GAWSCHI: Gaze-Augmented, Wearable-Supplemented Computer-Human Interaction

Vijay Rajanna,* Tracy Hammond†
Sketch Recognition Lab. Dept. of Computer Science & Engineering
Texas A&M University, College Station, Texas

Abstract

Recent developments in eye tracking technology are paving the way for gaze-driven interaction as the primary interaction modality. Despite successful efforts, existing solutions to the "Midas Touch" problem have two inherent issues: 1) lower accuracy, and 2) visual fatigue that are yet to be addressed. In this work we present GAWSCHI: a Gaze-Augmented, Wearable-Supplemented Computer-Human Interaction framework that enables accurate and quick gaze-driven interactions, while being completely immersive and hands-free. GAWSCHI uses an eye tracker and a wearable device (quasi-mouse) that is operated with the user's foot, specifically the big toe. The system was evaluated with a comparative user study involving 30 participants, with each participant performing eleven predefined interaction tasks (on MS Windows 10) using both mouse and gaze-driven interactions. We found that gaze-driven interaction using GAWSCHI is as good (time and precision) as mouse-based interaction as long as the dimensions of the interface element are above a threshold (0.60" x 0.51"). In addition, an analysis of NASA Task Load Index post-study survey showed that the participants experienced low mental, physical, and temporal demand; also achieved a high performance. We foresee GAWSCHI as the primary interaction modality for the physically challenged and a means of enriched interaction modality for the able-bodied demographics.

Keywords: Eye tracking; Gaze interaction; Multi-modal interaction; Wearable devices; Quasi-mouse; Foot-operated device; Midas Touch

Concepts: •Human-centered computing → Interaction devices; *Interaction paradigms; Interaction techniques; Accessibility;*

1 Introduction

Gaze-driven interaction is a highly promising human-computer interaction modality among the recent multimodal interaction methods [Prasad et al. 2014]. Gaze points are the manifestation of visual attention [Duchowski 2007]; an ability to leverage gaze accurately, while inducing no cognitive load on the user will lead to highly contextual and rich interactions. In addition, eye tracking has a wide application area; it is specifically suitable for the applications that make limited use of keyboard input but rely heavily on the mouse input.

*e-mail: vijay.drajanna@gmail.com URL: www.vijayrajanna.com
†e-mail: thammond@gmail.com URL: srl.tamu.edu

ETRA '16, March 14-17, 2016, Charleston, SC, USA
ISBN: 978-1-4503-4125-7/16/03
DOI: http://dx.doi.org/10.1145/2857491.2857499

Despite the discernible advantages of gaze-assisted interactions there is still no widespread adoption of this technology as the primary interaction modality. In this work we present "GAWSCHI," a framework that offers a precise and effective solution to the "Midas Touch" problem [Jacob 1991]. Unlike existing solutions to the "Midas Touch" problem, which rely only on the gaze input, we use a supplementary wearable device, along with gaze input to execute a user's command at the point of regard. The wearable device is a quasi-mouse that is operated with the user's foot. The quasi-mouse communicates with the central GAWSCHI system, running on the computer, over Bluetooth. This wearable design lends itself to a small form factor, allowing it to be attached to the user's footwear (shoe), or placed anywhere on the floor to enable effortless and hands-free interactions. GAWSCHI uses a table top eye tracker from "The Eye Tribe" for tracking user's gaze on the screen. The Gaze interaction server, the central controlling system, runs on the computer; controls both the eye tracker and the quasi-mouse. The activation of the intended commands like click, hold, release, etc., are executed by looking at the object of interest and executing the corresponding commands through the foot-operated quasi-mouse. GAWSCHI uses the foot-operated quasi-mouse, when just using an explicit button, either on the keyboard or mounted on the desk, is another design option. The interaction design of quasi-mouse is in accordance with the primary goal of GAWSCHI: "Provide a hands-free, immersive interaction, where the user is able to simply lean back in the chair and perform mouse based actions." However, an explicit button will force the user to reach-out to the button, demanding hand movements and forward bending, which defeats the interaction principles of GAWSCHI.

