Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/31/20230122

Interaction mode enables user perception recognition and perception optimization: An AI human-computer interaction study

Yizhe Shen

School of Software Engineering, Nanjing University of Information Science and Technology, Nanjing, China

202183290526@nuist.edu.cn

Abstract. Artificial intelligent (AI) has various ways of human-computer interaction, but most of them overlook the recognition of human perception. If the interaction mode is combined with psychology, the user's mood change can be identified by the user's subtle expression, movement change and voice tone change, so as to provide corresponding services and improve the user experience. Statistical analysis of human responses to different situations in cognitive psychology, incorporating them into human-computer interaction methods. The current human-computer interaction modes in products tend to be standardized, and focusing user experience on user perception will bring special experiences to users. Emotional recognition is a cross disciplinary discipline with broad application prospects, but it has not yet reached a mature stage and requires corpus enrichment, theoretical strengthening, and method innovation. The era of artificial intelligence is leading a new wave of technological progress, and emotion recognition, as an important topic in the field of artificial intelligence, can help computer intelligence recognize human emotions and make human-computer interaction more friendly. In the near future, research on emotion recognition technology will make greater progress and be better applied to practical products.

Keywords: interaction mode, perceptual recognition, cognitive psychology.

1. Introduction

With the continuous development of artificial intelligence technology, the interaction between humans and AI will become increasingly close. In the future, the interaction between humans and AI is not only a simple application scenario, but also involves deeper levels of interaction such as thinking and emotions. Emotional interaction is another important direction for future human-computer interaction. At present, AI can recognize human emotional states through emotion recognition technology, but AI is still unable to truly understand human emotions and express their own emotions. In the future, with the continuous development of AI technology, AI will become closer to the emotional level of humans, and the emotional interaction between humans and AI will also become more and more profound. For example, in the field of education, AI can better understand students' emotional states through emotional interaction, thus facilitating better teaching.

Now AI human-computer interaction mainly converts user's commands into Machine code through voice, visual action and keyboard button input for the machine to execute. Speech recognition has

DOI: 10.54254/2755-2721/31/20230122

become very mature in artificial intelligence fields such as navigation and mobile search. Visual actions can also be widely used in areas such as facial recognition [1]. However, overly extensive recognition also has many drawbacks, making it difficult to meet users' high-quality requirements. It is necessary to optimize interaction methods and develop them in a more detailed direction.The combination of interaction mode and Cognitive psychology can identify the change of user's mood by the change of user's subtle expression, movement and voice tone, so as to provide corresponding services. Focusing user experience on the user's perception experience will bring special experience to the user [2].

Develop a new system to capture details such as changes in user expressions, movements, and sounds, as well as changes in input speed. For example, if a user shows a blush or other state during use, it will be recognized as excitement or anger. Then, analyze the user's speech rhythm and voice to identify specific emotional states and save data on the user's psychological changes, To fully preserve the differences in user perception and preferences [3].

Emotional recognition, as the cornerstone of intelligent human-computer interaction, its main task is to establish a computing system that can analyze and recognize human emotions from various modalities such as speech, text, and video, achieving humanized communication between humans and machines.The era of artificial intelligence is leading a new wave of technological progress, and emotion recognition, as an important topic in the field of artificial intelligence, can help computer intelligence recognize human emotions and make human-computer interaction more friendly. authors believe that in the near future, research on emotion recognition technology will make greater progress and be better applied to practical products.

2. The combination of human-computer interaction and cognitive psychology

2.1. The combination of speech recognition and cognitive psychology

Identify changes in the user's mental state during use and change the way services are provided by changing the tone and frequency of the voice spoken to the user. Figure 1 classified speech features and analyze users' emotions based on the characteristics of each category.

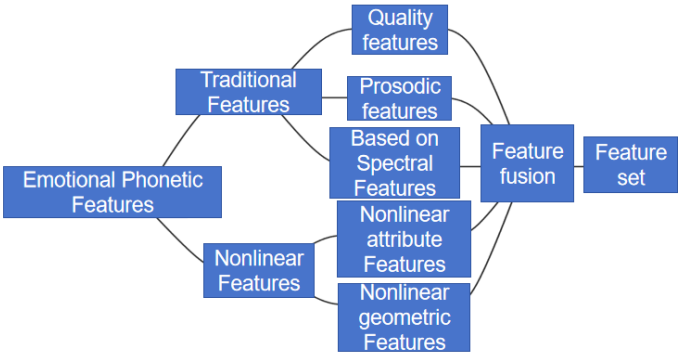

Figure 1. Collection of emotional speech features.

