Towards Multi-modal Interaction with Interactive Paint

Nicholas Torres, Francisco R. Ortega(✉), Jonathan Bernal, Armando Barreto, and Naphtali D. Rishe

Florida International University, Miami, FL 33196, USA
{ntorr054,fortega,jber006,barretoa,rishen}@fiu.edu,
fortega@cs.fiu.edu

Abstract. We present a Multi-Modal Interactive Paint application. Our work is intended to illustrate shortcomings in current multi-modal interaction and to present design strategies to address and alleviate these issues. In particular, from an input perspective use in a regular desktop environment. A serious of challenges are listed and addressed individually with their corresponding strategies in our discussion of design practices for multi-modality. We also identify areas which we will improve for future iterations of similar multi-modal interaction applications due to the findings identified in this paper. These improvements should alleviate shortcomings with our current design and provide further opportunities to research multi-modal interaction.

Keywords: Multi-modal · Interaction · Multi-touch · Modality

1 Introduction

New input devices (e.g., Leap Motion, Intel RealSense, Microsoft Kinect, Tobii EyeX) are changing the 3D user interface landscape [1], by enabling more intuitive (or "natural") interaction. This interaction is more intuitive as social action and interaction is inherently multi-modal and represents a significant portion of human activity. For example, Stivers et al. defined face-to-face interaction as multi-modal through the composition of the vocal and visuospatial modalities [2]. However, combining various modern input devices has created many challenges for user interaction research. Many of these devices enable unimodal interaction (vision-based, touch-based, speech recognition, etc.) as such the difficulty comes from attempting to combine the disparate modalities that typically are ignorant of one another.

Regardless of the difficulty, multi-modality has been the focus of a continuous research effort by the Human-Computer Interaction (HCI) community due to the promising improvements to user interaction. For example, Jourde et al. created an editor using their Collaborative and Multimodal (COMM) notation, allowing users to identify relationships between them and their devices [3]. In another

© Springer International Publishing AG, part of Springer Nature 2018
M. Antona and C. Stephanidis (Eds.): UAHCI 2018, LNCS 10907, pp. 299–308, 2018.
https://doi.org/10.1007/978-3-319-92049-8_22

study, Prammanee et al. proposed an architecture to discern the modalities of different devices [4]. In this paper, we present design strategies for multi-modal interaction and the challenges faced during the construction of Multi-Modal Interactive Paint (MIP).

2 Background

Modality can be defined as the mode in which something is experience, expressed or exists but is also used in the context of sensory perception. In this paper, we use both because each definition is applicable to our uses. For example, a touch gesture is an expression which is also a modality, but touch is also a sensory perception. With this definition, we can state that multi-modality is a combination of two or more modalities that can be a composition of touch, vision, speech recognition, etc. For example, there may be the utilization of touch and vision when identifying attributes of an object such as shape, color, and texture. The set of modalities that we discuss are those who mainly have the capability for active interaction such as touch and vision. Others are also useful for interaction even though they do not provide direct input to an application but we will not discuss those at any length in this document as our primary focus is on input devices and the previously discussed modalities.

This area interests us because multi-modal interaction is the default state of social interactions. In social situations, we incorporate various modalities when engaging others. The two central modalities in social situations such as face-to-face as discussed previously in the introduction are vocal and visuospatial. Stivers et al. define visuospatial as consisting of manual gestures, facial expressions, and body posture [2]. These elements of the visuospatial modality may be culturally dependent. For example, a hand with the index and middle finger forming a “V” shape has a significant difference in American and British culture. However, regardless of differences in the precise details of these elements all social actions and interactions are comprised of the visuospatial modality and the vast majority incorporate the vocal modality when face-to-face. This multi-modal interaction is regardless of cultural, racial or national background. Such a universal, inherent and instinctual form of interaction is why multi-modal interaction shall be a significant and critical area of research in the field of HCI.

3 Motivation

The motivation of MIP is to create a fun application that can be used to test multi-modal interaction (from an input perspective) while developing strategies to lessen the challenges in this type of interaction. Most importantly, we are interested in improving user interaction towards a more intuitive experience.

4 Multi-modal Interactive Paint

This section provides information about MIP.

