Natural Human-Robot Interaction Paradigm for a New Generation of Intelligent Robots

Xinyang Wang

{xinyangwang@arizona.edu}

Computer Science, University of Arizona, The United States

Abstract. The purpose of the human-computer interaction (HCI) technology is to introduce a smoother and more or less natural interaction experience between machines and human beings. This paper expounds on the fundamental concepts, the current research state as well as the future research directions of HCI. The article provides us with the introduction to the fundamental concepts and process of development of HCI. Also, it introduces various areas of modern HCI technology, such as new forms of interaction, brain-computer interface (BCI), multimodal technology, and emotion recognition. HCI technology despite the results realized so far in various fields continues to encounter a series of challenges. Thus, this article indicates the barriers and constraints experienced in human-computer interaction technology, including the issue of stability of BCIs, the confines of immersive technology interaction, and the technical problem of multimodal information fusion.

Keywords: Human-computer interaction, Multimodal technology, Emotion recognition

1 Introduction

Today with the rapid development of computer technology and artificial intelligence, humans' lives are inseparable from computer technology in all aspects. Under this trend, they will continue to develop at a high speed. These developments are not only reflected in our mobile phones, computers, and various electronic devices, but also in various fields such as healthcare, transportation, education, finance, and industrial manufacturing. Robotics is an important interdisciplinary field that integrates computer technology, mechanical engineering, and artificial intelligence. The development of Robotic mainly includes three stages: 1) Reproduction robot, the initial stage robot, which totally relied on the human instructions to operate and cannot adapt the environment and make changes accordingly. 2) Visual robots, through sensor technology and computing capabilities, could perceive the environment and the ability to adjust and operate based on environmental information. 3) Intelligent robots, due to the development in fields such as computing technology, machine learning, artificial intelligence, and perception technology, these robots have begun to possess stronger autonomous decision-making capabilities. Robots no longer merely perform fixed tasks; instead, they can interact with the environment and learn new tasks. The development of

ICIAAI 2026, May 22-24, Singapore, Singapore

DOI 10.4108/eai.22-5-2026.2365335

computer technology and robotics cannot be separated from the discipline of Human-Computer Interaction (HCI). Human-Computer interaction is a discipline that studies the interactions between human and machines. The development of this field has become a necessary requirement of the current era. Since the 21st century, with the rapid development of information technology, human-computer interaction technology has become an important direction for the future development of the artificial intelligence industry. There are three stages of the development of human-computer interaction technology: 1) Command-line interaction stage. This stage is not user-friendly. Users often need to remember many input rules for the programming languages to interact with machine. 2) Graphical interface interaction stage. After the user learns the interaction content from the graphical interface, they can interact with the machine through mouse clicks or input devices such as keyboards. This stage makes human-computer interaction more intuitive and convenient. 3) Natural language interaction. Users do not need to learn specific machine operation instructions. The machine can understand the natural language input by the users and provide feedback after logical reasoning. At this stage, both user experience and efficiency are better.

Tracing the origin of a field is not an easy task, and this is true for the discipline of human-computer interaction as well. Tracing back the history of human-computer interaction, the concept of human-computer interaction originated in 1946 with the electronic general-purpose computer, namely the electronic computer called Electronic Numerical Integrator and Calculator (ENIAC), which was approximately the size of a small town. Many concepts related to computers were not invented by a single expert or in a single paper but developed gradually over time. Windows95 is a good example of HCI. The Macintosh served as the foundation for the widely used Windows95 graphical interface. The Macintosh originated from the Xerox PARC research center. And the Macintosh was based on the early research conducted by the Stanford Research Laboratory and the Massachusetts Institute of Technology (MIT).

