design (CAD), etc.) and 3D gestures. 3D gestural interfaces are still not widely used, but provide an advantage over their 2D counterparts in that they can enable the use of gestures that are performed naturally in the context of completing a specific task. This interaction can range from simple pointing [Bolt 1980], through to actual manipulation of on-screen objects [Bolt 1992; Sowa 1999] and sign-language recognition [Bauer 2002].

The main drawback in the adoption of gestures in HCI is the technology involved. Although the cost of available hardware for use in 2D gesture capture is relatively low, the cost of technology to capture the more sophisticated 3D gestures, such as data gloves, is much greater (ranging from \$500 to \$14000). Such technologies are also obtrusive, often requiring something to be worn by the user. The positioning of sensors and cables can make users feel restricted in their movements. However, these issues are being resolved, as recent technological developments have introduced less intrusive and more affordable devices facilitating the use of 3D gestures in everyday computer interaction.

Besides the hardware-related issues, attention must also be paid to the type of gestures used and how they are interpreted. In certain cases such as sign language, depending on the actual language being used (e.g., American versus British sign language), a specific sign might have a completely different meaning. Other issues include how “natural” the gestures that users have to perform have to be, and how much training is required in order to effectively use a gesture-based system.

### 3. RELATED WORK

Using gaze and gesture as a mechanism for HCI has also been investigated within other work.

The EagleEyes system [Gips 1996] allows the user to control the computer through eye or head movement. Electrodes placed on the user’s head measure the electrical potentials caused by eye movement, which in turn are translated into cursor movement on the screen. In order to simulate mouse clicks, dwell clicking is implemented (a user’s gaze dwells on a point for a certain period of time). The learning curve of the system is fairly small, and it is even being successfully used by people with severe disabilities. A key drawback, however, is the intrusive element of using electrodes. The Visual Mouse system [Farid 2002] also uses gaze to move a cursor, but seeks to be less intrusive by making use of eye scanners that are placed below the screen. Similar dwell clicking has been implemented, with the system being successfully used as a means to select and play video streams. Both EagleEyes and Visual Mouse are “look-and-dwell” systems and consequently, the interaction is limited to simple point-and-click mechanisms. This type of interaction does not support the rich range of activities that users come across in their everyday interactions with commercial packages. However, research has investigated ways to overcome both the Midas Touch problem, including solutions that efficiently utilize the “dead” space around objects, and the resolution of the gaze-tracking system. More recent systems have utilized viewers, gaze to adapt their experience. The iTourist system [Qvarfordt 2005] is an experimental tourist information system that adapts what it displays, depending on where the user is gazing on the screen. Magic Asian Art [Park 2006] is an interactive painting that alters its appearance based on where the viewer looks (e.g., if the viewer gazes at the sky, it will start snowing).

The Digital Baton [Marrin 1997] allows users to use a baton (embedded with motion detectors) to control computer music in a similar way to how a conductor leads an orchestra. The gestures that

are performed can be used to execute operations such as exactly-timed actions of individual notes, as well as and high-level functions like shaping volumes. WorldBeat [Borchers 1997] takes this further, allowing batons to be used not only for conducting and instrument playing, but also for interacting with a graphical user interface. The user can use the baton to navigate and select objects on a screen in order to access and use different applications. Although Digital Baton and WorldBeat achieve what they set out to do, they still rely on the user having to learn an artificial interaction approach (i.e., creating gestures with a baton). WorldBeat probably has the lowest learning overhead due to its being developed for a museum where children are the main users, with the interaction simplified accordingly.

A number of multimodal systems have also been developed. Bolt and Herranz [1992] developed a system that makes use of speech, gaze, and free-hand gestures as a means of interaction. Unlike systems such as Digital Baton, users do not make use of pointing devices to provide input and instead they simply perform gestures with their hands. These gestures are interpreted by the software and can be used in conjunction with other interaction modalities to manipulate objects on screen. For example, gaze can be used to select an object and gestures can then be used to rotate it. Bolt's system is a good example of how the different modalities can be used to form a two-phased interaction "dialog" with the computer; however, it is limited to only the manipulation of on-screen objects. Gandalf [Thorisson 1998] also makes use of speech, gaze, and gestures to allow a user to interact with a graphical humanoid agent. The Gandalf agent is capable of conducting a fluid unscripted dialog with the user, and can respond using speech, gaze, and gestures. The system is limited, however, in that the user is not able to perform any activities other than to solely interact with the agent.

Multimodal interfaces have also been investigated within virtual reality environments. Shared virtual-table environment (SVTE) systems have utilized gaze and gestures to help in activities such as group video-conferencing [Chen 2000; Aoki 1999; Gibbs 1999]. In this case, gaze and gesture interactions help to establish who is communicating with whom within a conference. These types of system have also been extended to more general, virtual shared working spaces [Kauff 2002].

These existing developments highlight a number of positive aspects, namely, that the combination of gaze and gesture can provide a natural yet potentially sophisticated way of interacting with a computer, and also that problems such as the Midas Touch can be overcome. What they also highlight, however, is that existing interfaces are not ideally suited for these interaction mechanisms. New interaction methods, that can more readily exploit the characteristics of gaze and gesture, need to be explored and developed.

#### 4. THE CONDUCTOR INTERACTION METHOD

This article presents an alternative interaction method, the conductor interaction method, which provides users with a more natural way of interacting with a computer. This method features both gaze and gesture interaction, and also uses tailor-made metaphors in order to help the user understand and exploit the interface. Primarily, we see this method as being most beneficial for users who have little experience with computers, enabling them to manipulate media objects easily. It could also be applied to a number of areas, including creative domains (e.g., where the user may create presentations, music, stories, lessons, etc.), and control domains, where the user may control an environment (e.g., the lights and sound in house or night club, or a surveillance system).

More specifically the CIM aims to provide the following.

—*A More Natural Interface.* The CIM will have an interface that utilizes gaze and gestures, but is nevertheless capable of supporting sophisticated activities. The CIM provides an interaction technique that is as natural as possible and is close to the human-human interaction methods with which users are already familiar. The combination of gaze and gestures allows the user to perform not only simple

interactions with a computer, but also more complex interactions such as the selecting, editing, and placing of media objects.

- A Metaphor Supported Interface.* In order to help the user understand and exploit the gaze and gesture interface, two metaphors have been developed. An orchestra metaphor is used to provide the environment in which the user interacts. A conductor metaphor is used for interacting within this environment. These two metaphors are discussed next.
- A Two-Phased Interaction Method.* The CIM uses an interaction process where each modality is specific and has a particular function. The interaction between user and interface can be seen as a dialog that is comprised of two phases. In the first phase, the user selects the on-screen object by gazing at it. In the second phase, with the gesture interface the user is able to manipulate the selected object. These distinct functions of gaze and gesture aim to increase system usability, as they are based on human-human interaction techniques, and also help to overcome issues such as the Midas Touch problem that is often experienced by look-and-dwell systems. As the dialog combines two modalities in sequence, the gaze interface can be disabled after the first phase. This minimizes the possibility of accidentally selecting objects through the gaze interface. The Midas Touch problem can also be further addressed by ensuring that there is ample dead space between media objects.
- Significantly Reduced Learning Overhead.* The CIM aims to reduce the overhead of learning to use the system by encouraging the use of gestures that users can easily associate with activities they perform in their everyday life. This transfer of experience can lead to a smaller learning overhead [Borchers 1997], allowing users to make the most of the system’s features in a shorter time.