4.1 An Overview

In the quest for multi-modal interaction, we asked ourselves what will be the best application that can demonstrate true multi-modal interaction? We concluded that a painting application would be the best demonstration of true multi-modal interaction. This conclusion was due to a painting application providing a complex environment for testing real-time interaction while being a fun application to use, and most importantly, it has demonstrable benefits in aiding children by helping to identify their psychological profiles [5], which can also help research in multi-modal interaction for education. In addition to the complexities inherent in multi-modal interaction, studying children's interactions provide an additional layer of complexity. This additional complexity is due to the developmental differences between adults and children with regards to cognition. The difference necessitates recognition algorithms which can account for the increase in complexity. The aforementioned helps to make the case that multi-modal interaction design strategies will be different if used by children versus adults [6].

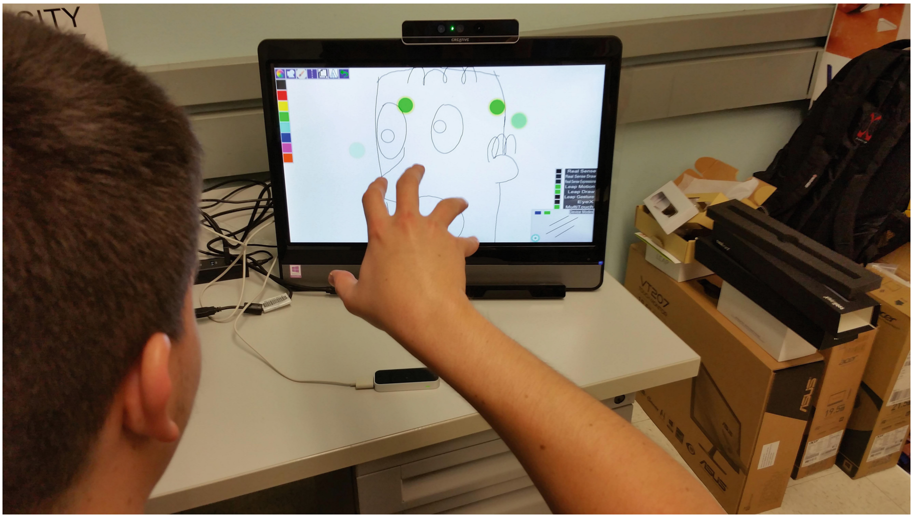

Fig. 1. Leap-motion and real-sense demo

This desire to create a proper demonstration of true multi-modal interaction lead to the creation of MIP. This application is a painting (drawing) application with multi-modal support, as shown in Figs. 1 and 2. The motivation to create MIP was due to the previous benefits discussed with regards to aiding children, among others. For example, painting in general not only helps develop a board psychological profile but can assist in identifying psychopathologies [5]. Currently supported modalities and their corresponding input device can be seen in Table 1.

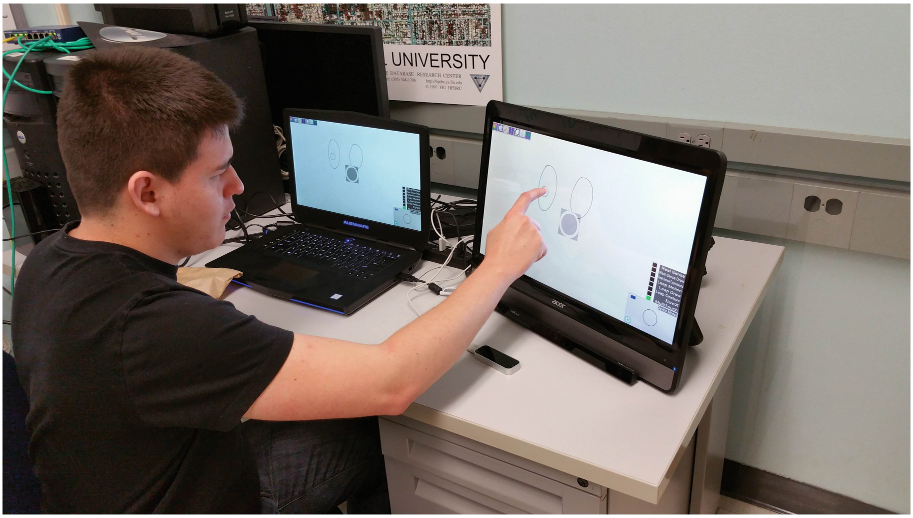

Fig. 2. Multi-touch demo

Table 1. Supported modalities and their corresponding input device

Modality	Input device
Touch	Multi-touch display
Visual	Tobii EyeX
Manual gesture	Leap motion
Facial gesture	Intel RealSense