2 Prior Work

Since Jacob presented his research on the potential of using eye tracking toward developing gaze-based interaction systems [Jacob 1991], there has been a significant research both in the development of accurate eye tracking systems and contemporary solutions to the "Midas Touch" problem. In the context of leveraging gaze as the primary interaction modality, while effectively addressing the "Midas Touch" problem, existing research can be classified across two broad categories.

2.1 Gaze-assisted Interaction on Desktop

Implementations like [Lankford 2000], [Kumar et al. 2007], etc., have demonstrated gaze-assisted interactions on the native interface of an operating system (Windows). However, such implementations continue to use dwell time, an external key, or magnification of the point of regard for target object selection. Hence, they fall short of the speed and accuracy of mouse-based interaction.

2.2 Gaze and Foot-based Interaction

Foot-operated computer input devices have always been well studied among the Computer-Human Interaction community. Pearson and Weiser [Pearson and Weiser 1986] conducted seminal work in this regard, as they presented the design of "Moles," foot-operated input devices similar to the mouse. Pakkanen and Raisamo [Pakkanen and Raisamo 2004] demonstrated the feasibility of feet input in non-accurate spatial tasks. Furthermore, Velloso et al. [Velloso et al. 2015], present a comprehensive survey of foot-operated interaction modality. Only Göbel et al. [Göbel et al. 2013], combined gaze and foot-input as an interaction modality. However, their implementation is specifically built for pan and zoom interactions; the authors mention that foot-interaction systems support coarse pointing interaction, and are not precise enough for pointing tasks.

Hence, an accurate, gaze-assisted interaction that integrates seamlessly with the native interface of an operating system to substitute mouse, while supporting most of the common interaction tasks still remains unresolved. GAWSCHI seeks to explore these limitations to *create a fully gaze and foot-based interaction modality to achieve precise "point and click" interactions*, while also supporting other interactions (double-click, click-and-hold, hold-and-release). Furthermore, GAWSCHI seeks to *improve the design of foot-operated devices through its wearable quasi-mouse that has a small form factor*. In addition, the design of the quasi-mouse only requires a *gentle press (minimal effort)*, with the foot, for executing the user commands (click), while still achieving a high performance.

3 System Architecture and Implementation

The goal of GASWCHI is to enable Gaze-Augmented Computer-Human Interaction that is noninvasive, inconspicuous, efficient, and accurate, while inducing no physical strain or cognitive load on the user. Hence, we believe any incremental development over the existing solutions, which use either dwell time or a blink to address the "Midas Touch" problem, will not help toward achieving a seamless integration, with an efficient and accurate gaze-mediated interaction on the native interface of a computer. In our approach, we move beyond the blink and dwell time based approaches for activation of the point of regard. We have invented a wearable, foot-operated quasi-mouse. The framework consists of three primary modules: 1) Gaze Interaction Server, 2) Eye Tracking Module, and 3) Foot-operated Quasi-mouse. A model of the system is shown in Figure 1.

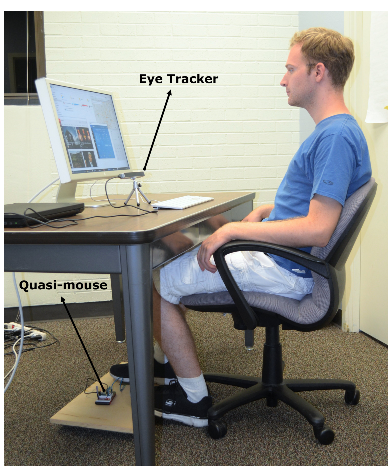
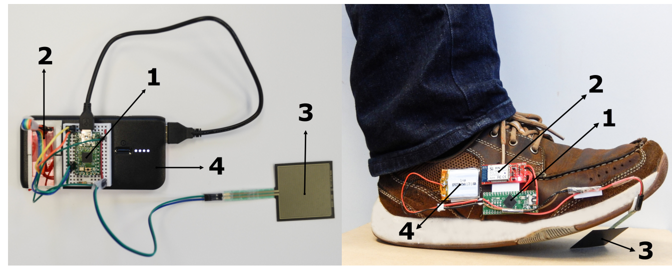