The early artificial interaction techniques were achieved through template matching. This is quite similar to the interaction techniques mentioned earlier in the command-line stage. The information input by the user and the feedback from the machine were relatively inflexible. A representative early interactive system was ELIZA, invented by MIT. With the development of graphical user interface (GUI), human-computer interaction has become more convenient and flexible. Apple Macintosh and Windows 95 are representative examples of human-computer interaction systems in the graphic user interface stage. Users no longer can only input rigid computer instructions; they can also interact with the computer by clicking on icons and windows. After this stage, with the development of speech recognition and speech generation technologies, the progress of human-computer interaction technology was also promoted. The interaction between humans and machines can gradually be achieved through natural language [1]. Some famous representations in this stage are Apple Siri, Microsoft Cortana, Amazon Alexa [2].

The journal "Science Robotics" has listed intelligence and human-like characteristics as one of the top ten challenges in the field of human-machine interaction [3]. At the current stage, intelligent and human-like human-machine interaction technologies are also an important research direction. The real world that is multi-modal, multi-source and information-rich, with various information sources. Therefore, simply conducting human-computer interaction through natural language communication is clearly insufficient. During the usage process, there often arise problems such as inaccurate machine feedback and inconsistent information.

Against this background, Ideal human-computer interaction could achieve the following goal: the machine can not only obtain information through natural language interaction but also acquire information from other sources. In other words, the machine can also recognize visual information and even detect human emotional information, which will greatly optimize the process of human-computer interaction. Therefore, the human-computer interaction technology that integrates multi-modal and multi-source information has crucial practical significance in the field of human-computer interaction.

Although human-computer interaction has achieved remarkable development in various related fields, it still inevitably faces a series of optimization challenges and technical problems at present. Given the broadness of the human-computer interaction field, the purpose of this research is to conduct a comprehensive review, introduce the basic definition of human-computer interaction, and analyze the development status, context and existing research issues of human-computer interaction in three related fields: new interaction media, BCI, multimodal technology, and emotion recognition. Furthermore, this paper makes an outlook on the development of human-computer interaction systems. The chapter arrangement of this paper is as follows: Chapter 2 mainly introduces the definition of human-computer interaction; Chapter 3 mainly discusses the current development status of human-computer interaction in these three fields; Chapter 4 analyzes the challenges and limitations faced by these three fields; Chapter 5 makes a final summary of the content of this article.

2 Human-computer interaction

Human-Computer Interaction (HCI) technology is designed to come up with a more natural and convenient way of interaction between machines and humans and thus enhances human life and work output.

The area of human-computer interaction is rather broad. The working definition of the concept of human-computer interaction was proposed by Sinha et al: HUMAN COMPUTER INTERACTION is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them [4]. Briefly, interaction, especially, human-computer interaction is very important in the perspective of computer science. One cannot separate computers and the concept of human-computer interaction. That is, putting it in another way, the concept of human-computer interaction could not exist without computers. The whole concept of human-computer interaction is a natural occurrence as computers appeared and will continue to change as the newest computer technology advances. Human-computer interaction is vital because people are capable of well working with an advanced-size of computing device. When people cannot manage or cannot understand how to manage a machine, such a machine will not have any value. Artificial intelligence is commonly applied in intelligent robotics, in artificial intelligence or certain areas of computer science to make computers more user friendly and responsive to become more responsive to its users. This is exactly the importance of the human-computer interaction technology research. A good example is human-computer natural interaction field in computer science. Human-computer natural interaction is one novel technological trend which is a result of interdisciplinary convergence of computer science with psychology [5]. The study

is meant to bring the interaction between man and computer closer to the daily behavior of man and man as well as the man and the environment.

3 Current development status of HCI

Based on the complexity and wide scope of the concept of human-computer interaction, this article will discuss the current development status in several aspects of this field: 1) Novel interactive media. 2) BCI. 3) Multimodal and emotion recognition.

3.1 Emerging interaction media

In recent years, with the rapid development of technology, many new human-computer interaction media have emerged. Immersive technology is a representative example. Virtual reality (VR), augmented reality (AR) and mixed reality (MR) are collectively referred to as extended reality (XR), and all of them belong to immersive technology [6].