#### 4.1 The Orchestra Metaphor

In order to provide users with an environment in which they can interact using gaze and gesture, an orchestra metaphor has been developed. The purpose of this metaphor is to graphically present to users the resources that are available for them to manipulate. The graphical representation of the resources and their positioning is done in a functional way to allow users to visually recognize and interact with them. For example, the image of a jukebox is associated with music, therefore inviting the user to interact with it to play some music. Because these resources can be visually recognized by the user, this metaphor can be used in gaze-based interaction.

The orchestral metaphor has its origins in the setup of a theater; the stage is the main area where the play is performed, and below this is the orchestra that consists of musicians playing their musical instruments under the guidance of a conductor, to supplement the play. The orchestra metaphor follows a similar idea to the theatrical setup. It uses the stage as the main area of interaction and presentation, and the orchestra represents the resources that are available to the user. There are no actors or musical instruments, but rather “media objects” which are visual representations of the media galleries (the groups of available resources) that they represent. These objects are placed on the two sides of the stage (see Figure 1), leaving the center stage available for the displaying and manipulation of these objects (e.g., combining them into a presentation). The functionality of the center stage is based on the actual metaphor, as the user should expect whatever happens to take place in that area of the screen. For example, if the user chooses an animation clip, it will be displayed in the central stage area.

Figure 1 provides an example of the orchestra metaphor with media-object galleries that could be used for the creation of multimedia presentations. For example, the photo gallery, depicted by a picture book, could represent a selection of images which the user can use in a presentation.

In order to interact with the orchestra metaphor it is proposed to use a *conductor metaphor*, as next defined.

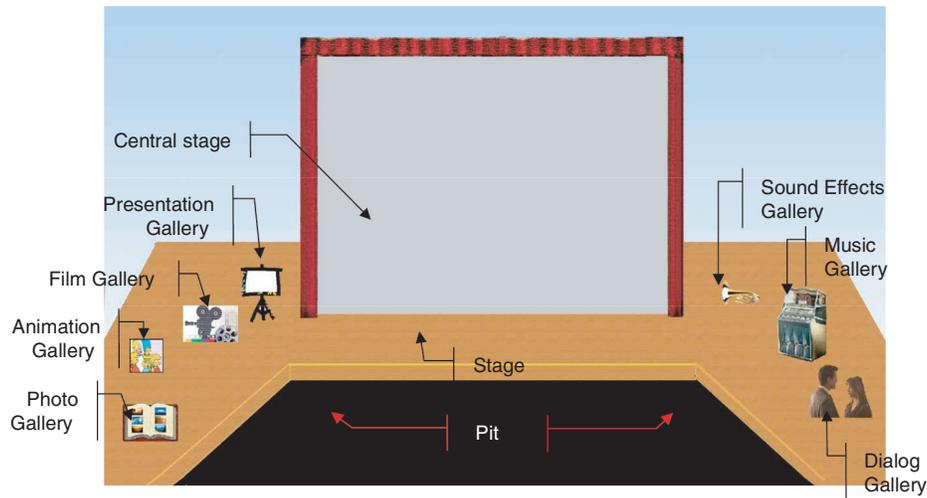


Fig. 1. The orchestra metaphor as used within the presentation conductor application.

## 4.2 The Conductor Metaphor

The conductor metaphor is used in conjunction with the orchestra metaphor, and is predominantly an interaction metaphor. Just as a conductor can interact with the musicians of an orchestra, the conductor metaphor allows a user to interact with the media objects and controls that are represented within the orchestra metaphor.

By breaking down the interaction of the music conductor, it is easy to identify the analogies between the way the music conductor and user (in the case of the orchestral metaphor) interact. The music conductor has the orchestra in front of him/her; the user has the media galleries and controls in front of him/her. The music conductor establishes eye contact with the musicians that he/she prompts to play their instrument; the user looks at the media gallery from which he/she intends to select an item. The music conductor uses bimanual (two-handed) gestures to guide the musician through the performance; the user uses bimanual gestures to interact with the media gallery objects. For example, when presented with a setup similar to that in Figure 1, the user may look at the “picture book” to select the photo gallery. Then using hand gestures, the user may select a picture of a forest fire and, with the appropriate gesture position, the picture on the central stage area where it will eventually be displayed.

The conductor metaphor itself provides a number of advantages. In particular, it is an interaction metaphor that has been designed specifically for gaze- and gesture-based interaction methods, and so is more suitable than other existing metaphors such as the desktop metaphor. The metaphor is also designed to be used in tandem with the orchestra metaphor. It is believed that together they can provide an interface that is easier to use and understand, especially for novice computer users.

## 5. THE PRESENTATION CONDUCTOR SYSTEM

In order to realize and evaluate the CIM, the presentation conductor system was developed to allow users to construct and display multimedia presentations using the interaction method.

Through gaze and gesture interfaces, the user (conductor) creates presentations by selecting and manipulating media objects. These media objects are grouped within media galleries and include images, videos, animations, and sounds. Depending on its type, each media object has a specific set of properties

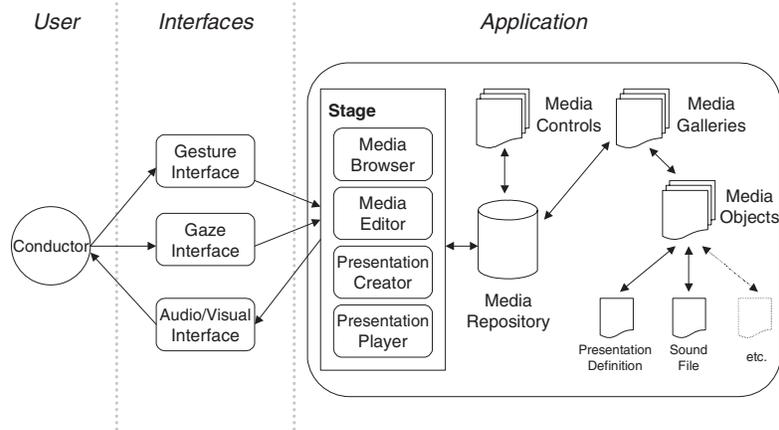


Fig. 2. Model of the presentation conductor.

that can be modified by the user using the appropriate media controls. For example, controls exist for altering the brightness of an image media object.

The “central stage” aspect of the interface represents the user’s work area and is where the selected media objects are manipulated and the presentation is displayed. Constructed presentations may be previewed at any point, or stored and displayed at a later date.

Figure 2 provides a model of the presentation conductor, which will be described in the following sections.

### 5.1 The User

The user represents the *Conductor* from the conductor metaphor. He/she interacts with the system to create a multimedia presentation through the gaze and gesture interfaces, while feedback is given on a display and through speakers. The conductor decides on the structure of the presentation and which media objects will be used to construct it. The conductor may also modify certain properties of these media objects, as well as store or discard a presentation.

