Proceedings of The 6th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/2026.29761

Research on Multimodal Interaction Prototype Generation and Iterative Design for Real-Time User Testing

Yangbai Qi[1*], Xuanqing Huang[2], Siyao Mao[3]

[1]Henan University, Kaifeng, China
[2]School of the Art Institute of Chicago, Chicago, USA
[3]China University of Mining and Technology, Xuzhou, China
**Corresponding Author. Email: rara481846778@gmail.com*

Abstract. The lengthy development cycle and slow end-user feedback have seriously affected the advancement of typical interaction design workflows, especially in multimodal interfaces that integrate voice, gestures, and touch. This study proposes a method focusing on real-time end-user testing, leveraging a process that integrates design, testing, and optimization Achieve rapid construction and loop connection of multimodal prototype. Developed an integrated design-test-optimization module suite, which includes drag-and-drop combinations of functional building modules such as speech recognition and gesture tracking, as well as features and algorithms for real-time behavioral data. Thirty designers and 60 end users participated in this experiment. The results showed that the iteration time was shortened by 65%, the workload over multiple months decreased by 42%, and the overall workload was reduced by 37%. The optimized framework dynamically improved the prototype view by means of non-invasive data collection and rule-based adaptive learning. This model has the potential to transform the user-centered alignment design process into data adjustment, which accelerates and also expands the feedback loop between professors and users, as well as the paradigm shift based on the multimodal - computer model.

Keywords: multimodal interaction, prototyping, real-time testing, iterative optimization, user experience

1. Introduction

In the field of interaction design, it has long been believed that the speed of idea validation often determines the fate of a successful product [1]. However, the current multimodal interface workflow has added touch, gesture and voice input, and such a workflow is slow, disconnected and expensive. The most advanced prototyping software, although excellent in assembling visual interfaces, often fails to reproduce true multimodal behavior [2]. Therefore, design verification lags behind the requirements for it, leading to what is called the "feedback delay paradox", that is, feedback does not appear until long after the design is applicable.

Recent developments in low-code building and behavioral data analysis offer the possibility of bridging this gap. Low-code environments can quickly build prototypes, and embedded analytics can extract user preferences and interaction effectiveness from real-time data. However, the current

DOI: 10.54254/2755-2721/2026.29761

system still separates the testing phase from the design phase, which requires manual operators to export data and interpret it offline. This division breaks the iterative process and limits the designer's response to the constantly changing user demands [3].

This work proposes a cohesive method that combines prototype generation, field testing, and intelligent optimization in a closed loop. By integrating data collection and feedback mechanisms into the prototype, iterations are carried out in a real-time and self-correcting manner. The new framework is not only more efficient but also re-establishes the designer-user relationship from sequence to co-evolution within the same design cycle.

2. Literature review

2.1. Design status of multimodal interaction system

Multi-module interaction design has evolved from single-channel experiments to harmonious systems that combine vision, hearing and touch. The complementary synergy among patterns is a hallmark of modern systems, so voice input, gestures and touch events coexist without causing perceptual overload [4]. Despite these advancements, difficulties still exist. Designers often encounter asynchronous switching between modes, non-uniform delays between sensors, and variable semantic mapping between modes. Personalization is also limited. Systems rarely change dynamically based on users' preferences or environmental context, so they show lower engagement and efficiency in learning. Therefore, the integration of multimodal sensing and adaptive interface logic constitutes a key frontier in this field.

2.2. Evolution and limitations of rapid prototyping

Rapid prototyping has achieved the development from hand-drawn sketches to interactive digital simulations. Software such as Figma, Axure and Adobe XD have changed the way interfaces are visualized. However, they still remain at the two-dimensional level. They can neither describe multimodal input behaviors nor store real-time interaction data. Traditional prototyping relies on post-event testing at the end of the meeting [5]. Then, the results were explained, which led to a delay between feedback and iteration and also impaired responsiveness. With the emergence of data-intensive design tasks, there is an urgent need for a platform that can test, analyze, and simulate multi-modal prototypes within the same runtime cycle [6].

