Available Online: 10 July 2025 DOI: 10.54254/2977-3903/2025.25079

Research on the application of speech recognition and applied psychology in Human-Computer Intelligent interaction

Zihan Zhang

Medical College, Inner Mongolia University of Science &Technology, Baotou, China

3421581450@qq.com

Abstract. Text and speech processing technologies encounter bottlenecks in understanding complex emotions within Human-Computer Interaction (HCI), with significant limitations particularly in the semantic parsing of intonation and tone. This paper employs a literature review approach to systematically synthesize core advancements in applied psychology and speech recognition technology within HCI, focusing on the intersection of emotion recognition technology and psychological intervention scenarios. The study indicates that the deep integration of these fields can significantly enhance intervention precision, especially in scenarios such as psychological diagnosis in children (e.g., emotion recognition in autism) and personalized Cognitive Behavioral Therapy (CBT) guidance. The research further points out that Artificial Intelligence (AI) has not yet fully grasped complex human emotions, necessitating the deeper embedding of psychological theories into speech emotion analysis frameworks to improve the ability to interpret semantically ambiguous expressions. This review provides a theoretical framework and a practical pathway for the application of AI technology and applied psychology in intelligent HCI.

Keywords: affective computing, Human-Computer Interaction (HCI), applied psychology, speech recognition

1. Introduction

In the era of big data, fast-paced lifestyles have spurred new demands for psychological adjustment. Increasingly, individuals turn to online emotional support and comfort, driving the convergence of artificial intelligence and applied psychology. In recent years, affective AI systems (e.g., intelligent chatbots, virtual companion systems) have achieved basic emotion recognition through voice and image technologies. Such systems utilize natural language processing (NLP), image processing, and other techniques to identify user emotions and provide feedback. However, their ability to interpret complex emotions remains superficial, particularly in parsing intonation and tone. Significant limitations remain when these systems process complex emotions. Their handling of intonation and tone is notably weak, failing to comprehend the deeper meanings conveyed by complex human vocal nuances. For instance, they exhibit limitations in understanding non-literal meanings such as puns, sarcasm, or internet slang embedded in speech prosody, making it difficult to interpret deep psychological states. How to leverage applied psychological theories to address the shortcomings in emotional understanding within speech recognition has become a key challenge for enhancing the authenticity of HCI. Based on this practical challenge, this paper centers on the cross-application of speech recognition and applied psychology, employing a literature review method to systematically synthesize latest advancements and theoretical debates in this field, elucidating the research status of applied psychological theories in emotion recognition. This study provides a theoretical foundation for subsequent research on speech recognition and applied psychology within HCI and offers guidance for deepening interdisciplinary collaboration.

2. Literature review

2.1. Speech recognition technology

Affective systems refer to those intelligent systems designed to exhibit emotions, recognize user emotions, or regulate users' emotional states [1]. These systems typically simulate multi-layered human verbal and non-verbal communication methods [1]. Examples include facial expressions, body posture, skin color response, grip strength, natural speech, and simulated emotions synchronized with speech prosody. The core task is the detection of user emotions. Systems achieve this through various pattern

Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

recognition tools (e.g., voice, facial recognition, speech recognition and analysis, posture, skin color), followed by pattern matching and emotion classification based on knowledge databases aligned with human emotional cognition.

2.2. Theoretical empowerment of applied psychology for speech interaction

Applied psychology supports HCI design through emotional cognition and social interaction theories. Early research by Klein proposed that computers possess the potential to simulate emotional intelligence to reduce user frustration laying groundwork for affective computing [1]. Cohen and Seider advanced multimodal emotion recognition research via facial expression [2]. Bänziger further expanded the cultural dimension of emotional processing [3].

Social interaction theory focuses on the mechanism design of anthropomorphic interaction. Bickmore's Embodied Conversational Agent (ECA). which simulates face-to-face interaction through voice, and other non-verbal means [4]. Kaul et al. classified the implementation forms of AI into diagnostic decision support systems, chatbots, AI-driven virtual avatars, and emotion analysts [5]. Chatbots and virtual avatars involve natural language processing, emotion analysis involves neural networks and deep learning, while AI decision support systems involve machine learning. Research by Terblanche et al. indicates that chatbots are as effective as human coaches in helping clients achieve goals and can replace human coaches using simple or formula-based methods, showing greater efficiency advantages especially when executing standardized processes (e.g., habit formation plans). However, significant limitations remain in complex scenarios requiring deep emotional empathy, such as trauma intervention [6].

The core bottleneck in current research lies in the insufficient computational expression of psychological theories and the semantic gap for complex emotions. Multimodal speech recognition and HCI emotional design are integrating, shifting HCI from function-oriented to emotional intelligence-driven paradigms.

3. Coreapplications of speech recognition and applied psychology

The convergence of artificial intelligence and psychology has fostered diversified psychological intervention scenarios, forming a multi-dimensional application system spanning from clinical diagnosis to personalized therapy, from educational assistance to societal-level mental health maintenance.