### 5.2 The Interfaces

There are three types of interface to support user interaction with the application.

The *gaze interface* allows the conductor to interact with the application through gaze. Gaze itself is not used to issue complex commands, but rather to select elements from the visual display. The gaze interface recognizes the user’s focus of attention on a specific object, and then informs the application to activate it. When used in conjunction with the gesture interface, the gaze interface allows the conductor to select and manipulate an object.

The *gesture interface* allows the conductor to interact with the application through gestures. Within the presentation conductor, gestures are used for complex manipulations, particularly those of manipulating the media objects and media controls.

Gestural interaction can only occur after an object has been activated through the gaze interface. The object can then be manipulated through a series of gestures, which are carried out by the conductor’s hands. The gesture interface recognizes the gestures that are being performed, and communicates this to the application.

The *audio/visual interface* provides feedback to the conductor from the presentation conductor system. The main function of this interface is to display the visualization of the orchestra metaphor (e.g.,

on a computer monitor or projection screen) and to provide audio output. This includes the preview of the media objects and the visualization of the presentation. Feedback is not provided by the gaze and gesture interfaces.

### 5.3 The Application

The conductor can create, manipulate, and present multimedia presentations with the application. As shown in Figure 2, the application focuses on providing a “stage” that acts as the conductor’s work area, as well as managing a media repository that stores the media galleries, their media objects, their media controls, and previously created presentation definitions.

A *media object* represents a stored file that contains multimedia data. This could be, for example, an image (JPEG, GIF, and so on), sound (WAV, MP3, etc.) video (MPEG, AVI, etc.), or even stored presentations (represented by presentation definitions, discussed later). New media types can be readily incorporated into the system should they become available.

A *media gallery* represents a directory of media objects of a specific type (e.g., images) or with specific characteristics (e.g., distinctive sounds such as a siren or breaking glass). The purpose of the media gallery is to categorize media objects so that the conductor can easily navigate through them. Individual media galleries are visually represented on the stage.

The *media controls* represent mechanisms for controlling specific properties of the various media objects. For example, a picture may have its brightness and contrast modified, while a sound its volume and pitch. A mapping exists between media objects and media controls, so that the correct media control is applied to a media object. Such a mapping could be derived from, for example, the file type of the media object (e.g., a volume control being associated with sound media objects with a file extension of .wav). Media controls are individually visually represented, and appear when the relevant media object is selected.

The *presentation definitions* are objects that capture the structure of a presentation (e.g., the media objects used, their location, when they appear and for how long, etc). The application is able to save these definitions as a file, thus allowing the presentations to be stored and reedited at a later date (similar in manner to a PowerPoint presentation file). However, as well as being a mechanism to capture and store created presentations, it also allows created presentations to be incorporated into new presentations. For example, a new presentation could be created that incorporates one or more previously created presentations. A presentation definition is comprised of packets that represent media objects and their role within the presentation.

The *media browser* module enables the conductor to browse the media galleries that are available to the application. By using both gaze and gestures, the conductor is able to select and manipulate media galleries and preview the media objects they contain. Gaze is used to select the desired media gallery visualization and then gestures are used to choose the media object from that gallery. Once a media object has been selected, it is passed on to the media editor where it can be tailored for the presentation.

The *media editor* module enables the conductor to apply the relevant media controls (derived from the mapping) to the selected media object. Depending on the number of media controls, a combination of gaze and gestures would be used to first select the desired media control and then to edit the media object’s properties.

The *presentation creator* module deals with the actual generation of the multimedia presentation, as well as handling its storage. It is this module where the actual presentation construction takes place and it is therefore the main module within the application. From here the conductor can browse the media galleries, edit media objects, store/view presentations, etc.

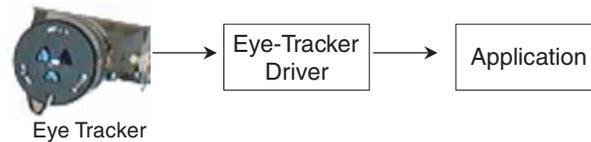


Fig. 3. Capturing and recognising user gaze within the conductor metaphor.

The presentation creator module itself has three main functions:

- Positioning of Media Objects on the Central Stage Area.* When the user has completed the editing of the media object's properties, the presentation creator positions the media object on the central stage (where other, previously positioned objects may exist). With gestures, the user may position the media object anywhere on the central stage for the purposes, of the presentation.
- Handling of Media Object Timing.* The presentation creator also handles the sequence and lifetimes of the media objects that exist in a presentation. A lifetime spans from the moment that a media object appears in the presentation until:
  1. the presentation ends;
  2. the media object is removed as part of the presentation sequence; or
  3. the media object has a time property that has expired (e.g., with sound and video clips).

The presentation creator keeps track of a presentation's sequence, which represents the points in the presentation when different media objects come into play.

- Storing of Presentations in the Media Repository.* The presentation creator is able to store created presentations in the media repository. This involves the creation of a presentation definition object, which is then stored.

The *presentation player* module is used to preview or play a presentation that is currently being edited. The conductor is able to move backwards and forwards through a presentation as well as stop it at any point, by using gestures.

The *media repository* module stores the data that is used by the system. It stores the media galleries, the media objects that make up the galleries, and the media controls.

## 6. IMPLEMENTING THE PRESENTATION CONDUCTOR

The presentation conductor was implemented using Visual Basic .NET, allowing a fairly complex interface to be developed with minimal effort. Moreover, VB .NET allows the GUI to be controlled at all points of an interaction, which is important, given that the user interacts using two different modalities. The media repository used by the application (to store the media objects, media galleries, etc.) is represented by the local file system.

### 6.1 Implementing the Conductor Metaphor

The capturing of gaze within the implementation of the conductor metaphor is performed using the process illustrated in Figure 3.

In order to capture the user's gaze, an LC Technologies VOG [LC Technologies 1986] eye tracker was used. This type of eye tracker works by measuring the relative positions of the glint and bright eye produced when infrared light is shone into the eye, resulting in a less intrusive interface for the user. A drawback of this is that it needs to be calibrated to a user's position, and should this change then recalibration is also required. To interface the eye tracker with the native on-screen mouse pointer, a

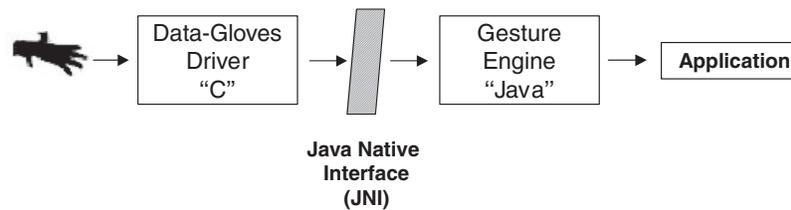


Fig. 4. Capturing and recognizing gestures within the conductor metaphor.