Figure 1: *A User Interacting with a Computer Using GAWSCHI*

3.1 Gaze Interaction Server

The Gaze Interaction Server, the central module, achieves desired interactions by mediating between the Eye Tracking Module and Foot-operated Quasi-mouse. The system is developed on the Eye Tribe SDK[1]; it runs on a computer to which both the eye tracking module and foot-operated quasi-mouse are connected over USB



(a) (b)

Figure 2: *Two Versions of the Quasi-mouse: (a) Quasi-mouse on the floor (b) Quasi-mouse worn on the footwear. 1) Microcontroller, 2) Bluetooth Modem, 3) Force Sensitive Resistor, 4) Battery*

and Bluetooth respectively. The gaze interaction server implements the algorithm to navigate the cursor on the screen by using smoothed gaze coordinates received from the eye tracker. In addition, the Gaze Interaction Server is also responsible for executing the commands issued by the user, at the point of regard. The commands received are encoded using a predefined single byte character, which are then decoded into an appropriate mouse command, and executed instantaneously at the current (x,y) mouse coordinates on the screen.

3.2 Eye Tracking Module

GAWSCHI uses "The Eye Tribe" tracker[1], an eye tracking system that provides a pair of (x,y) screen coordinates, based on where the user is looking. The Eye Tribe tracker is a table top eye tracker that is placed on a tripod, below the monitor. To work with the eye tracker, the user position is adjusted such that the face is centered in front of the monitor at a distance of 45 - 75 cm[1]. The eye tracker computes (x,y) coordinates by extracting the information from the user's eyes and face while s/he works on a computer. Prior to using the eye tracker with GAWSCHI, the system is calibrated for each user to develop a unique model for the user's eye characteristics.

3.3 Foot-operated Quasi-mouse

A significant contribution of this work is the development of a wearable quasi-mouse. The user executes commands at the point of regard on screen by operating the quasi-mouse. The quasi-mouse allows for interactions like click, double-click, click-and-hold, and hold-and-release by pressing on a flexible pressure pad (3.5" x 1.75"). The user's action (pressing the pressure pad) is encoded into an appropriate character (numeric) based on the amount of pressure applied, and the encoded command is delivered to the Gaze Interaction Server over Bluetooth. The quasi-mouse is built with three main components: 1) Teensy Microcontroller[2], 2) Bluetooth Modem (BlueSMiRF)[3], and 3) Force Sensitive Resistor[3]. A pictorial depiction of the quasi-mouse is shown in Figure 2.

4 Experiment Design

GAWSCHI was evaluated with a comparative user study to understand its efficacy in substituting the mouse, specifically for "point and click" interactions, while performing common interaction tasks on a computer. Questions that we sought to answer are as follows:

[1] theeyetribe.com
[2] www.pjrc.com
[3] www.sparkfun.com/products

1. Can gaze-augmented interaction, supplemented with a wearable device for executing the user commands, perform as good (time and precision) as mouse-based interaction?
2. What effects does GAWSCHI have on the mental, temporal, and physical demand of the users? How is their performance impacted?
3. How well is the design of GAWSCHI accepted by the users?