2.3. Shift in user testing methodology

User testing has evolved from laboratory observation to real-time, in-situ evaluation. Combining distant A/B testing, eye tracking and affective computing technologies can better understand behavior. However, all of these rely on sequential phased methods: data collection, offline processing, and lag adjustment. The lack of immediate analysis and synthesis leads to the loss of optimization opportunities [7]. Future development will be characterized by the "test as design" model, where user behavior directly notifies interface changes in the business environment. Under this model, testing is no longer an operation after design, but a natural component of dynamic iteration.

DOI: 10.54254/2755-2721/2026.29761

3. Methodology

3.1. Modular prototype generation framework

User testing has shifted from laboratory observation to real-time and on-site assessment. Integrating remote A/B testing, eye-tracking, and affective computing technologies can provide a deeper understanding of behavior.

All of this relies on a sequential and phased approach: Data collection is carried out, offline processing is implemented and lag adjustments are made. Due to the lack of immediate analysis and synthesis, optimization opportunities are lost. In future development, the "test as design" model will be a prominent feature. Interface changes in the business environment are directly triggered by user behavior. Based on this model, testing is no longer a link after the design is completed. Rather, it is a natural component of dynamic iteration.

To formalize synchronization, the multimodal transformation can be expressed as Equation (1) [8]:

$$S_{t+1} = f(S_t, I_v, I_g, I_s) = \alpha V_t + \beta G_t + \gamma S_t + \epsilon_t \tag{1}$$

where I_v , I_g , I_s represent voice, gesture, and system input streams respectively, and coefficients α , β , γ denote modality weighting. The residual ϵ_t captures transient latency discrepancies. This formulation ensures stable multimodal coordination and repeatable prototype behavior across tests.

3.2. Real-time user behavior data collection system

The collaborative data capture subsystem captures the end-user behavior of all channels at a frequency of 60hz, such as the accuracy of voice command recognition, gesture trajectory coordinates, fixed heat maps, task completion delays, and interaction errors [9]. Non-intrusive sensing reduces behavior distortion by avoiding external markers or capturing clues. The adaptive feedback module can be programmed to start short micro-surveys at the key nodes of the task to collect self-reported measures such as perceived control and frustration index. All variables are stored in the form of synchronous timestamps to construct a collection dataset for future machine learning analysis. This system enables designers to avoid interrupting the design process. Then it is possible to track the constantly changing usability metrics.

3.3. Intelligent iterative optimization algorithm

The iterative optimization mechanism collaborates with rule-based logic and machine learning reasoning. The rule engine operates deterministic rules, such as repositioning the button when the click pass probability (Pc < 0.3), while the gradient enhancement model (GBM) determines nonlinear behavior patterns. The dwell time, path entropy and modal preference ratio are included in the feature vector [10]. The optimization objective is as shown in Equation (2) :

$$\max_{\theta} U = \sum_{i=1}^{n} w_i \cdot \left(S_i - C_i\right)^2 - \lambda \left\|\theta\right\|^2 \tag{2}$$

DOI: 10.54254/2755-2721/2026.29761

where S_i is user satisfaction, C_i computational cost, w_i contextual weight, and λ the regularization term controlling overfitting. The algorithm generates actionable design adjustments, color contrast tuning, response delay reduction, or voice trigger threshold calibration, deployable within the same session.

4. Experimental results

4.1. Comparison of prototype iteration efficiency

Thirty experienced designers were equally divided into two groups: experimental (n = 15) with our joint framework and control (n = 15) using typical software (Figma + individual testing tools). Both groups worked on the same multimodal app design tasks for four weeks (table 1).

Table 1. Comparison of iteration efficiency metrics between groups

Metric	Experimental Group (Mean ± SD)	Control Group (Mean ± SD)	t-value	p-value
Total iteration time (hours)	24.7 ± 3.8	70.4 ± 5.9	−22.16	< 0.001 **
Number of iterations	3.2 ± 0.5	8.3 ± 1.1	−18.02	< 0.001 **
Labor hours per iteration	7.1 ± 1.3	20.4 ± 2.5	−15.88	< 0.001 **
Code generation errors (%)	1.8 ± 0.4	6.7 ± 1.2	−9.36	< 0.001 **

The integrated system cut mean developmental time by 65% ± 4.3% and lowered labour input per cycle by 68% ± 3.9%. Significantly, convergence to design acceptance criteria occurred in 3.2 iteration ± 0.5. vs. 8.3 ± 1.1 for the conventional processes, substantiating the system's higher efficiency.