3.1. Application of speech emotion analysis in psychological diagnosis

In the field of mental health diagnosis, the integration of speech recognition and applied psychology is breaking through the limitations of traditional assessments. Intelligent speech-based diagnostic systems guided by clinical psychology theories provide improved services by analyzing features such as speech disfluencies and pitch fluctuations. Mumtaz and colleagues proposed establishing an early depression screening model by analyzing features like speech disfluency frequency and pitch fluctuation rate, achieving significantly higher accuracy compared to traditional scales [7]. Researchers Qiu H.D. et al. pointed out that AI's uninterrupted 24/7 service and efficient data processing capabilities give it significant advantages [8]. Simultaneously, AI reduces subjective human bias through standardized algorithms, providing more rational and patient-specific recommendations.

For children with immature, complex emotions and language skills, multimodal fusion technology (e.g., joint analysis of vocal crying features and facial expressions) can improve emotion recognition accuracy. A study used convolutional neural networks to analyze the spectral characteristics of children's cries and frown frequency, achieving precise assessment of postoperative pain levels [9]. This achievement not only fills a gap in child pain assessment tools but also demonstrates the clinical application potential of AI in recognizing non-verbal emotional expressions, providing empirical medical context evidence for subsequent affective computing models.

This technology is not only applied in medical settings but also extends to emotion recognition in children with autism. Another study built an autism screening model based on speech emotion analysis, enabling early warning of childhood autism by identifying abnormal vocalization frequencies and speech disfluency patterns [7]. AI' ability to connect to databases enhances data processing and analysis capabilities, enabling deep mining of vast mental health data. It collects comprehensive information, ranging from an individual's daily behavior patterns and language expression habits to physiological indicators like sleep quality and heart rate variability. One intelligent companion system accesses a psychological dialogue library; upon detecting negative intonation, it automatically triggers CBT guidance. Using machine learning algorithms, it accurately identifies early symptoms and potential risk factors for various mental illnesses. Based on individual patient characteristics and psychological data, it builds dynamic mental health prediction models and customizes personalized treatment plans to improve therapeutic efficacy.

3.2. Personalized psychological support through speech interaction

Affective computing, as a crucial interdisciplinary domain integrating AI and applied psychology, seeks to enable machines to recognize, interpret, and express human emotions with greater precision. For instance, by refining speech synthesis algorithms, the intonation of robots speech can dynamically adjust in respond to user' emotional fluctuations, thereby enhancing emotional resonance. On one hand, improving sensor technologies and machine learning algorithms enhances the machine's accuracy in recognizing emotional cues, such as human facial expressions, speech prosody, and gestures. On the other hand, leveraging applied psychology research on emotional expression and communication helps machines respond to human emotions in more natural and appropriate ways, achieving HCI with greater emotional resonance. Developing systems like intelligent customer service agents and companion robots that can perceive user emotions and adjust accordingly enhances user experience across various interaction scenarios, strengthening trust and emotional connection between humans and machines.

Applied psychology deeply analyzes individual psychological traits, personality types, life experiences, and psychological needs. Subsequently, AI utilizes this information to develop personalized intervention plans for each individual. For example, with cognitive behavioral therapy (CBT), taking a patient with social anxiety as an example, AI analyzes the frequency of self-negating words in their speech to generate personalized dialogue training scenarios. When the system detects anxiety-related vocabulary in the patient's speech, it automatically transitions to a virtual reality-simulated micro-social scenario, guiding the patient through exposure therapy with voice prompts. Based on speech emotion analysis outcomes, AI generates personalized multimedia content. If the patient's speech spectrum shows high anxiety features (e.g., increased high-frequency components), the system automatically matches soothing music and guides progressive muscle relaxation training via voice prompts. Combining with psychotherapist guidance, this facilitates personalized exposure therapy, improving treatment effectiveness.

3.3. Social mental health monitoring and intervention

Societal-level mental health monitoring employs the analysis of emotional keyword distribution in collective speech, integrated with micro-expression simulation technology, to enable early warning of psychological crises and intelligent allocation of intervention resources.. Leveraging AI technology enables large-scale mental health screening and early intervention, identifying potential psychological problems promptly, alleviating workforce pressures associated with shortages of professional mental health personnel, and thereby enhancing society's overall mental health status. By tracking changes in individual real-time data, the system can predict the likelihood of psychological issues occurring months or even years in advance to enable timely intervention. For instance, by analyzing users' language styles and emotional word frequency on social media, combined with browsing history and interaction behavior data, natural language processing techniques and deep learning models are used to assess users' psychological states and detect potential psychological crises early.

Applying this technology in the education sector can provide personalized learning support and psychological counseling to students in different regions, helping to bridge educational gaps and benefit more students. For example, identifying learning anxiety through speech emotion analysis (e.g., reduced pronunciation clarity, increased pauses) can lead to recommending suitable learning resources. An intelligent English learning system can analyze "confusion intonation" in student speech in real-time, deliver grammar explanation videos, and generate personalized pronunciation correction plans. Simultaneously, it can timely identify and intervene in situations potentially leading to social problems triggered by psychological factors, such as anxiety and depression. By early detection of individual psychological crisis signals (e.g., abnormal occurrence of suicidal vocabulary in speech), it assists psychological intervention teams in precise interventions, reducing the incidence of extreme events, enhancing social psychological safety protection capabilities, and maintaining social stability and harmony.