Users can use any the modalities provided by the input devices as well as any of the built-in tools. Currently, built-in MIP toolset is a composition of basic shapes, random color selection, save and undo operations, layers, and image mirroring. The set of basic shapes includes circles, squares, and triangles which each support activities such as filling and transparency Additionally, the modalities described previously add additional functionality which can trigger upon performing a specific set of gestures. For example, the Intel RealSense camera has support for facial gestures, such as smile and puffy face. The former gesture allows shapeshifting, and the latter changes shape color. The Leap Motion device is capable of triggering the *Save* feature upon performing a double-tap gesture; performing rotation gestures as well as left and right swipes triggers the shape selection capability. The Tobii EyeX tracker enables the capability of drawing using the gaze of the user. The Tobii EyeX tracks the user's point of gaze on the computer screen and draws at the position where the user is looking. Perhaps, the most important feature is the coordination of the input devices enabled by MIP. For example, if the user has a multi-touch display and a Leap Motion, the latter will be used only for specific functions in the system by default. However, these functions can be overridden by the user as they may also decide to turn it into a drawing device. This flexibility allows for a rich set of interaction options for which can provide a tailor-made experience for the user matching their exact preferences if they so choose. Moreover, the number of responsibilities of a device may decrease if more devices become available to the system.

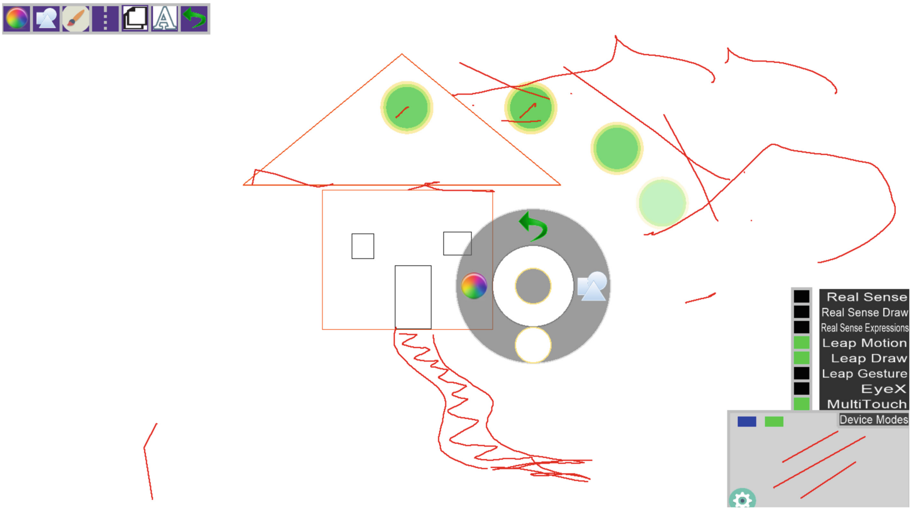

Fig. 3. Multi-modal canvas

Currently, MIP restricts all of the modalities provided by the input devices to 2D painting. However, the application makes use of LibCinder and OpenGL as core components. These components provide the possibility of adapting MIP for 3D painting. MIP for 3D painting would enable other industrial applications in addition to the benefits to psychologists that 2D painting brings. The industrial applications may include graphics professionals who paint onto 3D models, architects who create 3D drafts and other professionals. Also, the same therapeutic benefits from MIP for 2D painting should hold in a 3D painting environment.

MIP is also useful as a baseline for future device enablement and will serve as a testbed for new input devices. This testbed will allow us to tune new input devices for modalities that we have previously used since we will already have previous experience with user interaction with that specific modality. Having this ability enables rapid enablement for other projects.