The phonetic features of human organs include the lungs, trachea, vocal cords, nose, mouth, and lips. The length and tension of the vocal cords determine the fundamental frequency of sound; The sound intensity depends on the size and intensity of the airflow; The sound resonates through the oral cavity, nasal cavity, and other changes in shape, and finally radiates through the lips; The resonance process of sound forms resonance peaks, and the frequencies of each resonance peak are determined by the size and shape of the resonant cavity.

Prosodic features: The most important phonetic features, such as speech speed, volume, and tone, amplitude, pitch frequency, duration, etc; Speech speed: Speech can reflect the emotional state of the

DOI: 10.54254/2755-2721/31/20230122

speaker: when a person's emotions are relatively excited, such as in an angry state, the speed of language expression is significantly faster. On the contrary, when a person's emotions are relatively low, such as in a sad state, the speed of language expression is significantly slower.

And through spectrum analysis, linear prediction cepstrum coefficient LPCC is based on the assumption that the speech signal is an autoregressive signal, and linear prediction analysis is used to obtain the cepstrum coefficient.

The pronunciation of different emotions can cause different changes in the vocal tract, leading to changes in the cepstrum of the vocal tract transfer function. The Mel frequency domain cepstral coefficient (MFCC) considers the different resolutions of the human ear in different frequency bands, fully integrating the auditory characteristics of the human ear.

Identifying user emotional changes requires a large amount of data collection, which can be achieved using a voice emotion database. Choosing a high-quality voice emotion database directly affects the quality of perceptual recognition. Because most of the data in the voice emotion database are private, and considering the capacity of voice and the number and professional level of testers, SUSAS was finally selected. SUSAS, including voice under simulation and actual pressure, is a voice emotion database created by the Robust Speech processing Laboratory of the University of Colorado at Boulder under the guidance of Professor John H.L. Hansen and sponsored by the Air Force Research Laboratory [4]. Its data sources include four fields, with 32 speakers and a language capacity of over 16000, covering a variety of stress and emotions, including anger, fear, depression, anxiety, fear, and other emotions. Figure 2 showed the process of collecting voice signals and converting them into emotions.

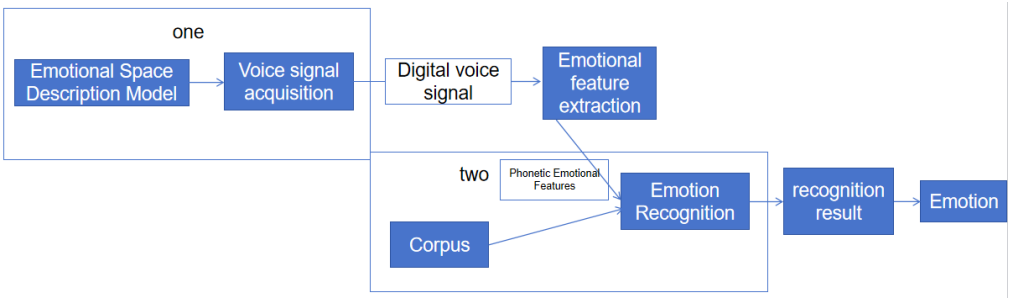

Figure 2. Emotional speech recognition process.

Integrating SUSAS into human-computer interaction speech recognition will change users' impression of indifference towards human-computer interaction. When users use the software, the software's response will change in real-time based on their voice emotions. The combination of voice and emotion can not only improve user satisfaction, but also be applied in psychological counseling. When users display emotions such as anger, fear, depression, anxiety, and fear when interacting with the software, the software can promptly soothe and stabilize the user's emotions [5]. If the emotional outburst is too severe, the software can promptly contact an offline psychological counselor or contact the police.

2.2. The combination of action recognition and cognitive psychology

By using micro expression change recognition and micro action recognition in psychology, users can identify changes in their mental state during use and change the way services are provided.