4. Discussion

4.1. Future directions

The integration of artificial intelligence and applied psychology is shifting from empowering single scenarios to constructing systemic ecosystems, presenting a three-tier progressive logic. First is the data-driven foundational layer (multimodal information collection). For example, constructing dynamic psychological profiles based on speech emotion features forms the data-driven foundational layer. In enterprise management, utilizing applied psychology theories on job burnout, integrating voice and physiological data to set reasonable performance indicators and incentive plans based on employees' personal goals and psychological needs can stimulate work enthusiasm and creativity. Second is translating psychological theories into algorithms.Utilizing learning analytics to continuously track student behavioral data during learning and develop personalized paths for each student is an example of the second tier of this logic. Thirdly, transcending the tool attribute and extending to socio-emotional governance represents the third tier. Developing emotional companion systems capable of simulating the voices of friends and family to enhance emotional trust through virtual reality aims to enhance emotional trust. Future development will

focus more on human-centered intelligent services. By integrating social media data, public opinion information with theories of social cognition and group dynamics, analyzing social hot topics, and predicting large-scale social behaviors, it can provide references for governments and social organizations in decision-making. Simulating group behaviors under different scenarios can assess policy implementation effects, identify potential social problems in advance, take preventive measures, and maintain social stability and harmonious development.

4.2. Challenges

The integration of artificial intelligence and applied psychology breaks through the functional limitations of traditional HCI. Interdisciplinary research has demonstrated practical value in areas such as emotion recognition and personalized services, but theoretical depth and technological maturity still need improvement. Furthermore, annotating emotional speech data is costly, typically requiring involvement of psychology professionals. Moreover, definitions of "emotion" vary across psychological studies, leading to inconsistent data annotation standards. Speech emotion data may reveal privacy issues related users' psychological states, which may spark controversy. Future efforts need to further establish technological ethics frameworks. Establishing principles of informed consent for emotional data collection and standards for algorithm transparency are essential to prevent technological misuse. Exploring cultural adaptation across diverse scenarios is also crucial, enhancing user emotional trust through the non-verbal interaction of embodied conversational agents. The deep integration of these fields must be guided by the principle of technological rationality serving humanistic values, aiming to build a more inclusive and humanistic future society [10].

5. Conclusion

This paper focuses on the cross-application of speech recognition and applied psychology. The study finds that while AI can achieve basic emotion recognition, deep semantic parsing of speech requires further refinement. Introducing applied psychological theories into AI can optimize algorithms' understanding of non-literal meanings through emotional cognition theories and develop embodied emotional feedback mechanisms based on social interaction theory. Research indicates that its effectiveness has been validated in multiple scenarios such as psychological diagnosis and personalized intervention. The study further points out that many challenges remain in HCI applications, such as the need for specialized technical personnel for integration, along with dilemmas like high costs and lack of standardization. Therefore, constructing an interdisciplinary methodological framework is necessary. However, this paper has certain limitations. For instance, it does not delve deeply into the theoretical exploration of applied psychology or the integration with emerging technologies, nor does it thoroughly explore multimodal fusion scenarios. Future research needs to further investigate the guidance of cognitive psychology theories for affective computing, to provide more humanistic technological support for digital mental health services.

References

[1] Klein, J. (1990). The potential of artificial intelligence in psychology. *Journal of Applied Psychology*, 75: 387-395.
[2] Cohen, J. F., & Seider, M. (2004). Facial expression and emotion. *Annual Review of Psychology*, 55: 591-614.
[3] Bänziger, T., & Scherer, K. R. (2005). Voice, speech, and communication. In: Handbook of nonverbal communication. *Oxford University Press.* pp. 161-187 .
[4] Bickmore, T., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *Transactions on Computer-Human Interaction.*12: 293-327.
[5] Kaul, V., Enslin, S., & Gross, S. A. (2020). History of artificial intelligence in medicine. *Gastrointestinal Endoscopy,* 92: 807-812.
[6] Terblanche, N., Molyn, J., De Haan, E., & Nilsson, V. O. (2022). The impact of coaching using a chatbot (ELEA) on managers' skills. *International Journal of Evidence Based Coaching and Mentoring,* 20: 20-35.
[7] Mumtaz, W., Ali, S. S. A., Yasin, M. A. M., & Malik, A. S. (2018). A machine learning framework involving EEG-based functional connectivity to diagnose major depressive disorder (MDD). *Medical & Biological Engineering & Computing*, 56: 233-246.
[8] Qiu H., & Wang X.. (2016). Application of artificial intelligence in psychology. *China Science and Technology Journal Database Research,* (12): 00040-00040.
[9] Yue, I., Wang, Q., Liu, B., & Zhou, L. (2024). Postoperative accurate pain assessment of children and artificial intelligence: A medical hypothesis and planned study. *World Journal of Clinical Cases*, 12: 681-687.
[10] Dave, R., Kaur, P., & Kaur, G. (2021). Artificial intelligence techniques for prediction of mental health disorders. *International Journal of System Assurance Engineering and Management,* 12: 254-263.