Table I. A Selection of the Gestures Used Within the Presentation Conductor

Gesture	Name	Function
	<b>Left-hand twist</b> All fingers are flexed Hand performs a twisting act from right-to-left and vice versa	Used to adjust properties of a selected media object, by turning a knob.
	<b>Right-hand Point</b> Index and thumb stretched, all other fingers are flexed No movement	Used to select an item from a list.
	<b>Right-hand Scroll Up and Down</b> All fingers are stretched Hand performs a vertical flex movement	Used to scroll through the contents of the active media-gallery.
	<b>Right-hand Flex and hold Up and Down</b> All fingers are stretched. Hand performs a vertical flex movement, and pauses at the extreme high or low	Used to adjust vertical position or height of a displayable media object
	<b>Left-hand vertical flat</b> All fingers are stretched Hand performs a horizontal movement	Used to adjust horizontal position to the right of a displayable media object
	<b>Both hands vertical flat, palms facing</b> All fingers are stretched Both hands stretched, vertical to the ground, palms facing each other	Used to proportionally decrease the size of a displayable object
	<b>Both hands flat</b> All fingers are stretched Both hands flat, parallel to the ground	Used to finalize an operation.
	<b>Both hands "shut the box"</b> All fingers are stretched Both hands stretched, vertical to the ground, palms facing each other, do an inwards movement as if shutting a box	Used to exit from the application

driver is used that maps gaze movement to mouse movement. This implementation is generic enough to allow it to be used with other applications besides the presentation conductor.

The capturing of gestures within the implementation of the conductor metaphor was performed using the process illustrated in Figure 4.

In order to capture the user's gestures, 5DT optical fiber-based gloves were used [5DT 2004]. These gloves are ideal for tasks that do not require very high precision of identification of hand movement. A special Java-based gesture engine was developed that can recognize performed gestures based on the data being output by the data-gloves driver. For this implementation the gesture engine was pre-programmed with a range of gestures (a selection of which are shown in Table I); however, new gestures can be easily added.

Once a gesture has been recognized it is then mapped to a keystroke. Mapping gestures to keystrokes provides a number of advantages, in particular the fact that keystrokes are an easy input for programming languages to handle, and also are easy to test during application development. As with the eye-tracking software, the gesture engine provides a generic interface, which means that means it can be easily incorporated into applications other than the presentation conductor.

## 6.2 Implementing the Orchestra Metaphor

The implementation of the orchestra metaphor for the presentation conductor features the stage and central stage, as discussed previously (see Figure 1). The stage is that part of the graphical interface where the objects that the user interacts with are placed, while the central stage is the part in which the results of the interaction are displayed, namely the edited elements of the presentation.

The range of “instruments,” that is, media objects, that are presented on the stage currently consist of the following.

- A *Music Gallery (Jukebox)*. This represents various sound objects and includes music and environmental sounds.
- A *Sound-Effects Gallery (Trumpet)*. This represents a series of sound-effect clips.
- A *Dialog Gallery (Two People Talking)*. This represents a series of short phrases.
- A *Film Gallery (Film Camera)*. This represents a series of films (i.e., digitized video sequences).
- An *Animation Gallery (Cartoon Characters)*. This represents a set of animations.
- A *Photo Gallery (Picture Book)*. This represents a series of pictures.
- A *Presentation Gallery (Projection Screen)*. This represents previously created and stored presentations.

## 6.3 Providing Two-Phased Interaction

As previously discussed, it was intended for the CIM to use a two-phased interaction process. To achieve this within the implementation of the conductor method, the interaction process uses each modality (gaze and gesture) in a specific way and in a particular function. The interaction between user and interface can be seen as a *dialog* that is comprised of two phases: selection and manipulation. In the first phase, the user selects the on-screen object by *gazing* at it (e.g., the user looks at the *picture book* object and activates it, as illustrated in Figure 5(a)). If the user looks elsewhere, the object is automatically deselected. In the second phase, with the *gesture* interface the user is able to manipulate the selected object (e.g., by flexing vertically his /her right hand the user can scroll through the list of images, as illustrated in Figure (5b)). This combination of interaction modalities provides two-phased interaction and can help overcome the Midas Touch problem.

## 6.4 Assisting in Reducing the Learning Overhead

It was also desired for the CIM to reduce the learning overhead required to utilize the interaction style and any respective applications. To achieve this, a small number of gestures were used that correspond to those that would normally be used either in performing a similar task or representing it when describing it in a pantomimic way; for example, adjusting the volume of an amplifier by turning the volume control knob, or how a policeman stops traffic by using a flat-hand gesture. By using gestures that are mostly pantomimes, it is intended that both cross-cultural misinterpretations and the learning overhead will be reduced. Table I presents a subset of the gesture vocabulary that was used within the presentation conductor.

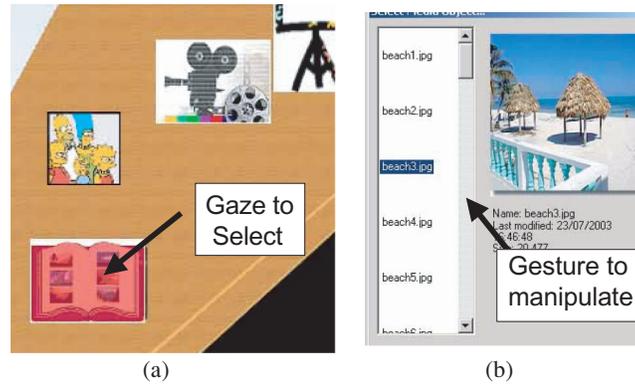


Fig. 5. How two-phased interaction has been implemented.

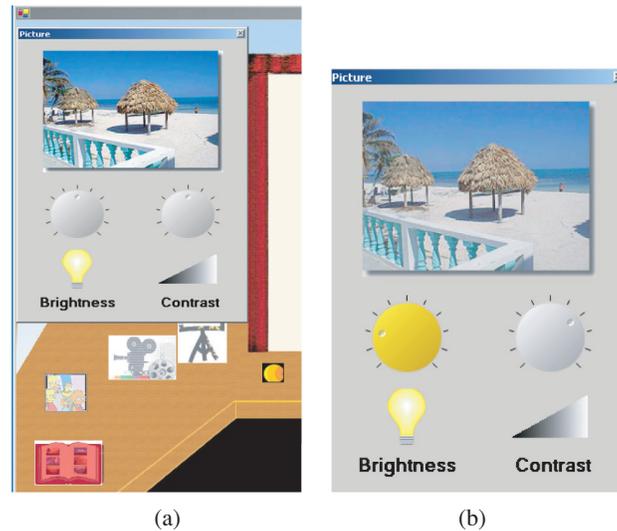


Fig. 6. (a) Media controls available via the media editor; (b) adjusting the brightness of the image.

## 6.5 The Conductor Interaction Method in Use

To demonstrate the CIM being used within the presentation conductor, we will now provide a brief example depicting how a hypothetical user (Jack) uses it to create a presentation.

Jack starts the presentation conductor application and is presented with the stage as it was shown in Figure 1. Jack decides he wants to add a picture of a beach to his presentation. He gazes at the picture-book icon on the stage. The application interprets this and highlights the icon in red (as previously shown in Figure 5(a)). Jack continues to stare at the icon and in doing so activates the media browser. In the browser, Jack scrolls through the list of pictures by flexing his palm up and down. A preview of each picture is shown and Jack manages to find a suitable beach picture (Figure 6). He points at it with the index of his right hand and the picture is loaded into the media editor.