Action recognition uses a camera to detect user actions, collect changes in the user's actions during interaction with the software, and analyze them based on the information from the body action emotion recognition dataset. The PAD emotion three-dimensional theory is mainly used in action emotion recognition. The model believes that emotions have three dimensions: pleasure, activation, and dominance, where P represents pleasure degree and represents the positive and negative emotional state of the subject; A represents Arousal non arousal, indicating the level of neurophysiological activation of the subject; D represents Dominance submassiveness, indicating the subject's state of control over the

DOI: 10.54254/2755-2721/31/20230122

situation and others. It is the state in which individual emotions are controlled and dominated. Spontaneous emotions from the inside out are dominant emotions such as anger, while passive emotions from the outside in are submissive emotions such as fear. At the same time, the values of these three dimensions can also be used to represent specific emotions and emotions. Research has shown that using the three dimensions of PDA can effectively explain human emotions. Mehrabian et al. used these three dimensions to explain the vast majority of variations in the other 42 emotion scales, and these three dimensions are not limited to describing subjective experiences of emotions. They have a good mapping relationship with external manifestations of emotions and physiological arousal. Figure 3 showed the PAD three-dimensional theory of emotions suggests that emotions have three dimensions: pleasure, activation, and dominance.

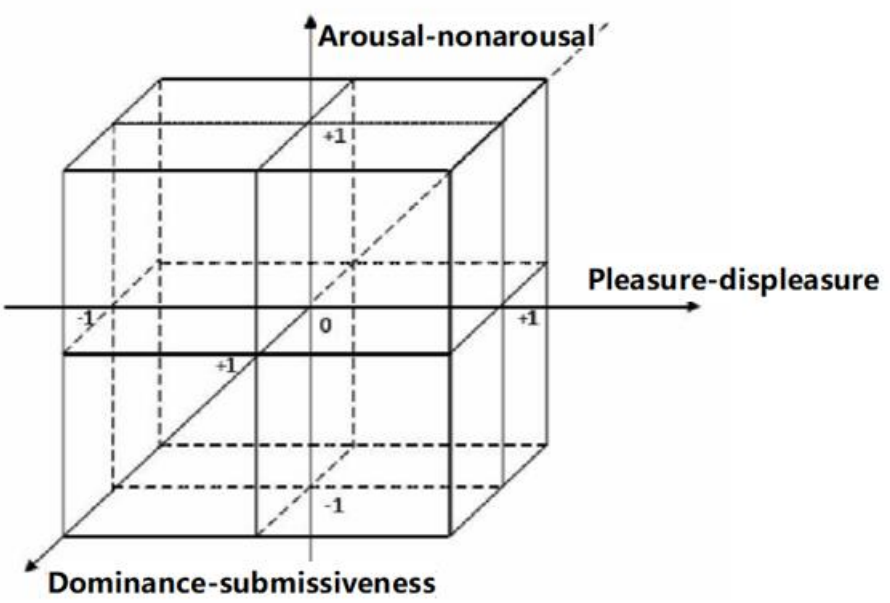

Figure 3. PAD emotional three dimensional theory [6].

The use of PAD three-dimensional emotional models can fully express and quantify human emotions and emotions [7], and can effectively identify the user's movement emotions.

2.3. Combination of image and text input and cognitive psychology

Users may also experience emotional changes when using image and text input, as well as changes in their text input speed and the type of input text and images. Based on these changes, data collection is used to analyze the characteristics of images and text. The features of images and text are refined into individual units and brought into the image and text emotional dataset. This can determine the user's psychological changes and provide corresponding feedback to the user. Figure 4 showed the process of emotional analysis in graphics and text.

DOI: 10.54254/2755-2721/31/20230122

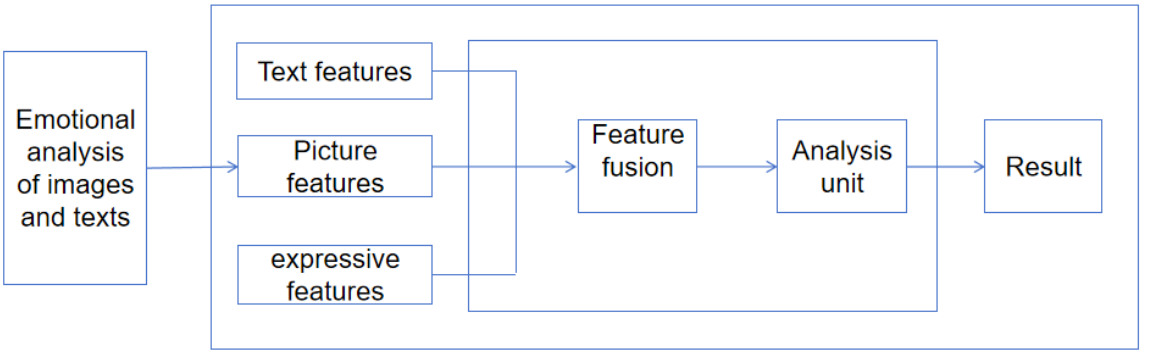

Figure 4. The process of emotional analysis in graphics and text [8].