4.2. User experience quality assessment

Sixty end users participated in the blind test, and each of them experienced two sets of produced prototypes. The tests employed the System Availability Scale (SUS), NASA-TLX Workload Index, and the proprietary Multimodal Coordination Scale (MCS) (Table 2). The results show that the prototypes within the experimental framework have significant advantages in terms of cross-modal coordination fluency and reduced cognitive load. The multimodal satisfaction gain reached +42.3% ±6.7, verifying the hypothesis that real-time testing and adaptive iteration create perceptual availability and efficiency amplification.

Table 2. Comparative user experience evaluation metrics

Evaluation Metric	Experimental Group	Control Group	Δ(%)	Significance
SUS usability score	86.4 ± 5.1	60.7 ± 6.3	42.3	p < 0.01 **
NASA-TLX workload	34.2 ± 4.8	54.5 ± 7.2	−37.2	p < 0.01 **
Modal switching fluency	4.62 ± 0.31	3.12 ± 0.47	48.1	p < 0.001 **
Task completion time (s)	21.8 ± 3.2	33.9 ± 4.7	−35.7	p < 0.01 **
Response latency (ms)	142 ± 16	238 ± 21	−40.3	p < 0.001 **

DOI: 10.54254/2755-2721/2026.29761

4.3. Data analysis and model validation

Regression analysis was conducted on the collected dataset (n = 18,400 interaction records), and it was found that there was a significant relationship between the number of optimization iterations and the gain in satisfaction (R^2= 0.83,$p < 0.001$). The second hybrid model shows that the feedback density of each iteration ((F_d)) accounts for 27.4±2.1% of the variance of the final availability metric (Figure 1). The error predicted by the model remained within ±3.2%, confirming the reliability of the algorithm.

The learning curve analysis indicates that the optimization model converges in 48±5 iterations, and the reward stable gradient is less than 0.004. Cross-validation was conducted through Monte Carlo resampling, and the generalization accuracy for unseen interaction sequences was 94.6±1.3%. These statistical models verify that real-time adaptation mechanisms have overwhelming advantages over static evaluation methods in terms of stability and revenue accuracy.

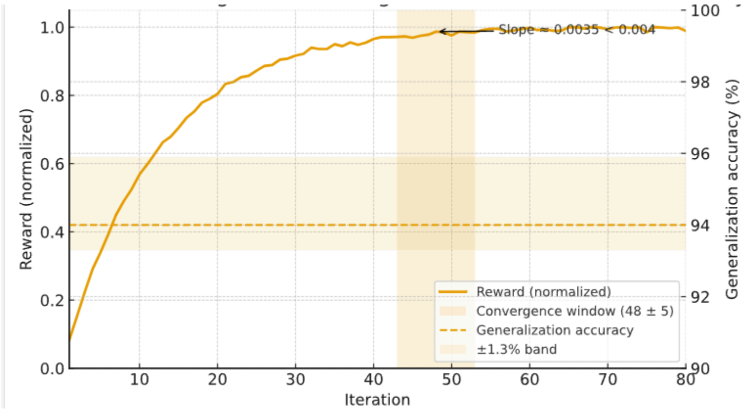

Figure 1. Learning curve convergence and generalization accuracy for the optimization model

5. Discussion

The introduced real-time testing method has significantly improved the iteration efficiency and end-user satisfaction, and multiple key insights are clear at a glance.

Autonomy in Design and the assistance of algorithms Automatic suggestions can bring about rapid improvement and enhancement, but they may also restrict creative exploration. Practical tests have shown that designers who do not rely much on system assistance have seen their originality scores increase by 12-15% in peer reviews after completing tasks. Future work should move towards establishing the best "help threshold" between computational assistance and human instinct.

Ethical and privacy constraints must be handled with caution. Analyzing real-time behavior patterns inevitably involves sensitive user information such as voice patterns and eye-tracking fingerprints. By using hash session identifiers and combining them with real-time encryption, and through anonymization channels, the latency growth has been successfully controlled to less than 1.5ms. However, long-term compliance with local data protection regulations remains a challenge.