We recruited 30 participants, belonging to diverse ethnic groups (USA, Mexico, Europe, India, China, and Korea) since the physiological structure of the eye differs between ethnicities impacting the eye tracking accuracy. The participants pool consisted of 26 males and 4 females, all either graduate or undergraduate students, aged between 20 and 43 ($\mu_{age} = 24.7$); none had previously used gaze-interaction. For the user study we used a 23" monitor, with a 1900 x 1200 resolution (19.5" x 12.19" screen size), and a PPI of 98.44 (Pixels Per Inch). Each user was first briefed about the experiment; subsequently presented with a demo of each required activity, using both the mouse and gaze-driven interactions. Furthermore, each participant was put through a practice session for a maximum of five minutes. Following the completion of the experiment, each user completed a NASA TLX survey and a quantitative evaluation of the system. We have excluded data collected from nine participants because of one or more of the following reasons: a) calibration failure, b) the participant was using thick spectacles for vision correction, or c) the participant could not complete all the tasks.

During the experiment each user performed eleven interaction activities on a computer (running Windows 10). The task execution was counterbalanced, with 50% of the users performing the eleven tasks first with the gaze-driven interaction, and subsequently performing the same tasks using the mouse; the remaining 50% of the users followed the reverse order. The interaction activities chosen are common ones, performed routinely on a computer by a user; the tasks were elicited from interviews and a pool of common tasks. The eleven tasks performed during the experiment are: 1) Click on the Windows start menu, open the News application, open the first article, and scroll through the article by clicking on the side scroll bar. 2) Click on the Windows start menu, open the Weather application, and report the current temperature. 3) Open Chrome web browser from the toolbar, log-in to Gmail (credentials pre-saved) by clicking on its bookmark, and open the first email. 4) Open Chrome web browser from the toolbar, open YouTube by clicking on its bookmark, play any video, pause, switch to full screen mode, and switch back to the normal screen mode. 5) Open Chrome web browser from the toolbar, open Google Maps by clicking on its bookmark, click on a point of choice, and switch to Google street view. 6) Open the "Images" folder on the desktop, and open an image. 7) Open the "Documents" folder on the desktop, and open the first pdf file. 8) Run the "Notepad" application by clicking on its icon on the desktop. 9) Click on the Windows start menu, open Calculator, and compute 5+7. 10) Open the Calendar application from the toolbar, and view today's schedule by clicking on the toolbar item 'day.' 11) Click on the "Task view" icon on toolbar, and switch to the top left task.

5 Results

During the user study, the time taken to perform each task, both with the gaze-driven interaction and mouse-based interaction, was recorded for each participant. Figure 3 shows the mean time taken to perform each task, and the standard deviation as an error bar, for both mouse and gaze-based interactions. Figure 3 also shows Δt, the time difference between the mean times to perform a task using gaze and mouse-based interactions. $\Delta t = GazeInteractionMean - MouseInteractionMean$. Δt compares the time taken by gaze-driven interaction to mouse-based interaction, hence a positive value of Δt indicates that gaze-driven interaction consumed Δt time more than the mouse-based interaction. Conversely, a negative value of Δt indicates that gaze-driven interaction consumed $|\Delta t|$ time less than the mouse based interaction.

Table 1: *Matched-pairs t-test, Two tailed, 95% Confidence Interval*

Task	Mean Time Diff $\bar{d}$	Standard Error $SE(\bar{d})$	t-stat	p
1	2.70	1.40	1.93	0.068
2	-1.39	1.09	-1.27	0.219
3	-3.32	0.91	-3.65	0.002
4	-5.38	1.47	-3.67	0.002
5	-1.52	1.31	-1.17	0.257
6	-1.48	2.04	-0.73	0.476
7	-0.73	0.59	-1.25	0.227
8	-2.43	1.24	-1.95	0.065
9	-5.52	1.34	-4.12	0.001
10	-1.49	1.21	-1.23	0.234
11	-0.59	0.40	-1.45	0.161