There are two media controls available to Jack: brightness and contrast (see Figure 6(a)). Jack decides he would like to darken the picture slightly. He gazes at the brightness icon. The application recognizes



Fig. 7. Placing the image on the central stage.

this and highlights the brightness control knob in yellow. Jack then twists his left hand and in doing so adjusts the brightness of the picture (Figure 6(b)). When Jack is satisfied he holds both his hands out flat and exits the media editor.

The beach image that Jack edited is now displayed on the central stage. The image is automatically centered on the display area and covers it completely. Jack may now adjust the size of the image and its position on the central stage. Jack decides he would like to shrink the picture slightly. He positions his hands horizontally, palms facing each other in front of him to initiate the adjustment of the image's height. He then flexes his left hand which is facing up, mimicking a "push" to the image. The image height decreases. Jack is satisfied with the height of the image and holds both his hands out flat to end the height adjustment action. Similar steps are performed to adjust the width of the image (using the relevant gesture for adjusting the width). When Jack is satisfied with the adjustment of the image he holds both his hands out flat to set the adjustments.

Although this is only a simple example, it demonstrates how the CIM is used within the presentation conductor. Having users make use of the presentation conductor in order to create presentations formed the basis of evaluation of the method.

## 7. EVALUATION

To assess the usefulness of the CIM as well as to obtain feedback on its application, a number of user studies were performed. These studies aimed to assess the implementation of the method and its learning overhead, in comparison with existing interaction approaches. Moreover, the studies investigate the appropriateness of the presentation conductor and the impact that the underlying technology has on the method's effectiveness. As one of the aims of the method was to assist inexperienced users in their interaction with the computer, groups of users possessing differing levels of computer experience were used in the studies. Due to the prototypical nature of the presentation conductor system, formative and qualitative evaluation approaches were primarily used in assessment.

The evaluation itself focused on users employing the presentation conductor to carry out a set of prescribed tasks. Three groups of users participated in the evaluation: a group of four experienced computer users, a group of three inexperienced computer users, and a control group that was also made up of three inexperienced computer users. The experienced user group was made up of colleagues



Fig. 8. Stage 1 setup.

from the Computing department who possessed extensive Microsoft PowerPoint experience (over 5 years of regular use). The inexperienced users included four first-year and two postgraduate students from noncomputing disciplines (music and psychology), and a member of the university's security staff. Although some of these users had used computers for tasks specific to their discipline, they had no or very little experience in using PowerPoint. The control group was used to identify whether the creation of a presentation with PowerPoint had any effect on how inexperienced users then interacted with the presentation conductor. Drawing upon users with different levels of PowerPoint experience was important, as it would help us to determine the impact that prior experience with other interaction techniques (for creating presentations) would have on a user's perception of the CIM.

The evaluation was broken down into the two stages next described.

*Stage 1.* The first stage focused on obtaining feedback on the CIM, and particularly, how it compared with existing interaction techniques. The groups were given a specific presentation to create by first using PowerPoint and then the presentation conductor system. The control group did not take part in the PowerPoint part of this experiment.

*Stage 2.* The second stage focused on obtaining feedback on the use of technology with the CIM. All three groups were again given the same presentation to create, but this time they had to use the presentation conductor system with a different set of devices.

All evaluation sessions were filmed from two angles, namely focusing on the display screen and on the user. Verbal feedback was also obtained as a result of investigator questions and from the users "thinking aloud" whilst performing the tasks. At the end of every stage the users were asked to fill in standardized questionnaires and were also interviewed. The questionnaires also included assessment activities related to the NASA Task Load Index [Hart 1987].

### 7.1 Stage 1: Evaluating the Conductor Interaction Method

For this stage the user sat in front of a workstation that was fitted with an LC Technologies eye tracker and a pair of 5th DT data gloves. Figure 8 illustrates the setup for this stage.

Each of the sessions for this stage consisted of up to six phases (five for the control group), as next presented.

- Preliminary Briefing*. This consists of a brief overview of the evaluation procedure, and an introduction to computer-based presentations.
- PowerPoint Tasks*. (These were not performed by the control group.) The inexperienced or experienced user was given a presentation to create using Microsoft PowerPoint.
- Briefing on the CIM and Presentation Conductor*. The user was provided with a brief description of the CIM and presentation conductor system. Following this, the user was presented with the application user guide and given 5–10 minutes to read it.
- System Calibration and Training*. The system was calibrated to the individual user and the gesture vocabulary was practiced.
- Presentation Conductor Demonstration*. The presentation conductor environment was activated and the investigator guided the user through the environment and instructions on how to perform various operations within the application.
- Presentation Conductor Tasks*. The user was then asked to create the same presentation as he/she had done with PowerPoint (the control group performing these tasks for the first time).

A semistructured evaluation approach was used where the users performed the tasks and were encouraged to give verbal feedback.

7.1.1 *Discussion of Sessions*. Overall feedback from users with regards to the CIM was mixed. All found it to be a novel approach, but a few were hesitant to embrace it as an interaction technique to replace what they currently use today. In particular, many of the experienced users became frustrated with hardware calibration issues and the learning overhead that was involved. This was less of an issue with the inexperienced users, and they compared the interaction approach favorably to existing techniques they had used.

*“I didn’t have a good time! Didn’t get what I wanted, and it took a long time.”* (Experienced user)

*“It was much more fun, much more exciting. There was a sense of doing something novel and seeing how such interaction techniques are capable of working and using them to achieve something; which leads to a belief that they will become more widespread.”* (Experienced user)

*“I liked it! Once you get used to it, you can do it really fast!”* (Inexperienced user)

*“Adjusting Controls was easier with the gloves, it was more natural for me.”* (Inexperienced user)

Users believed that the interaction method would be best suited for application domains that relied heavily on manipulation, and suggested alternative applications including browsing, music mixing, gaming, and designing. One user also pointed out that the CIM would be a more suitable means of human-computer interaction for people with disabilities (particularly for those who already heavily use gestures, e.g., deaf people). In this context it could be used to help control systems such as home lighting, television, or audio; systems that may traditionally be difficult for people with disabilities to use.

*“The use of hands could help people with disabilities.”*(Inexperienced user)

*“... in a kitchen environment e.g. cook-book browsing.”*(Inexperienced user)

*“... some kind of sorting through objects, or browsing / searching; perhaps a specialist desktop application of some kind.”* (Experienced user)

Most users understood the use of the Orchestra and conductor metaphors, and found their use appropriate for the nature of the tasks they had to complete. Users found that the Orchestra metaphor assisted them in using the interaction methodology and liked the idea of the central interaction space.

Some experienced users, however, commented that the metaphor was just a different type of “desktop.” In terms of the actual implementation, it was mentioned as to that the icons could be larger in size so as to be easier targets at which gaze. The small size of the icons made it harder for some users to gaze-select the media galleries on the stage, or to focus on them for the required length of time.

*“I can’t seem to be able to hold my focus on [the media gallery] for long. Maybe it would have been easier if it was bigger . . . .”* (Experienced User)

It was observed that most users (especially the inexperienced ones) became very involved and *assumed the role* of the conductor. The majority of the gestures were found to be natural in the context in which they were used, although some were quite difficult (and painful) to perform, highlighting differences in people’s wrist flexibility. It was also noted that when some users performed a gesture with one hand, their other hand would also perform some sort of “recognizable” gesture without their realizing it. Consequently, the system would recognize a different set of gestures than the user actually intended. This highlights that either some of the gestures need to be altered, or perhaps mechanisms need to be added to the implementation to prevent gesture misinterpretation.