4.2 User Interface

The MIP application's user interface consists of two major and ever-present components: (i) the Mode Buttons, (ii) the Mode Box. Both of these can are present in Fig. 3 as well as a couple of contextual menus.

- **Mode Buttons** – Top left corner of the User Interface (buttons from left to right). Enables the user to specify their input and organize their painting.
  - **Color Change Button** – Allows the user to make a color selection in the same manner they make in unimodal applications such as Microsoft Paint.
  - **Change Button** – Allows the user to make a shape selection, where users can choose from Lines, Circles, Rectangles or Triangles. A shape is filled (with the color selection) or unfilled (transparent) based on the user inputted value of the Toggle Fill Button.
  - **Buttons** – This button has several buttons nested which become available to the user upon selection. The buttons are: (i) Toggle Fill Button (enable or disable transparency), (ii) Line Size Increase Button, (iii) Line Size Decrease Button, (iv) Transparency Increase (increase opacity), (v) Transparency Decrease (decrease opacity).
  - **Toggle Symmetry** – Enforces symmetric input by reflecting input over a user selected axis.
  - **Layer Visualization** – Allows users to select different layers for painting. This function is similar to those found in popular applications such as Krita and Adobe Photoshop.
- **Mode Box** – Bottom right corner of the User Interface. Informs the user of currently selected settings. For example, if the user is entering a line, a set of lines will be shown in the box in the color that the user currently has selected. Also, the box display information of devices at the beginning of the MIP application's execution (launch) and allows the user to select devices that they would like to enable or disable. Finally, there is a cog icon in the bottom left corner of the Mode Box that allows diagnostic information to be displayed in MIP such as Frames Per a Second (FPS). All of these present in Fig. 3.

5 Design Practices for Multi-modality

During the design of MIP, we were able to find many challenges that provide insight into the design and research directions for multi-modal interaction. Some of these challenges are more apparent than others which are only found while conducting live testing of a multi-modal application and even then only with some modalities. The following list summarizes some of those challenges and their suggested solutions or areas of further research interest:

- **InfraRed (IR) technology** – having multiple devices with infrared emitters creates a problem if their respective fields of view (of the camera) intersect their counter emitters. The best solution is to use at most one IR device (e.g., Leap Motion) and provide other means of recognition – in particular, the use of optical tracking. For example, Hu et al. accomplished gesture recognition by using two optical cameras, achieving accuracies of 90% when detecting the position of fingers and hands [7].
- **Gestures and Children** – As previously discussed children provide a unique challenge when trying to form an interactive experience in not only multi-modal interactive experiences but even unimodal interactions. Lisa Anthony and Colleagues have advanced the field in this area. They provide the following recommendations, which have been useful in our application [6]:
  - It is essential to prevent unintended touch interactions
  - Use platform-recommended sizes (and larger whenever possible)
  - Increase the area for the widgets
  - Align targets to the edges
  - Use tailored gesture recognition for children
- **Gesture Discovery** – it is best to keep gestures simple whenever possible and the total number small. One option for finding the best gestures is to use a Wizard-of-oz experiment, such as the one done by Wobbrock et al. [8]. However, there is a divide in the 3D User Interface community in their acceptance of this approach. Differences in gesture discovery can also be seen when a set of gesture is developed by users as seen in Balcazar et al. [9].
- **Paper Prototyping:** Paper prototyping may be beneficial at the early stages of development to showcase different ideas as shown in Figs. 4 and 5. Paper prototypes allow for rapid iteration and feedback when creating a user interface as in the figures. This rapid iteration allows less development time to be spent refactoring boundary class code and more time spent improving code quality. Also, having this provides materials which can be shown during the requirements elicitation phase of development.
- **Modes** – What role does each input device undertake for user interaction and how do the input devices cooperate? This question is fundamental when devising a multi-modal application that relies on several unimodal input devices, and a solid plan of action must be developed to coordinate the devices to provide meaningful user interaction. To address this question, we created different modes depending on the available devices and the preferences of the user. Having different modes allows for more dynamic modality as devices can come online or offline at any time, as such the responsibility of a device at these events is a crucial component of multi-modal interaction. However, this is only one question, and there are others that must be addressed and answered (see examples in the previous section). We believe that the discovery of the right combination of responsibilities between input devices and the interaction they drive will be an important research direction.