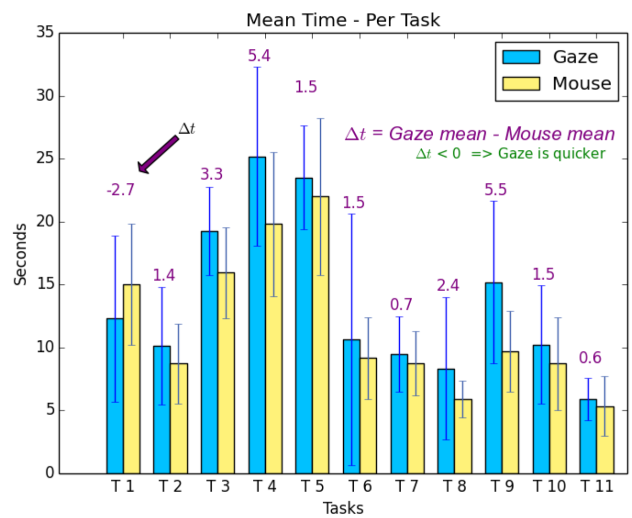


Figure 3: *Mean Time Taken for Each Task, Std Dev, and* Δt

To better understand the statistical significance of per task mean time differences, we performed a matched-pairs t-test for each task since our study involves the same participants performing a set of eleven tasks, using both mouse and gaze based interactions. Our null hypothesis is: "For a given task, there is no time difference between the gaze and mouse-interaction modalities." Results from the two tailed matched-pairs t-test with a confidence interval of 95% ($\alpha = 0.05$) is shown in Table 1. We observe that the p-value >0.05 for all the tasks, except for tasks 3, 4, and 9. Hence (except for 3, 4, and 9), we fail to reject the null hypothesis that there is no time difference between the two interaction modalities, and infer that the gaze-driven interaction using GAWSCHI is as good as mouse-based interaction for the majority of tasks. For tasks 3, 4, and 9, we reject the null hypothesis and accept the alternate hypothesis that there exists a time difference between the two interaction modalities; the **reasons for rejection of the null hypothesis** are elaborated in the discussion section (6).

6 Discussion

Through the system evaluation, we have tried to answer the questions listed in section 4. The experiments conducted and the results of statistical tests show that gaze-based interaction using GAWSCHI is as good as mouse-based interaction in a majority of the tasks (question 4.1). However, GAWSCHI is slower than the mouse on tasks 3, 4, and 9. Further analysis of the dimensions of all UI elements involved in all the tasks reveal that for a user to interact with GAWSCHI the dimensions of the UI element must be above 0.60" x 0.51", which is the smallest dimension of a UI element that users were able to comfortably interact with in our study. As noted, GAWSCHI takes more time on tasks 3, 4, and 9. The reasons include: a) Tasks 3 and 4 are browser-based tasks, and the dimensions of most of the interface elements in these tasks are less than 0.60" x 0.51", and b) Task 9, where users perform an addition task on a calculator, the proximity of numbers and a minimal gaze drift lead to higher time consumption. In all the other tasks, where the gaze interaction matches the performance of mouse, the dimensions of UI elements are higher than the threshold. Also, the UI elements are well spaced; hence, no misclicks were observed.

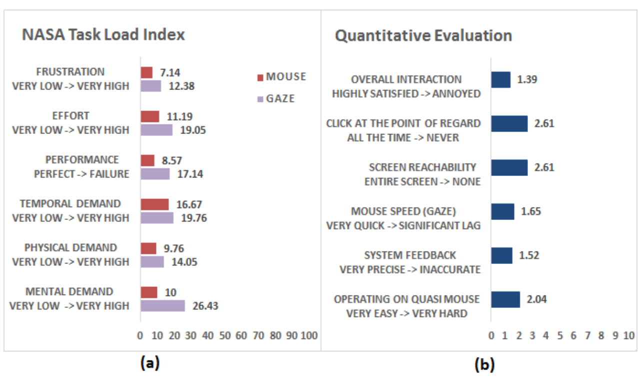


Figure 4: *(a) NASA Task Load Index for Mouse and Gaze Interaction (lower is better), (b) Quantitative Evaluation for Gaze Interaction (lower is better)*

