Multimodal Smart TV Interface for Multisensory Learning: Scenarios and Performance Evaluation

Sheng-Ming Wang, *Member, IEEE*, Winston Yang and Yu-Rong Lin

***Abstract*—This paper discusses the design and performance assessment for tangible interaction natural user interface of smart TV(TIxNUI). Through a systematic application of design thinking methodologies, we identified key features for TIxNUI. Subsequently, we conducted user training and performance evaluations using a prototype of the system. The results suggest that gestural modalities are effective: hybrid modals that include gaze and voice could be viable.**

I. INTRODUCTION

The availability of gesture input was expected to provide a wealth of possibilities for multimodality in education. As of today, few examples of their applications are found in classrooms. Education is still predominantly conducted through visual and auditory media; lectures and videos are still effective, given the combination of contents and users in most classrooms. Consequently, researchers are focusing their efforts on specific use cases and scenarios designs for multimodality in education [1]. This paper will discuss the specific scenarios for natural user interface (NUI) for education. Accordingly, it presents TIxNUI features that are needed to provide affordances in those scenarios. Subsequently, we will discuss our findings in training students to use a TIxNUI prototype that combines gestural recognition with voice and eye tracking.

Body gestures have been an integral part of our communications. Many studies have shown that gestures can enhance education by increasing motivation for students. Additionally, learning with gestures encourages communication and collaboration. For example, Shakroum proves that using Kinect in learning systems increases the students' *intrinsic motivations* which led to better learning outcome[2]. Ghisio argues that body performance is integral to many fields of education, such as music and math. Wherever spatial and temporal reasoning is applicable, multimodality should be used to created multisensory learning experience[3].

In recent years, the focus on NUI research is on the design and orchestration of learning activities, where NUI is part of a larger system. NUI could be deployed to provide motivation in specific areas along the learning path. Additionally, they are used in situations where collaboration and socialization is desirable. An area of high interest for NUI application is special education. Researchers have been creating serious games that allow students to use gestures to participate in activities of learning, playing, and diagnostics. Many studies suggest that special needs students became more aware of their surroundings, experienced a higher level of focus and become more engaged with teachers and peers. NUI could also be used for early detection of learning difficulties in children [4]. Troubles in spatial temporal cognitions are early signs of learning disabilities.

NUI and multimodalities are often applied in scenarios of classroom orchestrations. In these teaching methodologies, lectures and readings are supplemented with activities that put new knowledge to test. Modalities are utilized where motivations and socialization is needed.

Researchers generally acknowledge that NUI are not ready for widespread use in every classroom. Gestural inputs suffer from machine errors and users' fatigue. Eye trackers still lack precision, although they excel in fast targeting. Speech input is still not reliable. In order for NUI to be applied effectively right now, they need to be part of a multimodality system. The deficiency in one modal could be compensated by another modal. Additionally, NUI should be applied only for short tasks. They tend to perform more slowly compared to traditional input devices, such as touchscreen and mouse, compounding the problems of user fatigue.

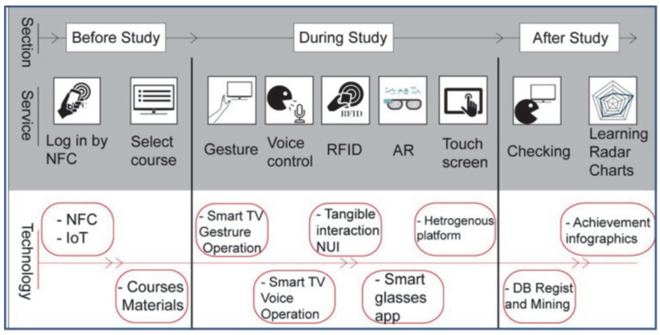

Figure 1. Service Blueprint of TIxNUI

II. TIxNUI SYSTEM REQUIREMENT

TIxNUI was envisioned through service blueprint approach to orchestrate touch points using a variety of modalities(see fig.1) . Based on Wang's previous classroom scenarios and the discussion of recent researches on NUI for education[1], we propose a multimodal learning system that could be deployed in the near future. This system will facilitate the implementation of classroom orchestrations and collect useful data for further research. The system should include features described below:

<u>Hybrid Gaze Modalities:</u> Tracking the gazes and eye paths of the students and teachers will provide a new modality for interactive learning. A combination of gaze and hand gestures will provide more efficient targeting compared to gestures alone. Eye contact could also provide critical data for special education.