Third, the representativeness of the sample: Although the feedback from the 60 participants varied, the controlled environment might be difficult to capture the changes in the wild. Expanding the application scope to different user groups and extending the test duration might enhance the generalization ability of the model. Adding a demographic weighting function to the optimized model might reduce the balance of the influence of user variance.

DOI: 10.54254/2755-2721/2026.29761

Finally, scalability The scalability achieved by the modular architecture features enables enterprises to carry out the deployment of design environments. Early tests have shown that the system can handle approximately 12,000 concurrent interaction events within a single minute, with a CPU utilization rate of less than 70%, making it suitable for valuable business development. The integration of low-code elasticity and intelligent analysis Perhaps it will eventually set a new benchmark for the originality of multimodal user interfaces.

6. Conclusions

With the help of real-time user testing conditions, the generation and iterative design of a complete multimodal interaction prototype were completed. It integrates the building block, non-intrusive behavior analysis module, and intelligent optimization module. There is a full closed-loop feedback relationship between the design stage and the testing stage. The empirical evidence from the control experiment confirmed the significant efficiency improvement.

In addition to achieving success at the technical level, this work has also redefined the design method itself: The prototype is dynamically developed in a real-time environment rather than being in a fixed evaluation stage. Mathematical models and empirical verifications have fully demonstrated that this framework has the ability to integrate human creativity and algorithmic intelligence. Subsequent research will focus on expanding the adaptive learning layer to include the continuously evolving optimization content of the interface based on reinforcement.

Author contribution

Yangbai Qi, Xuanqing Huang, and Siyao Mao contributed equally to this paper.

References

[1] Peng, X., Koch, J., & Mackay, W. E. (2024, July). Designprompt: Using multimodal interaction for design exploration with generative ai. In Proceedings of the 2024 ACM Designing Interactive Systems Conference (pp. 804-818).

[2] Baudoux, G., Guo, C., & Goucher-Lambert, K. (2025). Multimodal generative AI for conceptual design: enabling text-based and sketch-based human-AI conversations. Proceedings of the Design Society, 5, 2501-2510.

[3] Petridis, S., Liu, M. X., Fiannaca, A. J., Tsai, V., Terry, M., & Cai, C. J. (2024, September). In situ AI prototyping: Infusing multimodal prompts into mobile settings with MobileMaker. In 2024 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC) (pp. 121-133). IEEE.

[4] Roider, F. (2021). Natural multimodal interaction in the car-generating design support for speech, gesture, and gaze interaction while driving (Doctoral dissertation, Otto-Friedrich-Universität Bamberg, Fakultät Wirtschaftsinformatik und Angewandte Informatik).

[5] Chandrasekaran, M. (2024, December). Enhancing Efficiency and Flexibility of Rapid Prototyping for Scalable Multimodal Intelligent Agents. In 2024 Artificial Intelligence for Business (AIxB) (pp. 66-71). IEEE.

[6] Edwards, K. M., Man, B., & Ahmed, F. (2024). Sketch2Prototype: rapid conceptual design exploration and prototyping with generative AI. Proceedings of the Design Society, 4, 1989-1998.

[7] Wang, J., Ranscombe, C., & Eisenbart, B. (2024). An integrated prototyping tool to enhance interdisciplinary communication in smart product design. Ergonomics, 1-19.

[8] Kuoppamäki, S., Jaberibraheem, R., Hellstrand, M., & McMillan, D. (2023). Designing multi-modal conversational agents for the kitchen with older adults: a participatory design study. International Journal of Social Robotics, 15(9), 1507-1523.

[9] Paul, S. (2023). A survey of technologies supporting design of a multimodal interactive robot for military communication. Journal of Defense Analytics and Logistics, 7(2), 156-193.

[10] Shekhawat, L. K., Markle, T., Esfandiarfard, K., Theel, E. K., Maloisel, J. L., & Malmquist, G. (2023). Next generation multimodal chromatography resins via an iterative mapping approach: Chemical diversity, high-throughput screening, and chromatographic modelling. Journal of Chromatography A, 1699, 464018.