The foundation of immersive technology lies in the methods, theories and interaction approaches provided by two-dimensional displays. Its aim is to introduce new ways for users to interact with computers, that is, to develop new human-computer interaction methods. Immersive technology goes beyond the static characteristics of the original two-dimensional environment and provides a three-dimensional experience.

The various interaction methods of immersive technology are very similar to those of humans, such as language, eye contact, gestures and touch, which naturally provides many possibilities for interaction. It can be mapped to various environments, circumstances, activities and applications based on the senses and communicate means of the user or not [7]. Moreover, immersive technology offers a form of communication that is different than what can be found in normal human communication. More precisely, with a more advanced approach, that is voice, head and gesture control in work [8].

3.2 BCI

BCI is an issue that is very popular in the sphere of human-computer interaction. It can transform the interaction of human beings with computers and may be an innovation in this field.

Brain-computer interfaces are able to read brain activity and translate it into instructions that the computers and other machine devices can execute such that with the help of their thoughts, one is able to control the machines/ devices. The brain-computer interfaces are classified to include both one-way brain-computer interfaces and two-way brain-computer interfaces. They are different in terms of their functions since one-way brain-computer interfaces are only able to receive brain signals as well as send signals to the brain whilst two-way brain-computer interfaces only can facilitate two-way flows of information hence allowing control of devices by the brain [9].

Brain-computer interface operation mode involves a series of procedures as follows: Firstly, electrodes monitor the signals of the brain of a person. The microcontroller of the brain-computer interface interprets these signals. This has the ability to remove the noise produced by external influences and the device itself. Lastly, the artificial neural network has the capacity to process data and adapt on the requirements with the use of its analytical powers that are then

applied in determining the relevant instructions. The brain-computer interface utilizes several types of signal processing algorithms, such as feature extraction algorithms, classification algorithms, pattern recognition algorithms and machine learning models [9].

3.3 Multimodal technology and emotion recognition

In the physical world, multimodal data can be of many different kinds, and the world is multimodal. Modality generally used is that of the how of an object existence or perception [10]. A simple illustration that is easy to comprehend is that a human being has the ability to gather information at a variety of modalities. Images, odors, and sounds may be viewed as the three different information types by sensing objects visually, smelling, and hearing. In order to be able to make sense of our real world, it is essential to analyze and reason multimodal data with the purpose of empowering artificial intelligence. The technologies of multimodality invented to address this need offer an avenue of using multimodal data and relating with machines [11] such as multimodal information perception technology and multi-source information fusion technology.

Emotions are complicated combination of physiological and psychological situations, consisting of different thoughts, feelings, and behaviors. Simply, emotional states are the various responses that a living being presents to external influences in certain circumstances. This involves physical indicators that include tension of the muscles and increase in heart rate and sweating and psychological factors including personal experiences and emotional reactions. These sorts of reactions to emotion are the basis behind what is known as emotion recognition since it seeks to determine the level of emotion that people are portraying in this manner. These kinds of information are to be combined with each other to be recognized as emotion. The multifaceted feelings are incapable of being represented with the help of a single medium [12]. Under the conditions of artificial intelligence, the emotional expression does not only appear within interpersonal relations but also extends to the whole sphere of the human-computer interaction. Multimodal data can have complex expression of emotions. Machines can recognize human emotions faster and more accurately by passaging voice, text, and images over the internet instead, this has become a research field [13]. The application of multimodal technology in emotion analysis is very important in human-computer interaction. It is capable of enhancing the precision and efficiency of communications, allowing artificial intelligence to comprehend the expression of emotions users are interested in receiving better.

