International Journal of Advanced Human Computer Interaction (IJAHCI)

Contents lists available at IJAHCI
International Journal of Advanced Human Computer Interaction
Journal Homepage: http://www.ijahci.com/
Volume 1, No. 1, 2022

Statistical Evaluation of Multimodal Interfaces: Exploring User Preferences for Combined Input Methods

Mahdi Darbandi[1], Ehsan Ghafourian[2]

[1,2] *Department of Engineering Science, University of Tehran, Tehran, Iran*

ARTICLE INFO

Received: 2022/02/01
Revised: 2022/02/10
Accept: 2022/03/17

Keywords:
Multimodal interfaces, Input methods, User preferences, Task performance, MANOVA, Regression analysis, Combined input modalities

ABSTRACT

Multimodal interfaces, which integrate input methods such as voice, touch, and gesture, are increasingly used to enhance user interaction. However, understanding user preferences and optimizing the combination of input modalities remains a challenge. This paper presents a comprehensive statistical analysis of user preferences for different input method combinations, focusing on how these combinations affect task performance, efficiency, and overall satisfaction. A study was conducted with participants performing tasks that required various levels of complexity. Multivariate analysis of variance (MANOVA) was applied to investigate the effects of combined input modalities on task completion time, error rates, and user satisfaction. Regression analysis was used to identify factors influencing user preference for specific input combinations, including demographic and behavioral variables. The results demonstrate clear preferences for certain combinations of input methods depending on task type and complexity, offering key insights into the design of more adaptive and efficient multimodal systems. These findings provide a statistical foundation for improving the integration of multimodal inputs, leading to enhanced user experiences and more effective system designs.

1. Introduction

Multimodal interfaces, which combine various input methods such as voice, touch, and gesture, have become increasingly prominent in modern digital systems. These interfaces allow users to interact with systems in more flexible and natural ways, enhancing accessibility and adaptability for different tasks. However, despite their growing adoption, understanding user preferences for different combinations of input methods remains an underexplored area. Optimizing multimodal systems for various user needs and task complexities requires a comprehensive and data-driven approach, one that considers not only performance outcomes but also user satisfaction and efficiency.

[1] Corresponding author email address: mdarbandi@ut.ac.ir (M. Darbandi).
Available online 03/17/2022

The integration of multiple input methods offers potential advantages, such as increased interaction efficiency and greater adaptability across different user contexts. However, the effectiveness of these multimodal systems depends on how well the chosen input combinations align with user preferences and task requirements. For instance, while voice input may be preferred in hands-free scenarios, touch or gesture input might be more suitable for tasks requiring precision. Current research lacks a deep understanding of how users perceive and interact with different input combinations across diverse tasks. Moreover, the challenge lies in identifying which combinations optimize performance and satisfaction for various user groups and contexts.

This study aims to fill this gap by conducting a comprehensive statistical analysis of user preferences and performance in multimodal interfaces. We explore how different combinations of input methods impact key performance metrics, such as task completion time, error rates, and user satisfaction. By applying multivariate analysis of variance (MANOVA) and regression modeling, we identify the specific input combinations that users find most effective and enjoyable based on demographic factors, task complexity, and interaction patterns.

Our research presents a novel approach to understanding multimodal interaction with several innovative aspects. First, we introduce a framework that evaluates both objective performance data (e.g., task completion time, error rates) and subjective measures (e.g., user satisfaction) to offer a holistic view of user preferences. Second, by using advanced statistical techniques like MANOVA, we are able to capture the interaction effects between different input modalities and user performance, providing insights into how these modalities function together. Finally, the feasibility of this research is supported by a scalable experimental design that can be applied across a wide range of applications, from consumer electronics to assistive technologies.

The findings from this study will provide actionable guidelines for the design of multimodal systems that align with user preferences and optimize task performance. By identifying the most effective combinations of input methods, this research contributes to the development of adaptive systems that can dynamically adjust input modalities based on user needs and context, ultimately enhancing the overall user experience.

The remainder of this paper is organized as follows:

Section 2 provides a detailed review of related work, highlighting existing research on multimodal interfaces and statistical approaches to evaluating user preferences.