Furthermore, we verify our hypothesis that the design of a gaze-interaction system, like GAWSCHI, that isolates user actions to a foot-operated quasi-mouse, demands minimal user efforts (question 4.2). Though comparison of NASA TLX scores for an entirely new interaction modality like gaze against a highly familiar modality like mouse is impractical, from Figure 4.a it can be observed that, despite TLX scores for gaze being marginally higher than mouse, each TLX score for gaze is still lower than a high workload threshold value of 40 as used in previous studies, e.g., [Knapp and Hall 1990]. Hence, though interaction using GAWSCHI is not simpler than mouse, we infer that users do not experience physical or mental workload (fatigue) when using GAWSCHI. Lastly, Figure 4.b shows a quantitative evaluation of the system, based on the Likert scale (lower score is better), on various design aspects. It can be observed that the users are highly satisfied with the overall interaction (question 4.3). In addition, users rated high for screen reachability, mouse speed, system feedback, quasi-mouse ease of use, and their ability to click at the point of regard.

7 Conclusion and Future Work

In this work we present GAWSCHI: a Gaze-Augmented, Wearable-Supplemented Computer-Human Interaction framework. The framework comprises of an eye tracker that drives mouse navigation on the screen using gaze; the user executes the intended command, like click, at the point of regard through foot-operated quasi-mouse. The quasi-mouse has a small form factor allowing it to be worn on the user's footwear or placed on the floor. Functionality supported by the quasi-mouse include click, double-click, click-and-hold, and hold-and-release. From a user study involving 30 participants, we found that the gaze-driven interaction using GAWSCHI is as good (time and precision) as mouse-based interaction as long as the dimensions of the interface element are above a threshold (0.60" x 0.51"). An analysis of NASA Task Load Index post-study survey showed that, when using GAWSCHI, the participants experienced low mental, physical, and temporal demand, while achieving a high performance. In the future work, we will explore the other potentials of GAWSCHI like learning effects, and the possibility of outperforming the mouse with increased familiarity.

Acknowledgements

We would like to thank the members of "Sketch Recognition Lab" for their support in the development of this project. Also, we are thankful to "The Eye Tribe" for providing us with an Eye Tracker, and "Microsoft Corporation" for providing us with a Surface Hub for Sketch, Gesture Recognition, and Eye Tracking research.

References

DUCHOWSKI, A. 2007. *Eye tracking methodology: Theory and practice*, vol. 373. Springer Science & Business Media.

GÖBEL, F., KLAMKA, K., SIEGEL, A., VOGT, S., STELLMACH, S., AND DACHSELT, R. 2013. Gaze-supported foot interaction in zoomable information spaces (interactivity). In *Proceedings of the Conference on Human Factors in Computing Systems - Extended Abstracts*, ACM.

JACOB, R. J. 1991. The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Transactions on Information Systems (TOIS) 9*, 2, 152–169.

KNAPP, B., AND HALL, B. 1990. High performance concerns for the trackwolf system (ari research note 91-14). *Alexandria,(VA): ARI*.

KUMAR, M., PAEPCKE, A., AND WINOGRAD, T. 2007. Eyepoint: practical pointing and selection using gaze and keyboard. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, 421–430.

LANKFORD, C. 2000. Effective eye-gaze input into windows. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, ACM, 23–27.

PAKKANEN, T., AND RAISAMO, R. 2004. Appropriateness of foot interaction for non-accurate spatial tasks. In *CHI'04 extended abstracts on Human factors in computing systems*, ACM, 1123–1126.

PEARSON, G., AND WEISER, M. 1986. Of moles and men: the design of foot controls for workstations. In *ACM SIGCHI Bulletin*, vol. 17, ACM, 333–339.

PRASAD, M., RUSSELL, M., AND HAMMOND, T. A. 2014. Designing vibrotactile codes to communicate verb phrases. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 11*, 1s, 11.

VELLOSO, E., SCHMIDT, D., ALEXANDER, J., GELLERSEN, H., AND BULLING, A. 2015. The feet in human–computer interaction: A survey of foot-based interaction. *ACM Computing Surveys (CSUR) 48*, 2, 21.