<u>Expert Gestures for Motivated Users:</u> In use cases where the users are highly motivated, gestural interfaces could be deployed effectively. Depth cameras are error prone, because they do not always recognize userss gestures. One of the solutions is to use gestures that are easily recognized by RGB-D cameras. These gestures are created by experts. As

compared user-elicited gestures that are generated by user centric approach, expert gestures require motivated users to be trained. One of our research goals, is to study the viability of NUI in scenarios that involve users who are willing to train.

Smart TV: The television is a convenient gateway to introduce new technologies and services, because of its ubiquity. In an separate study to identify opportunities to improve the smart TV, Wang conducted an Quality Function Deployment study which concluded that smart user interface with adaptive UI layout would yield most benefit to users[5].

Familiar Interactions: Many studies on user elicited NUI gesture design have found that users do not always want to imagine original gestures derived from naive physics or real-world metaphors. They incorporate experiences from existing modalities on other devices. Mid-air mouse and touch screen gestures are often found on these experiments. One of the goals of our study is to assess if borrowing familiar interactions for a new modality would improve the user's performance.

III. PERFORMANCE EVALUATION

In order to verify the effectiveness of a multimodal system, a variety tests would need to be carefully designed and conducted. Each of the modality would need to be studied by themselves, and in combinations with one another. We initiated the investigation with the user training for gesture input. Among the available modalities, hand gesturing is the most complex and crucial, because it affords spatial precisions not available with gaze and voice. Users' proficiency with gestural input will be the foundation for their success in hybrid modalities.

A. Apparatus and Methodologies

Five testers were recruited for the tests. They are chosen based on their high affinity to learn and adopt to new technologies. The experiment is conducted on a smart TV, using an Intel Realsense Camera SR300 for NUI development and a Tobii Pro Glasses 2 for collecting users interaction data. Users' operation data were collected and studied through the MORAE software[6]. The content was an educational video library. The experiments are conducted in two stages. The first stage is training. Users are taught expert gestures using graphical illustrations. They could only proceed to the second stage, after passing a test. The second stage is in-system performance assessments. The users performed several tasks in the following modes: 1) mouse 2) gesture only 3) mid-air mouse + gesture. Subsequently, they completed system usability scale (SUS) surveys.

Fig. 2 TIxNUI prototype and samples of eye tracking data

B. Results and Discussion

Subjects using gesture alone consistently performed faster than the traditional mouse. Although the SUS scores were generally low, performance was acceptable. The mid-air cursor + gesture is the slowest. This could be the result of training. While we methodically trained the test subjects on the hand gestures, they had no practice on manipulating the cursor with their arms.

The gaze heat maps suggest that test subjects feel insecure about their memories on the gestures. Four out of the five testers constantly glanced at the *help* button on the top navigation bar, which would display an overlay window that indexes all the gestures. If there were an active gesture guidance system, either overlaid on the screen or projected onto another surface, the results in gesture inputs would improve.

IV. CONCLUSION AND FUTURE STUDIES

Results from this pilot study suggest that modalities in TIxNUI could be effective, if motivated users are properly trained. In future studies, performance assessments will be conducted on hybrid modalities that combine gestures with gaze and voice. Additionally, collaborative studies with educators will be conducted to test the system in specific scenarios.

Reference

[1] S.-M. Wang, S.-X. Lin, and C. Huang, "Smart TV tangible interaction natural user interface design in the scenario of future classroom," *International Journal of Advanced Computer Science and Applications,* vol. 6, no. 2, 2015.

[2] M. Shakroum, K. W. Wong, and C. C. Fung, "The influence of Gesture-Based Learning System (GBLS) on Learning Outcomes," *Computers & Education,* vol. 117, 2018.

[3] S. Ghisio, P. Alborno, E. Volta, M. Gori, and G. Volpe, "A Multimodal Serious-Game to Teach Fractions in Primary School," in *20th ACM International Conference on Multimodal Interaction*, Boulder, CO, 2017, pp. 67-70, New York, NY: ACM, 2017.

[4] E. Chatzidaki, M. Xenos, and C. Machaira, "A Natural User Interface Game for the Evaluation of Children with Learning Difficulties," presented at the 6th International Workshop on Child Computer Interaction, Glasgow, Scotland, UK, 2017.

[5] S.-M. Wang and C. J. Huang, "Service Design for Developing Multimodal HumanComputerInteraction for Smart Tvs," *International Journal of Advanced Computer Science and Applications,* vol. 6, no. 2, p. 8, 2015.

[6] "Morae from TechSmith.," ed.