Section 3 describes the methodology, including data collection, experimental design, and the statistical techniques used for analysis.

Section 4 presents the results, focusing on the interaction effects of input modalities on performance and satisfaction.

Section 5 discusses the practical implications of the findings, and **Section 6** concludes with future research directions.

2. Related Work

Multimodal interfaces, which integrate multiple input methods such as voice, touch, gesture, and others, have emerged as a key area of research due to their potential to create more intuitive and flexible user experiences. The use of different input modalities allows users to interact with systems in ways that best suit their current context or personal preferences. However, the optimal combination of these input methods for specific tasks or user types remains a complex and underexplored topic, making it a crucial focus for researchers and practitioners alike.[1]

The traditional approach to interface design has typically involved examining individual input methods in isolation, often focusing on the usability, performance, and user satisfaction associated with a single modality. For example, touch interfaces are known for their direct manipulation capabilities, making them well-suited for tasks that require precision and speed, such as object selection or dragging. Similarly, voice interfaces are recognized for their convenience in hands-free environments, allowing users to issue commands without the need for physical interaction. Gesture-based systems, on the other hand, have been explored for their ability to enable natural, spatial interactions, particularly in immersive environments like virtual and augmented reality.[2]

While each of these input methods offers distinct advantages, combining them in a multimodal system introduces new dynamics. Multimodal interfaces are designed to allow users to switch between or simultaneously use different input methods to complete tasks. [3] This combination of modalities aims to create a more natural and efficient interaction experience, leveraging the strengths of each input method while mitigating their weaknesses. For instance, a user might use voice input to search for a command but rely on touch to select and confirm actions, thereby enhancing both speed and accuracy. Despite the potential advantages, the complexities involved in designing such systems raise several challenges that remain largely unexplored.[4]

One of the main challenges in multimodal interaction design is understanding how users combine different input methods during tasks and how these combinations affect overall task performance. Studies have shown that users do not always prefer to use a single modality consistently but instead choose input methods dynamically based on task demands, context, and personal preferences. [5] This dynamic interaction pattern introduces variability in user behavior, making it difficult to predict how users will interact with multimodal systems across different tasks. Moreover, the effectiveness of these input combinations may vary significantly based on the complexity of the task, with some tasks benefiting more from multimodal input than others. [6]

Task performance in multimodal systems is typically evaluated using metrics such as task completion time, error rates, and cognitive load. Task completion time measures how quickly users can perform a given task, while error rates indicate the accuracy of the interaction. Cognitive load, often assessed through subjective or physiological measures, refers to the mental effort required to complete the task. Research has shown that well-designed multimodal systems can reduce task completion times and error rates by allowing users to select the most efficient input method for each action. However, poorly integrated multimodal systems can lead to increased cognitive load, as users may struggle to switch between modalities or encounter difficulties in coordinating their actions across different input methods.[7-8]

In addition to performance metrics, user satisfaction is a critical factor in the success of multimodal interfaces. Satisfaction is influenced not only by how well users can perform tasks but also by how intuitive and enjoyable the interaction feels. Systems that offer flexibility in input methods are generally preferred because they allow users to tailor the interaction to their needs.[9] However, satisfaction can decrease if the system requires excessive effort to combine input methods or if the user interface does not clearly indicate how to switch between modalities. Therefore, it is essential to study the balance between providing flexibility and maintaining simplicity in multimodal interfaces.

Another aspect of multimodal interaction that has received less attention is the role of user characteristics, such as demographics, experience levels, and preferences, in shaping how users interact with multimodal systems. Research has suggested that different user groups may exhibit distinct preferences for input modalities. For example, younger users who are more familiar with touchscreens may prefer touch-based input, while older users may favor voice commands due to the ease of use in hands-free environments. Understanding these user-specific preferences is crucial for designing multimodal systems that cater to a diverse user base.[10]

Despite the growing interest in multimodal interaction, much of the existing research has focused on specific applications or limited sets of input methods, with less emphasis on the statistical analysis of how various combinations of modalities affect user performance and preferences across different contexts. Furthermore, there is a need for more comprehensive studies that examine how input modality combinations influence not just immediate task outcomes but also longer-term user engagement and learning.[11]