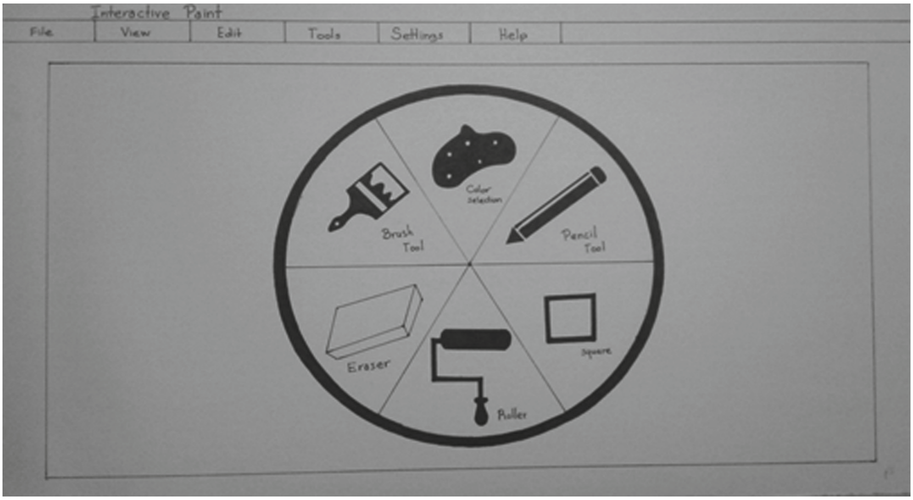

Fig. 4. Radial menu

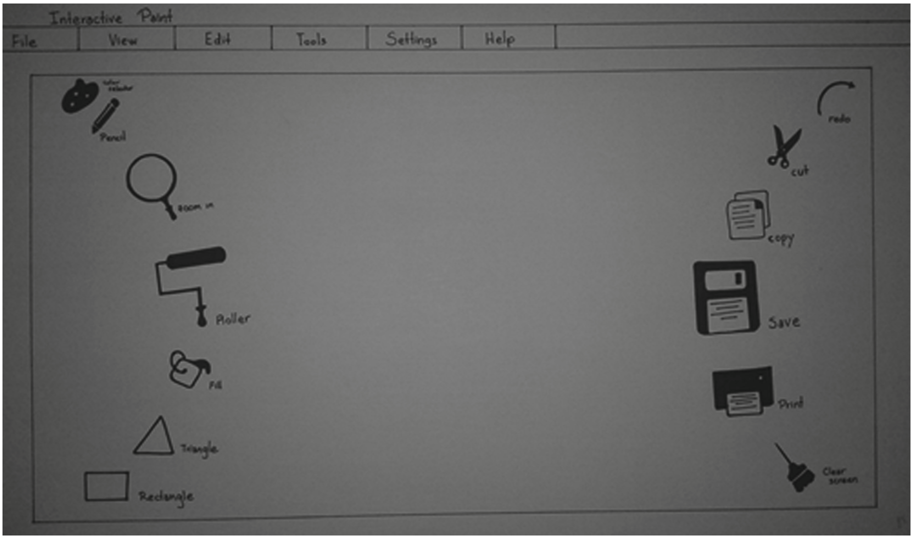

Fig. 5. Double-sided menu

6 Conclusion and Future Work

We have presented the Multi-Modal Interactive Paint application which combines several unimodal input devices to present the user with a cohesive multimodal interactive experience. As discussed previously the MIP application has the possibility of assisting children by helping psychologists develop psychological profiles and identify psychopathologies. MIP helps psychologists by providing an interactive painting experience for children with an intuitive form of input. By accepting that children are one of the targeted audiences MIP can prepare for interaction by more general audiences because children are a more difficult audience for multi-modal interaction because their level of cognitive development is vastly different from adults. There is an increased challenge because it is difficult to extract meaningful information modalities of touch and manual gestures when children interact with the application.