The operation process of multimodal technology generally follows this: Firstly, perform preprocessing and feature extraction for a single modality. Different modalities, such as image modality, speech modality, text modality, have different preprocessing and feature extraction methods. For example, the image normalization preprocessing method for image modality and the Gabor filter feature extraction method for texture. Then, through various fusion strategies, such as early fusion, late fusion, mixed fusion, to perform feature fusion, thereby improving the model's accuracy in judging multi-modal features.

4 Challenges and future prospects

The development of human-computer interaction is still facing limitations and challenges in various fields at present.

4.1 Emerging interaction media

Due to the increasing diversity of application environments, designers of immersive technologies are confronted with the issue of sustainability. Although XR technology has made significant progress and achieved a certain level of natural interaction, researchers have been unable to apply human interpersonal communication to all interactions in virtual content. Moreover, immersive technology integrates different levels of reality and virtuality, and the interaction paradigm highly depends on the nature of the user's interaction content. The interaction methods between real objects and virtual items will also change. These inconsistencies in XR technology bring interaction paradoxes to users and pose some design challenges for providing the best user experience.

Spittle et al. [9] investigated the acceptance and application methods of extended XR input in different fields, highlighting the current lack of a unified solution. However, the research results also offer hope: based on trends related to devices, tasks, and use cases, the most suitable interaction methods can be predicted. The future of immersive technology is promising. If the interaction paradox in some cases can be resolved to improve the interaction accuracy of users in virtual environments and combined with new multimodal technologies, it can bring users a better interaction experience.

4.2 Brain-computer interaction

Although BCI technology has achieved some research results, it is still not mature and is in the early stage of development. The first limitation faced by BCIs is their poor universality and high cost. Due to the differences in brain activity among users, neural networks and other learning algorithms are commonly applied in signal processing. Correspondingly, these systems require a long period of training for the BCI to accurately interpret the instructions of a specific user [14], and each user's BCI is unique. Additionally, BCI still faces a challenge: stability. Noise and environmental interference can significantly affect the efficiency of BCI systems. Two common types of noise are acoustic noise and electromagnetic field interference. Outside the idealized laboratory environment, acoustic noise, such as various sounds produced by people or machines near the BCI user, human conversations, etc., can interfere with the collection and processing of electroencephalogram (EEG) signals [14]. Common sources of electromagnetic field interference include radio waves emitted by mobile phones or Wi-Fi routers, which can also affect the neural activity information recorded by EEG. Therefore, adequate shielding measures are very important.

BCI still has great potential in the future. Brain-computer interface technology has been proposed for application in various fields such as medicine, neuroscience, education, human-computer interaction, and entertainment.

Neural prosthetic technology is an interesting advancement in the field of brain-computer interaction, presenting a new treatment concept for disabled individuals. Especially for the spinal cord injured and paralysis people, there will be hope for them if it is applied widely. The external equipment controlled by the disabled people will be operated by its own brain signal through the brain computer interface technology. With the development of new technology, overcoming the above technical problems and enhancing the anti-noise performance, and the brain-computer interaction technology can realize more stable application.

4.3 Multimodal technology and emotion recognition

The present multimodal technologies have a number of challenges, such as:
1) Dataset construction. It is not easy to develop good quality multimodal datasets due to privacy such as the possibility of revealing personal details.
2) Modality alignment. The matching of heterogeneous information of text, audio, and video data is not an easy task, because various modalities are not always matched by time and meaning structures.
3) Lightweight models. To achieve effective multimodal models, one should balance between performance, costs in computational and storage, particularly in limited resources contexts.
4) Contextual information. A lack of contextual information inhibits the accuracy of multimodal sentiment analysis because correct interpretation is often based on the history of communication and the context around it.
5) Modality missing. Missing modalities can be due to sensor failure, signal corruption or other factors and drastically affect the performance and robustness of the model. However, multimodal and emotion recognition issues have massive application options. It can be inferred that in the future, multimodal technology and emotion recognition will be deployed to sustain a broad range of practice applications. As an illustration, the technologies can promote the investigation of people involved in a crime through studies of their emotional and psychological state, thereby facilitating a more effective approach to interrogating a suspect in other case, these technologies can be used to study the emotional state of a customer and make the interaction with him or her more personal. With the ability to adequately deal with the current technical challenges, the future uses of multimodal emotion recognition are projected to turn to be not only more extensive but also more sophisticated.