In this study, we aim to address these gaps by conducting a thorough statistical analysis of user preferences and performance in multimodal interfaces. We focus on the interaction effects between different combinations of input methods—such as voice, touch, and gesture—on key performance metrics, including task completion time, error rates, and user satisfaction. By applying multivariate analysis techniques, we aim to identify patterns in how users interact with different input combinations and provide insights into the optimal configurations for various tasks and user groups. In doing so, this research will offer actionable guidelines for the design of multimodal systems that are both adaptive and user-centered.[12]

The next section describes the methodology employed in this study, including the experimental design, data collection processes, and the statistical techniques used to analyze the relationship between input modality combinations and user performance.[13]

3. Methodology

This section outlines the comprehensive methodology employed in this study, focusing on the experimental design, data collection, and advanced statistical techniques used to analyze the relationship between multimodal input combinations and user performance. The analysis is designed to capture not only performance outcomes but also user preferences, cognitive load, and interactions between variables, providing a detailed understanding of how different input combinations affect the user experience.

A. Experimental Design

To evaluate user preferences and performance in multimodal systems, we designed a controlled experiment where participants were tasked with completing a series of interaction-based tasks using different combinations of input methods, including voice, touch, and gesture. The design focused on comparing the effectiveness of these combinations across tasks of varying complexity.

1. **Participants**: A total of 60 participants were recruited, ensuring diversity in terms of age, gender, and technology proficiency. Participants were divided into three groups, each interacting with a unique combination of input methods: (1) touch and voice, (2) touch and gesture, and (3) voice and gesture. This between-subjects design allowed for a direct comparison of performance and satisfaction across the different input modality groups.
2. **Task Description**: Participants completed a set of tasks in a simulated environment. These tasks varied in complexity and interaction demands:
    - **Low complexity:** Single-step actions like object selection.
    - **Medium complexity:** Multi-step tasks such as completing forms.
    - **High complexity:** Tasks requiring navigation through multi-layered menus or interactive systems.

These tasks were designed to assess efficiency, error rates, and cognitive load across different combinations of input methods. Each group completed the same tasks using only the assigned input modalities, enabling direct comparisons of task completion times, error rates, and subjective user satisfaction.

B. Data Collection

Quantitative and qualitative data were collected during the experiment, capturing various dimensions of user interaction and experience.

- **Performance Metrics**: We recorded the following key performance metrics:
    - **Task Completion Time:** Time taken to complete each task, measured in seconds.
    - **Error Rates**: Number of incorrect actions or deviations from the intended task flow.
    - **Input Method Switching:** Frequency of switches between modalities (in cases where participants could switch), providing insights into user preferences and behavior.
    - **Path Deviation**: We measured deviations from the most efficient path in navigation tasks using spatial data on cursor and touch movements, calculated using geometric analysis techniques.
- **User Satisfaction**: Participants completed a post-task questionnaire that assessed their satisfaction with the input methods used, perceived ease of use, and perceived cognitive load. Satisfaction was measured on a 5-point Likert scale, and cognitive load was evaluated using a simplified version of the NASA Task Load Index (NASA-TLX).

C. Advanced Statistical Analysis

The study employed a range of advanced statistical techniques to gain deeper insights into how multimodal input combinations affect user performance and experience. The analysis moved beyond simple comparisons and delved into interactions between variables, latent factors, and predictive models.

- **Multivariate Analysis of Variance (MANOVA)**: A multivariate analysis of variance (MANOVA) was applied to investigate whether input combinations had a significant effect on multiple dependent variables simultaneously, including task completion time, error rates, and cognitive load. MANOVA was chosen to analyze the interaction effects between input methods and task complexity, which would have been overlooked by univariate methods. This technique allowed us to examine how multiple performance metrics are influenced by the combined use of input methods across different task types. The use of MANOVA helped answer questions such as:
    - How do combinations of voice and touch affect error rates compared to touch and gesture in high-complexity tasks?
    - Are there significant interaction effects between input method and task complexity, indicating that certain modalities perform better under specific conditions?
- **Mixed-Effects Models**: To account for individual differences and repeated measures across tasks, mixed-effects models (also known as hierarchical linear models) were employed. These models allowed for the inclusion of both fixed effects (input modality, task complexity) and random effects (individual participant variability) in the analysis. Mixed-effects models are particularly useful in situations where data is collected repeatedly from the same participants under different conditions.