We have also identified software design and research problems for multi-modal interaction with our painting application. The simplest among these for others to test when designing software is the paper prototyping as the tools for doing so are widely available across the globe. This rapid and iterative process should have the capability to provide improved multi-modal interactions as it helps developers consider their ideas in a more tangible medium. The problem with IR cameras that we identified should allow other developers to avoid similar mistakes by avoiding multiple IR cameras and instead sticking to a single IR camera, such as the Leap Motion. While this may be a well-known problem, it is imperative to keep it in mind when developing multi-modal technologies with IR cameras. In particular, the need for non-IR cameras are critical in multi-modal interaction. One of the most intriguing research problems we have identified is finding the most appropriate combination of input device responsibility. A finer tuned set of responsibilities will enhance user interaction.

6.1 Future Directions

Future work includes addressing essential questions, such as the mode-switching of devices, how devices interact with each other, and the best approach for user interfaces when dealing with multi-modal applications. Our highest priority is to conduct formal user studies to identify any issues with the interaction in MIP further and record notable observations to drive an more intuitive form of interaction. Formal studies will be conducted similarly to other OpenHID studies such as those found in CircGR [9] The next version of an improved MIP is at https://goo.gl/3oxWv0. In the next version of MIP, we shall improve the user interface of the application. Also, we shall expand the list of input devices to include the Microsoft Kinect, MicroChip 3D multi-touch, active pen and motion sensors. Furthermore, the use of more stereo-vision (i.e., without IR technology) is critical. Another intriguing avenue of exploration is exploiting the vocal modality as discussed in [2] it is a commonly utilized modality that leads to natural and intuitive interaction. Techniques can be adapted from Adler et al. where the ASSIST system is defined in detail and enables natural conversation with the system [10]. This modality is significant as it can allow users with disabilities to benefit from the multi-modal interaction if other modalities are unavailable.

Acknowledgments. We acknowledge the support by the National Science Foundation under Grant Nos. I/UCRC IIP-1338922, AIR IIP-1237818, SBIR IIP-1330943, III-Large IIS-1213026, MRI CNS-1532061, OISE 1541472, MRI CNS-1532061, MRI CNS-1429345, MRI CNS-0821345, MRI CNS-1126619, CREST HRD-0833093, I/UCRC IIP-0829576, MRI CNS-0959985, RAPID CNS- 1507611. Also, we acknowledge Garrett Lemieux and Andrew Mitchell for their help with the first prototype.

References

1. Ortega, F.R., Abyarjoo, F., Barreto, A., Rishe, N., Adjouadi, M.: Interaction Design for 3D User Interfaces. The World of Modern Input Devices for Research Applications, and Game Development. CRC Press, Boca Raton (2016)
2. Stivers, T., Sidnell, J.: Introduction: multimodal interaction. Semiotica **156**, 1–20 (2005)
3. Jourde, F., Laurillau, Y., Nigay, L.: COMM notation for specifying collaborative and multimodal interactive systems. In: Proceedings of the 2nd ACM SIGCHI (2010)
4. Prammanee, S., Moessner, K., Tafazolli, R.: Discovering modalities for adaptive multimodal interfaces. Interactions **13**, 66–70 (2006)
5. Khorshidi, S., Mohammadipour, M.: Children's drawing: a way to discover their psychological disorders and problems. Int. J. Ment. Disord. **14**, 31–36 (2016)
6. Anthony, L., Brown, Q., Tate, B., Nias, J., Brewer, R., Irwin, G.: Designing smarter touch-based interfaces for educational contexts. Pers. Ubiquit. Comput. **18**, 1471–1483 (2013)
7. Hu, K., Canavan, S., Yin, L.: Hand pointing estimation for human computer interaction based on two orthogonal-views. In: Pattern Recognition (ICPR), pp. 3760–3763 (2010)
8. Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-defined gestures for surface computing. In: The SIGCHI Conference, pp. 1083–1092. ACM Press, New York (2009)
9. Balcazar, R., Ortega, F.R., Tarre, K., Barreto, A., Weiss, M., Rishe, N.D.: CircGR: interactive multi-touch gesture recognition using circular measurements. In: Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces, ISS 2017, pp. 12–21. ACM, New York (2017)
10. Adler, A., Davis, R.: Speech and sketching for multimodal design. In: ACM SIGGRAPH 2007 Courses, SIGGRAPH 2007. ACM, New York (2007)