*“Now I really feel like a Maestro!”* (Inexperienced user)

Users indicated that they would have liked a visual representation of their hands during the gesture training activity, to help them understand what they were doing and when they were getting the gestures wrong.

*“Perhaps you could do with some visual representation of the gestures at this stage [gesture training].”* (Experienced User)

*“More visual feedback [of gestures] would have been helpful.”* (Experienced User)

It was observed that the context of the interaction would often help the users to remember what gesture to use (e.g., turning the brightness knob using a left-handed twist gesture). This further illustrates how choosing a gesture vocabulary that reflects the desired actions can significantly affect the reduction the learning overhead.

*“Oh yes, so now I push it like this [performs gesture] to move it to the left.”* (Inexperienced user)

*“Oh yes, that gesture makes sense . . . .”* (Experienced user)

Overall, users found the learning overhead for the CIM to be greater when compared to other existing interaction techniques. This was particularly the case for experienced users (possibly due to their bias towards those interaction techniques with which they were already familiar). On the other hand, the majority of inexperienced users believed that the learning required was no different from any of the other interaction techniques they had used. Although this implies that the CIM might not be a particularly easy interaction interface to use, it does suggest that it is on a par with existing ones (certainly for inexperienced users). This was found to be promising, especially when one considers that the implementation of the method is but a prototype and that it, and the method, could be further refined.

**7.1.2 NASA Task Load Ratings.** In addition to the feedback provided as a result of the performed tasks, users were also asked to complete NASA Task Load (TLX) [Hart 1987] questionnaires after every experiment. The TLX can be used to measure the workload that users perceive themselves to be under, and was used to help quantifiably analyze the qualitative feedback. Figures 9 and 10 present a summary of the results for this stage and depict the average group ratings for each attribute (the lower the rating, the better). For example, the experienced user group found using the presentation conductor with the eye tracker to be particularly physically demanding.

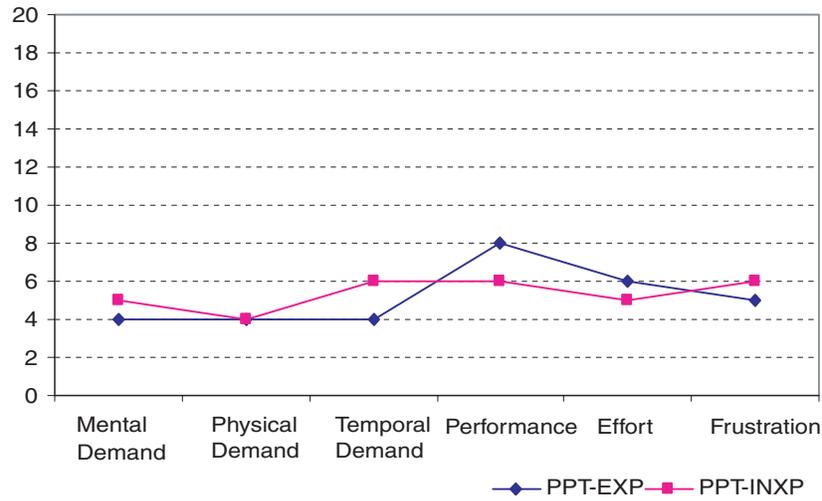


Fig. 9. Task load when using PowerPoint.

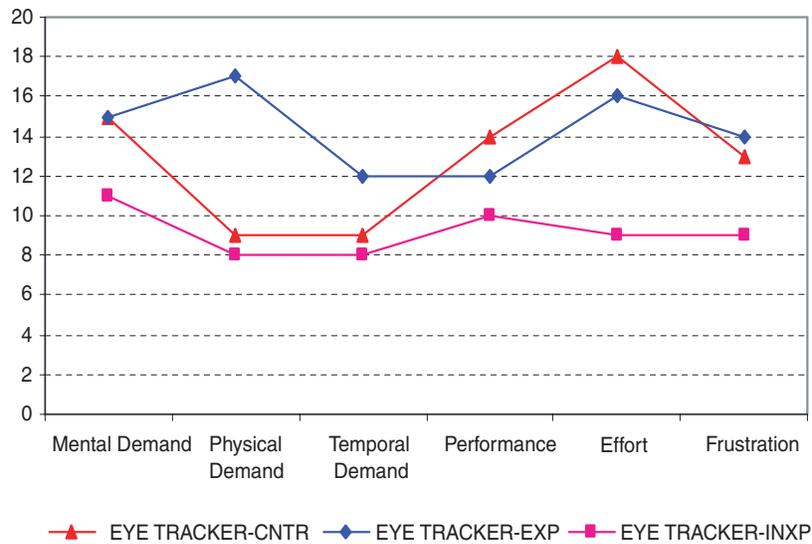


Fig. 10. Task load when using the CIM with the eye tracker.

The results of the TLX ratings correlate with the provided verbal feedback. They show how both inexperienced and experienced users found the CIM to be difficult to use with the eye-tracker setup. When compared to the rating provided for using PowerPoint, these differences are significant.

The TLX ratings also illustrate notable differences between experienced and inexperienced users. To an extent this is to be expected, as more experienced users will be used to working in a WIMP-based environment. However, it does highlight how the CIM is likely to be most beneficial for people with limited computer experience, or who are unable to interact using traditional techniques (e.g., disabled people).

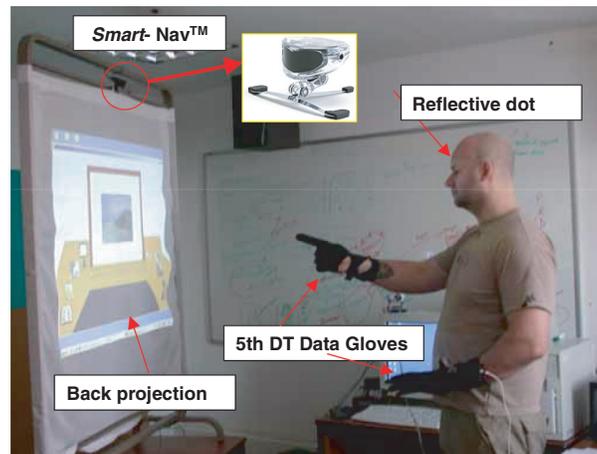


Fig. 11. Stage 2 setup.

## 7.2 Stage 2: Evaluating the Impact of Technology

For this stage of the evaluation it was decided to assess whether head pointing (instead of gaze) in combination with a large screen would have any effect on the usability of the CIM. In particular, it was desired to see whether interacting whilst standing up, where a greater freedom of movement exists, would have any bearing on the CIM. To achieve this, the user stood in front of a projection screen that provided a back projection of the workstation display. A Smart-Nav™ [NaturalPoint 1997] optical tracking tower was fixed on top of the screen frame and a reflective patch was placed on each user's forehead. The tracking tower emits infrared light which is reflected back, processed, and translated into cursor movement. The pair of 5th DT data gloves were again used for this stage. The setup for this stage of evaluation is illustrated in Figure 11.