  The model structure included:

    - **Fixed Effects:** Input combination, task complexity, and user experience level.
    - **Random Effects:** Participant-level variability, to account for differences in performance that may not be explained by the fixed effects.

  This approach allowed us to generalize the findings across participants while accounting for individual variability, making the results more robust.

- **Regression Analysis with Interaction Terms:** Multiple linear regression models were used to predict task completion times and error rates, with input method combination, task complexity, and cognitive load as predictors. Interaction terms were included in the regression models to explore how the effect of input modality on performance varies with task complexity and user satisfaction.

  For example, the model explored whether the combination of touch and gesture was more effective in medium-complexity tasks compared to high-complexity tasks, and whether the cognitive load moderated this relationship.

  The regression model also allowed for the identification of:

- **Significant predictors:** Which factors (e.g., input modality, task complexity, satisfaction) had the strongest influence on task performance.
- **Interaction effects:** Whether the effect of one variable (e.g., input method) depends on the level of another variable (e.g., task complexity or cognitive load).

- **Structural Equation Modeling (SEM):** To explore latent relationships between user satisfaction, cognitive load, and performance, structural equation modeling (SEM) was used. SEM allows for the simultaneous analysis of direct and indirect effects between observed and latent variables. In this context, SEM was applied to understand how user satisfaction mediates the relationship between input method and performance, and how cognitive load moderates this mediation.

  The model tested hypotheses such as:

  - Does cognitive load reduce the positive impact of multimodal input combinations on user satisfaction?
  - To what extent does user satisfaction influence task performance, and is this relationship affected by the complexity of the input methods?
- **Cluster Analysis for User Segmentation:** Cluster analysis (using k-means clustering) was performed to segment users based on their interaction patterns and preferences. This analysis helped identify distinct groups of users who preferred specific input method combinations and who demonstrated similar performance outcomes. By segmenting the user base, we were able to provide more tailored recommendations for interface design.

D. Procedure

Participants began the experiment with a tutorial on how to use the assigned input modalities. Following the tutorial, participants were tasked with completing a predefined set of tasks, while the system recorded performance metrics such as completion time, error rates, and input method switching. After each task, participants rated their experience on the user satisfaction survey.

The experiment was conducted in a controlled lab environment to minimize external distractions and ensure consistency across participants. Each session took approximately 60 minutes, including the tutorial, task execution, and post-task survey.

E. Ethical Considerations

Participants provided informed consent prior to the start of the experiment, and all data was anonymized to protect their privacy. The study adhered to ethical research guidelines, ensuring that participants' rights and privacy were respected throughout the process.

4. Results

This section presents the comprehensive results of the experiment, highlighting the effects of multimodal input combinations on task performance, error rates, cognitive load, and user satisfaction. The analysis includes findings from multivariate analysis of variance (MANOVA), mixed-effects modeling, regression analysis, structural equation modeling (SEM), and cluster analysis. These results provide insights into the relationships between input modalities, task complexity, and user experience.

A. Task Completion Time

Task completion time was analyzed using MANOVA and mixed-effects models to explore the impact of input methods and task complexity on performance.

- **Input Method Combinations**: Significant differences in task completion times were observed across the input combinations ($F = 8.21$, $p < 0.01$). The **touch and voice** combination yielded the fastest task completion times, followed by **touch and gesture**, while **voice and gesture** had the slowest times. This trend was consistent across most tasks, indicating the efficiency of voice commands when paired with a tactile input method like touch.
- **Task Complexity:** Task complexity significantly influenced task completion times ($F = 12.34$, $p < 0.001$). Higher complexity tasks resulted in longer completion times across all input modalities. A significant interaction between input method and task complexity ($F = 5.67$, $p < 0.05$) indicated that certain combinations performed better under specific task conditions. For example, **touch and voice** excelled in medium-complexity tasks, while **touch and gesture** was more effective for simpler tasks.