5 Conclusion

In conclusion, this article focuses on the development area of artificial interaction, with particular attention to emerging interaction media, brain-computer interaction, multimodal technology, and emotion recognition. Firstly, the article briefly introduces the basic concepts and historical information of human-computer interaction and conducts detailed discussions on the three related fields, elaborating on the current applications and development status. Subsequently, the article focuses on presenting some limitations and current challenges in the three related fields and discusses the potential application directions and development prospects of each field in the future. In the current era of rapid technological development, human-computer interaction and the complementary computer technology, artificial intelligence has great potential for development and application value. This article aims to help readers who have not conducted in-depth research in the field of human-computer interaction to quickly understand the current development status of the field.

References

[1] Xiao, C., Fan, Y., Xu, L., et al.: Human-centered natural spatial language interaction between human and robot. Chinese Journal of Applied Psychology. pp. 319–331. (2019)

[2] Kepuska, V., Bohouta, G.: Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home). Proc. IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), 2018, pp. 99–103.
[3] Yang, G. Z., Bellingham, J., Dupont, P. E., et al.: The grand challenges of science robotics. Science Robotics. pp. eaar7650. (2018)
[4] Sinha G, Shahi R, Shankar M.: Human computer interaction[C]//2010 3rd International Conference on Emerging Trends in Engineering and Technology. Goa, India: IEEE, 2010
[5] Tian, F., Yu, C.: Preface to the special issue on new progress in natural human-computer interaction. Journal of Software. pp. 2925–2926. (2021)
[6] Mohanty, R. R., Selly, P., Brenner, L., Vyas, S., Nelson, C. R., Moats, J. B., Gabbard, J. L., Mehta, R. K.: From discovery to design and implementation: A guide on integrating immersive technologies in public safety training. IEEE Transactions on Learning Technologies. pp. 387–401. (2025)
[7] Bach, B., Sicat, R., Beyer, J., Cordeil, M., Pfister, H.: The hologram in my hand: How effective is interactive exploration of 3D visualizations in immersive tangible augmented reality? IEEE Transactions on Visualization and Computer Graphics. pp. 457–467. (2018)
[8] Bai, H., Sasikumar, P., Yang, J., Billinghurst, M.: A user study on mixed reality remote collaboration with eye gaze and hand gesture sharing. Proc. CHI Conference on Human Factors in Computing Systems, 2021, pp. 1–13.
[9] Spittle, B., Frutos-Pascual, M., Creed, C., Williams, I.: A review of interaction techniques for immersive environments. IEEE Transactions on Visualization and Computer Graphics. pp. 3900–3921. (2023)
[10] Nizam, S. S. M., Abidin, R. Z., Hashim, N. C., Lam, M. C., Arshad, H., Abd Majid, N. A.: A review of multimodal interaction technique in augmented reality environment. International Journal of Advanced Science Engineering and Information Technology. pp. 1460. (2018)
[11] Peksa, J., Mamchur, D.: State-of-the-art on brain-computer interface technology. Sensors. pp. 6001. (2023)
[12] Baltrušaitis, T., Ahuja, C., Morency, L. P.: Multimodal machine learning: A survey and taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence. pp. 423–443. (2019)
[13] Zhu, X., Guo, C., Feng, H., et al.: A review of key technologies for emotion analysis using multimodal information. Cognitive Computation. pp. 1504–1530. (2024)
[14] Abdullah, S. M. S. A., Ameen, S. Y. A., Sadeeq, M. A. M., et al.: Multimodal emotion recognition using deep learning. Journal of Applied Science and Technology Trends. pp. 52–58. (2021)