The second stage was broken down into two phases in which all groups took part, detailed as follows.

- Technology Briefing, Calibration, and Training.* An overview of the different technologies used in this stage was provided. The Smart-Nav™ technology was calibrated to the user, who then practiced using the head tracker. Due to the different posture used (now standing), it was also necessary to recalibrate the hand-gesture software.
- Presentation Conductor Tasks.* The user was then asked to complete the same set of tasks that they performed in stage 1, but this time using the new setup and technology.

**7.2.1 Discussion of Sessions.** In both stages of the evaluation, none of the users had used the interaction technologies before, however, overall they were enthusiastic about them. Nonetheless, technical issues did arise. The eye tracker was problematic due to the amount of calibration involved and the ease with which this could, in turn, be lost. Users who wore glasses found it particularly hard to calibrate the eye tracker, and in two cases calibration could not be achieved at all. In addition, differing size of user hands affected the ability of the data gloves to recognize gestures, highlighting the fact that a “universal” data glove size is not really a feasible possibility.

*“The eye tracker didn’t work for me. The gloves failed to find certain positions . . .”* (Experienced user)

The loss of eye-tracker calibration happened quite frequently within the evaluation and was commonly caused by users fidgeting and readjusting their seating position, a user’s body moving out of

position whilst performing gestures, and users naturally turning their head to focus on an on-screen object rather than just moving their eyes. Some of these problems could be tackled by using alternative eye-tracking technologies that support head movement, however, these are typically even more intrusive to the user.

The optical tracking tower was better received and clearly the preferred technological setup. Removing the tracking technology from the user's person not only reduced the likelihood of technical difficulties, but also allowed the user much greater freedom of movement.

*"In the beginning I preferred the eye tracker, but when I saw I could move easier I liked this one better."* (Experienced User)

The greater freedom also helped many users in performing the hand gestures and reduced the accidental occurrence of other gestures. The fact that users were now able to move their eyes independently of their head also allowed them to make use of their peripheral vision. On the negative side this meant that if a user looked at an object without moving his/her head, this would not be detected by the optical tracking tower. However, due to the large screen used for the experiment this was not noticed as being a significant problem.

A key benefit of the optical tracking tower was the fact that it did not require user calibration. Consequently, it typically only took a few minutes for users to familiarize themselves with the technology, a direct contrast to the longer period required for the eye tracker. The use of the large screen was also found to be more favorable. The sizes of the displayed media gallery objects were larger and consequently users found them easier to select.

*"This [big screen setup] is far better."* (Experienced User)

Overall, the majority of users found the interaction and use of the CIM to be a smoother process with the optical tracking tower setup, rather than with the eye tracker. This setup also benefited from producing far fewer technical issues.

**7.2.2 NASA Task Load Ratings.** In stage 2 the users were also asked to fill in TLX questionnaires. The results of these showed that, in contrast to the eye-tracker setup, the TLX ratings for the optical tower setup (*large screen*) were much lower, indicating that the choice of technology can have a significant impact on the effectiveness of the CIM. Furthermore, the inexperienced users provided ratings that were comparable to those they had provided when using PowerPoint. This is promising, considering that the technology and implementation of the method can still be further refined. Figure 12 summarizes the TLX results of this stage.

## 8. CONCLUSIONS AND FUTURE WORK

This article has presented a novel interaction method, the conductor interaction method. It has been argued that existing interaction techniques have three notable failings: They are typically unnatural and require the user to learn a new interaction method, they are often inconsistent in their semantic language, and they can restrict creativity and be oversimplistic. The CIM seeks to provide an alternative interaction technique that tackles some of these failings. To achieve this it draws upon human-human interaction techniques and makes use of a two-phased multimodal interaction mechanism, namely gaze for selection and gestures for manipulation, as well as incorporating this in a metaphor-based environment.

A description of the CIM has been provided, including detailing of the conductor and orchestra metaphors that are central to its operation and form the basis of the gaze/gesture interaction. To help realize the interaction method, a prototype system has been developed that allows users to create

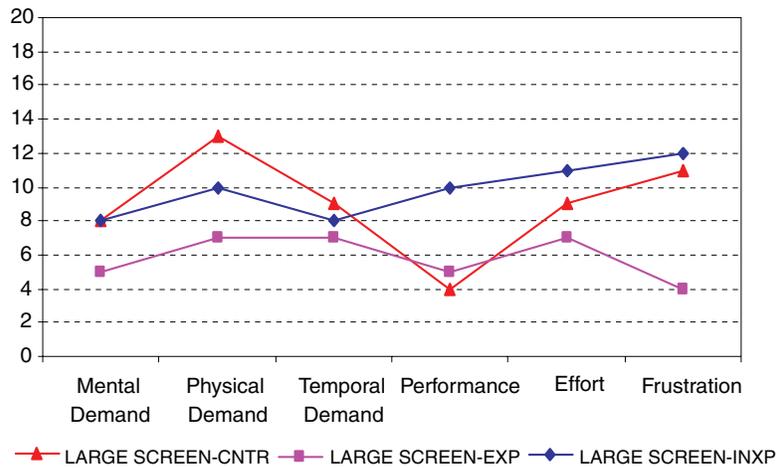


Fig. 12. Task load when using the CIM with the large screen/head tracker.

multimedia presentations. This prototype was used as the basis for the evaluation that has been carried out on the methodology.

The evaluation has shown that the CIM, and a gaze/gesture-based approach, can provide a viable alternative interaction technique. However, it is probably most suited to users with little or no prior computer experience, and to application domains that, in their noncomputing form, typically involve hand manipulation to a great extent. Although the feedback was not overwhelmingly positive, it is promising that an alternative interaction method, while at the prototype stage, was generally well received. The evaluation also highlighted the difference that interaction technologies can have on the CIM; in particular, large screen/head-tracking-based approaches proved to be more effective than the more static eye-tracker alternative. The key issue here was the intrusiveness of the technology not only in terms of equipment, but also in terms of aspects such as calibration and the loss thereof.

The evaluation did highlight areas of the methodology and implementation that could be further refined, namely the following.

*Improved method training support.* Users suggested that they would have preferred to have a visual representation of their hand movement, and feedback on the correctness (or otherwise) of a gesture. Possible solutions include the creation of a virtual model of the user's hands to check whether gestures are being performed correctly, an on-screen visual representation of the user's hands, or a more comprehensive gesture training session.

*Refining the metaphors.* Evaluation feedback indicated that the orchestra metaphor would benefit from using larger icons for the galleries, not only to make them more distinguishable but also to help in their selection using gaze-based technologies. Feedback on the conductor metaphor indicated that users found some gestures hard to relate to. One solution would be to allow user-defined gestures, allowing for a more tailored interaction experience.

*Experimenting with different technologies.* Different technologies can be investigated to see whether some of the hardware issues can be resolved, as well as to further assess the impact of technology on the applicability of the CIM. Computer vision and the use of position trackers on hands are two areas that can be explored for gesture recognition. This would allow for hand position to be determined, which

could prove beneficial to gesture refinement. Likewise, the latest gaze/head-tracking technology can be assessed.