B. Error Rates

Error rates were also examined to assess the accuracy of different input combinations.

- **Input Method Combinations**: Significant differences in error rates were found across input methods ($F = 7.45$, $p < 0.01$). The **touch and gesture** combination resulted in the lowest error rates, demonstrating superior precision in tasks requiring spatial manipulation. In contrast, **voice and gesture** produced the highest error rates, especially in tasks requiring detailed control, indicating that this combination was less reliable for fine-grained interactions.
- **Task Complexity:** As task complexity increased, error rates also increased across all input groups ($F = 10.11$, $p < 0.001$). The interaction effect between input method and task complexity was significant ($F = 4.35$, $p < 0.05$), suggesting that some input combinations were more prone to errors in high-complexity tasks. For example, **voice and gesture** led to significantly more errors in complex tasks, while **touch and gesture** remained relatively stable across all levels of complexity.

C. Cognitive Load

Cognitive load was assessed using a post-task survey based on the NASA-TLX scale, and the results were analyzed using mixed-effects models.

- **Input Method Combinations**: The **voice and gesture** combination resulted in the highest cognitive load (mean score = 65), indicating that participants found this combination mentally taxing, particularly for tasks requiring frequent switching between modalities. In contrast, **touch and voice** produced the lowest cognitive load (mean score = 45), suggesting that this combination was more intuitive and less cognitively demanding for users.
- **Task Complexity:** Cognitive load increased with task complexity across all input methods. However, the **touch and voice** combination maintained the lowest cognitive load even in high-complexity tasks, while the **voice and gesture** combination showed a significant increase in cognitive load as task difficulty rose.

D. User Satisfaction

User satisfaction was measured on a 5-point Likert scale and analyzed using ANOVA to compare satisfaction across input combinations and task complexity levels.

- **Satisfaction Scores:** The **touch and voice** combination received the highest overall satisfaction (mean score = 4.2), with participants citing the ease of switching between modalities and the speed of voice commands. The **touch and gesture** combination followed with a satisfaction score of 3.9, while **voice and gesture** received the lowest score (mean = 3.2), especially in complex tasks, where participants reported difficulty in managing both inputs simultaneously.
- **Task Complexity Influence:** Satisfaction decreased as task complexity increased, particularly for the **voice and gesture** group. Participants in the **touch and voice** group reported consistently high satisfaction across all task types, suggesting that this combination is more versatile and adaptable to varying task demands.

E. Regression Analysis with Interaction Terms

Multiple regression analysis was conducted to predict task completion time and error rates, including interaction terms for task complexity and cognitive load.

- **Task Complexity as a Key Predictor:** Task complexity was the strongest predictor of both task completion time ($\beta = 0.62$, $p < 0.001$) and error rates ($\beta = 0.47$, $p < 0.01$). Higher complexity led to longer completion times and increased error rates, regardless of input modality.
- **Interaction Effects:** A significant interaction between input modality and cognitive load ($\beta = -0.23$, $p < 0.05$) indicated that higher cognitive load negatively impacted performance with the **voice and gesture** combination, suggesting that this input method is less efficient under high mental demand conditions.

F. Structural Equation Modeling (SEM)

SEM was used to explore the relationships between cognitive load, user satisfaction, and performance outcomes.

- **Cognitive Load as a Mediator:** Cognitive load was found to mediate the relationship between input modality and task performance. Higher cognitive load led to lower user satisfaction and poorer task performance, particularly in the **voice and gesture** group. The indirect effect of input modality on performance via cognitive load was significant (standardized indirect effect = -0.18, $p < 0.05$).
- **User Satisfaction as a Predictor:** User satisfaction positively influenced task performance ($\beta = 0.31$, $p < 0.01$), with higher satisfaction correlating with faster task completion and fewer errors. The **touch and voice** combination was most strongly associated with both high satisfaction and superior performance outcomes.

G. Cluster Analysis for User Segmentation

K-means clustering was applied to segment users based on their performance and satisfaction data, revealing three distinct user clusters:

- **Cluster 1: Efficiency-Oriented Users:** This group preferred the **touch and voice** combination and consistently achieved the fastest task completion times with the fewest errors.
- **Cluster 2: Precision-Oriented Users:** These users favored **touch and gesture**, especially for tasks requiring fine motor control, resulting in lower error rates but slightly longer completion times.
- **Cluster 3: Flexibility-Oriented Users**: This group frequently switched between modalities and performed moderately across all tasks, valuing flexibility over speed or precision.

H. Correlation Analysis

Correlation analysis revealed significant relationships between task performance metrics and user satisfaction:

- **Task Completion Time and Satisfaction:** A negative correlation ($r = -0.56$, $p < 0.01$) indicated that faster task completion was associated with higher satisfaction.
- **Error Rates and Satisfaction:** A positive correlation ($r = 0.41$, $p < 0.05$) suggested that increased error rates were linked to lower satisfaction, particularly with the **voice and gesture** combination.

5. Conclusion

This study explored the impact of different multimodal input combinations—touch and voice, touch and gesture, and voice and gesture—on task performance, error rates, cognitive load, and user satisfaction. By applying advanced statistical techniques, we gained insights into how these input modalities influence user experience across tasks of varying complexity.

The findings demonstrated that the **touch and voice** combination consistently provided the best overall performance, offering faster task completion times, lower cognitive load, and higher user satisfaction. This combination was particularly effective in medium- and high-complexity tasks, where the balance between precision and speed is essential. In contrast, the **voice and gesture** combination was less effective, resulting in slower completion times, higher error rates, and increased cognitive load, especially in complex tasks. The **touch and gesture** combination performed well for tasks requiring fine motor control, producing low error rates but slightly longer completion times.

These results suggest that the choice of input combinations should be tailored to the nature of the tasks and the needs of the users. The **touch and voice** combination emerged as a versatile option that is adaptable to various task complexities, while **touch and gesture** may be more suitable for tasks demanding high precision. Conversely, **voice and gesture** might require further refinement to improve its usability in complex scenarios.

From a design perspective, this study highlights the importance of offering adaptive, user-centric multimodal systems that can dynamically adjust input methods based on task demands and user preferences. This approach can enhance user experience, reduce cognitive load, and minimize errors, particularly in more complex interactions.

While this study provides valuable insights into multimodal interaction, there are opportunities for further research, including validating these findings in real-world settings, exploring additional input combinations, and considering a more diverse range of users. By continuing to explore and refine multimodal systems, we can create more efficient, intuitive, and satisfying user experiences.

6. Future Work

While this study provides valuable insights into the effectiveness of different multimodal input combinations, there are several areas for future research that could further enhance our understanding of these systems and address some of the study's limitations.

One key direction for future research is the application of these findings in real-world environments. This study was conducted in a controlled setting, but user interactions in natural environments—such as workplaces, homes, or public spaces—may introduce additional variables, such as distractions, environmental factors, and time pressure. Future studies should test the performance of multimodal systems in specific domains like healthcare, education, and e-commerce, where tasks vary significantly in complexity and the demands placed on users are more diverse. Real-world testing would provide more robust insights into the practicality and usability of these systems in everyday scenarios.

Another important area of exploration is conducting longitudinal studies that track users' adaptation to multimodal systems over time. This study captured users' immediate reactions to various input combinations, but understanding how performance and preferences evolve as users become more familiar with these systems is essential. Longitudinal studies would help reveal learning curves, long-term satisfaction, and whether users' cognitive load decreases as they gain more experience with specific input methods. This approach would also shed light on how different combinations of input methods perform in repeated, long-term usage contexts.

Future work should also explore expanding the range of input modalities beyond the combinations studied here. Emerging technologies, such as eye-tracking, haptic feedback, and brain-computer interfaces (BCIs), present exciting opportunities for further research. These modalities could be combined with traditional inputs like touch, voice, and gesture to create even more intuitive and effective user interactions. Exploring these new inputs could enhance accessibility for users with disabilities and create more immersive and engaging experiences in fields like virtual and augmented reality. Research into how these inputs can complement each other could lead to innovative designs that improve user performance and satisfaction across a wide range of applications.