*Investigating different application domains.* Feedback suggested other possible application domains that the method could be used with, including browsing, music mixing, gaming, and designing. Such domains heavily rely on manipulation (often hand-based), where the use of a two-phased interaction approach could be appropriate. Another area of investigation is the open question of whether the CIM can be beneficial to people with disabilities. A large proportion of people with disabilities, particularly the deaf, use gestures to interact in their everyday life. As well as relying on signing and visual contact, deaf people also use eye contact to establish the communication channel. This combination of gaze and gestures mirrors the two-phase interaction used within the CIM, suggesting that the technique could be an HCI method that they could easily adapt to and utilize.

In addition to these issues, the evaluation itself was fairly limited in its scope (e.g., in the number of participants) and we intend to perform more extensive, larger-scale evaluation studies. One possibility may include timing how long it takes for participants to complete a set of given tasks. Carrying out these studies as well investigating the aforementioned points will form part of the methodology's continuous refinement.

## REFERENCES

- 5DT. 2004. Data glove. <http://www.5dt.com/>.
- ARGYLE, M. 1996. *Bodily Communication*, 2nd ed., Routledge, London.
- AOKI, T. 1999. MONJUnoCHIE system: Videoconference system with eye contact for decision making. In *Proceedings of the International Workshop on Advanced Image Technology (IWAIT)*.
- BAUER, B. AND KRAISS, K. F. 2002. Towards an automatic sign language recognition system using subunits. In *Proceedings of the International Gesture Workshop, on Gesturure and Sign Language in Human-Computer Interaction*. Lecture Notes in Computer Science, vol. 2298, Springer, Heidelberg, Germany.
- BAUML, B. J. AND BAUML, F. H. 1997. *Dictionary of Worldwide Gestures*. Scarecrow Press. MD.
- BENFORD, S., SNOWDON, D., GREENHALGH, C., INGRAM, R., KNOX, I., AND BROWN, C. 1995. VR-VIBE: A virtual environment for co-operative information retrieval. In *Proceedings of the Conference, Eurographics*. Maastricht, The Netherlands.
- BOLT, R. 1980. "Put-that-there": Voice and gesture at the graphics interface. In *Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques*, Seattle, WA.
- BOLT, R. AND HARRANZ, E. 1992. Two-Handed gesture in multi-modal natural dialog. In *Proceedings of the 5th Annual ACM Symposium on User Interface Software and Technology (UIST)*.
- BORCHERS, O. 1997. WorldBeat: Designing a baton-based interface for an interactive music exhibit. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Atlanta, GA.
- CHEN, W. C. 2000. Toward a compelling sensation of telepresence: Demonstrating a portal to a distant (static) office. In *Proceedings of the IEEE Visualization Conference*, Salt Lake City, UT.
- DIX, A. J., FINLAY, J. E., AOWD, G. D., AND BEALE, R. 2004. *Human-Computer Interaction*, 3rd ed., Prentice Hall, Hertfordshire, UK.
- FARID, M. M. AND MURTAGH, F. 2002. Eye-Movements and voice as interface modalities to computer systems. In *Proceedings of the SPIE Regional Meeting on Optoelectronics, Photonics and Imaging (OPTO)*, Galway, Ireland.
- FARID, M. M., MURTAGH, F., AND STARCK J. L. 2002. Computer display control and interaction using eye-gaze, *J. Soc. Inf. Display*.
- GIBBS, S. J. 1999. TELEPORT-Towards immersive copresence. *Multimedia Syst.* 7, 3, 214–221.
- GIPS, J. AND OLIVIERI, P. 1996. EagleEyes: An eye control system for persons with disabilities. In *Proceedings of the 11th International Conference on Technology and Persons with Disabilities*, Los Angeles, CA.
- HART, S. G. 1987. Background description and application of the NASA task load index (TLX). In *Proceedings of the Department of Defence Human Engineering Technical Advisory Group Workshop on Workload (NUSC6688)* Newport, RI, 90–97.
- INSTANCE, H. AND HOWARTH, P. 1994. Keeping an eye on your interface: The potential for eye-gaze control of graphical user interfaces. In *Proceedings of the Conference on Human-Computer Interaction, People and Computers*, Glasgow, 195–209.
- JACOB, R. J. K. 1991. The use of eye movements in human-computer interaction techniques: What you look at is what you get. *ACM Trans. Inf. Syst.* 9, 3, (Apr.).

- KAM, K. S. 2002. Definitions “dwell” vs. “gaze”. In *Eye-Movement Mailing List*, <http://www.jiscmail.ac.uk/files/eye-movement/introduction.html>.
- KENDON, A. 1992. The negotiation of context in face-to-face Interaction. In *Rethinking context: Language as an Interactive Phenomenon*, Duranti and Goodwin, Eds. Cambridge University Press.
- KAUFF, P. AND SCHREER, O. 2002. An immersive 3D video-conferencing system using shared virtual team user environments. In *Proceedings of the 4th International Conference on Collaborative Virtual Environments (CVE)*, Bonn, Germany, 105–112.
- LC-TECHNOLOGIES 1986. VOG. <http://www.eyegaze.com/>.
- MARRIN, T. AND PARADISO, J. 1997. The digital baton: A versatile performance instrument. In *Proceedings of the International Computer Music Conference*, Thessaloniki, Greece.
- NATURAL POINT. 1997. Optical tracking systems. <http://www.naturalpoint.com/>.
- PARK, E., KIM, B., SALIM, W., AND CHEOK, A. D. 2006. Magic Asian art. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Montréal, Québec, Canada, 255–258.
- QVARFORDT, P. AND ZHAI, S. 2005. Conversing with the user based on eye-gaze patterns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Portland, OR, 221–230.
- SALVUCCI, D. D. 1999. Inferring intent in eye-based interfaces: Tracing eye movements with process models. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Pittsburgh, PA.
- SHNEIDERMAN, B. AND PLAISANT, C. 2005. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley.
- SIBERT, L. E. AND JACOB, R. J. K. 2000. Evaluation of gaze interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, The Hague, The Netherlands.
- SOWA, T., FROHLICH, M., AND LATOSCHIK, M. E. 1999. Temporal symbolic integration applied to a multimodal system using gestures and speech. In *Proceedings of the International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction*, Gif-Sur-Yvette, France. Lecture Notes in Computer Science, vol. 1739. Springer, Heidelberg, Germany.
- THORISSON, K. R. 1998. Real-Time decision-making in multimodal face-to-face communication. In *Proceedings of the Autonomous Agents Conference*, Minneapolis, MN.
- VERTEGAAL, R. 1999. The gaze groupware system: Mediating joint attention in multiparty communication and collaboration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Pittsburgh, PA.
- VERTEGAAL, R., SLAGTER, R., VAN DER VEER, G., AND NIJHOLT, A. 2001. Eye gaze patterns in conversations: There is more the conversational agents than meets the eyes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Seattle, WA.
- ZHAI, S., MORIMOTO, C., AND IHDE, S. 1999. Manual and gaze input cascade (MAGIC) pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Pittsburgh, PA.

Received August 2007; accepted August 2007