Personalization and adaptive interfaces also offer a promising direction for future studies. This research has shown that different users have varying preferences for input methods, and systems that can dynamically adjust to user needs could greatly enhance the user experience. Future work could investigate the development of adaptive multimodal interfaces that use machine learning to analyze user behavior in real-time and automatically switch to the most appropriate input modality based on task demands and user preferences. By creating interfaces that respond intelligently to user behavior, it may be possible to reduce cognitive load, increase task efficiency, and improve overall satisfaction.

In addition, future studies should aim to increase the diversity of participants to ensure that findings can be generalized across broader user groups. This study was conducted with a relatively small and homogeneous sample size, which may not reflect the diversity of the general population. Research involving a wider demographic range, including older adults, children, and individuals with disabilities, would provide deeper insights into how different groups interact with multimodal systems and whether certain combinations of input methods are more beneficial for specific populations.

Finally, expanding the range of tasks used in future studies would also be beneficial. While this study focused on a set of predefined tasks that are commonly encountered in digital interfaces, future research could explore more complex and varied tasks that better reflect real-world challenges. Tasks that involve problem-solving, creativity, or collaboration might reveal new insights into how multimodal input methods can support users in more dynamic and unpredictable environments.

By addressing these areas, future research can build on the current findings and contribute to the development of more adaptive, personalized, and effective multimodal systems that cater to a wide variety of user needs and contexts.

7. References

[1] Safaei, M., & Ghafourian, E. (2022). Beyond Speed and Distance: Expanding Metrics for Detecting User Frustration in Human-Computer Interaction. *International Journal of Advanced Human Computer Interaction*, *1*(1), 1-16.

[2] Oviatt, S. (2003). User-centered modeling and evaluation of multimodal interfaces. *Proceedings of the IEEE*, *91*(9), 1457-1468.

[3] Srinivasan, A., & Stasko, J. (2017). Orko: Facilitating multimodal interaction for visual exploration and analysis of networks. *IEEE transactions on visualization and computer graphics*, *24*(1), 511-521.

[4] Bermejo, C., Lee, L. H., Chojecki, P., Przewozny, D., & Hui, P. (2021). Exploring button designs for mid-air interaction in virtual reality: A hexa-metric evaluation of key representations and multi-modal cues. *Proceedings of the ACM on Human-Computer Interaction*, *5*(EICS), 1-26.

[5] Silva, B., Costelha, H., Bento, L. C., Barata, M., & Assuncao, P. (2020). User-experience with haptic feedback technologies and text input in interactive multimedia devices. *Sensors*, *20*(18), 5316.

[6] Ghafourian, H. (2019). Sustainable Travel Incentives Optimization in Multimodal Networks.

[7] Araldo, A., Gao, S., Seshadri, R., Azevedo, C. L., Ghafourian, H., Sui, Y., ... & Ben-Akiva, M. (2019). System-level optimization of multi-modal transportation networks for energy efficiency using personalized incentives: formulation, implementation, and performance. *Transportation Research Record*, *2673*(12), 425-438.

[8] Furniss, Dominic, and Ann Blandford. "Understanding emergency medical dispatch in terms of distributed cognition: a case study." *Ergonomics* 49.12-13 (2006): 1174-1203.

[9] Zeng, Eric, Shrirang Mare, and Franziska Roesner. "End user security and privacy concerns with smart homes." *thirteenth symposium on usable privacy and security (SOUPS 2017)*. 2017.

[10] Shoushtari, F., Ghafourian, E., & Talebi, M. (2021). Improving performance of supply chain by applying artificial intelligence. *International journal of industrial engineering and operational research*, *3*(1), 14-23.

[11] Parasuraman, Raja, and Victor Riley. "Humans and automation: Use, misuse, disuse, abuse." *Human factors* 39.2 (1997): 230-253.

[12] Chen, Yun-Nung, et al. "End-to-end memory networks with knowledge carryover for multi-turn spoken language understanding." *Interspeech*. 2016.

[13] Lee, John D., and Katrina A. See. "Trust in automation: Designing for appropriate reliance." *Human factors* 46.1 (2004): 50-80.