For the judgment of image and text, two image and text sentiment datasets, Twitter-15 and Twitter-17, are used. These two are multimodal datasets that contain text and corresponding images of the text. The dataset annotates the target entity and the emotional tendencies expressed in its image and text. The entire data scale is Twitter-15 (3179/1122/1037) striped image tweets, Twitter-17 (3562/1176/1234) striped image tweets, and sentiment annotations classified into three categories. The basic categories are pleasure, sadness, and anger. Quantify images and text into emotions, and bring them into the dataset to obtain the user's approximate emotional state. The software can respond to the user's needs in an interactive state as soon as possible, meeting their needs.

3. Optimizing user experience through perception

Figure 5 showed people make choices, generally, perceptual choice precedes rational choice.

Figure 5. Perceptual choice and rational choice.

When users choose to use software, they usually make emotional choices before rational ones, so their first impression of the software is very important. The first glance of visual interaction between users and software will be the aesthetic choice in their minds. If the aesthetic meets the standards, it will greatly enhance users' expectations for use. Coupled with subsequent perceptual recognition, it will further enhance users' user experience and strengthen the relationship between users and the software.

Often, well-designed software can better attract users' desire to use it. If an interactive software has a beautiful appearance and bright and gentle colors, it can not only bring users a good emotional experience, but also make a qualitative leap in the value of the software itself, improve quality, and attract users. Based on the analysis of influencing factors designed at three levels: instinct level, behavior level, and reflection level (see Figure 6).

DOI: 10.54254/2755-2721/31/20230122

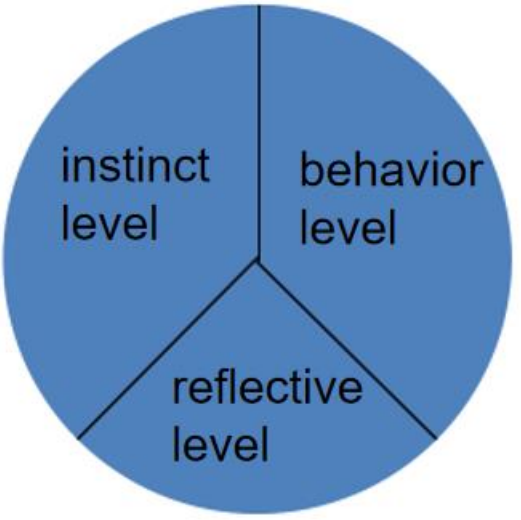

Figure 6. Three levels of influencing factors.

Divide users' choices of software into three layers: instinctive layer, behavioral layer, and reflective layer. The instinctive layer represents the subjective judgment of human instincts on the visual effects of interface colors and patterns. Therefore, in order to meet the user's instinctive needs, exquisite graphics and attractive colors are needed. The behavior layer represents the functionality of the software, including its functionality, comprehensibility, practicality, and other features. The reflective layer goes further on the instinctive level, representing an individual's reflective aesthetic. This layer includes what users believe using the software can bring to users, such as meeting their needs to establish their self-image and social status [9]. By optimizing the software at three levels, users' expectations for the software can be significantly increased.

Perception affects users' emotions through their recognition of software, and in most cases, cognition is a rational judgment that logically determines the quality of a software; Emotions, on the other hand, are a product of emotions and a manifestation of software satisfaction. Cognition and emotion complement each other and cannot be separated. To provide users with a good experience, it is necessary to provide a good cognitive experience, so that users can see it as if they are using it. Then, the subsequent cognitive experience can stabilize the user's sense of experience and make them feel the value of interaction.

4. Discussion

With the rapid development of artificial intelligence, there is a significant growth in mobile applications and startups in emotional computing, among which mobile applications are currently mainly used in industries such as business negotiations, healthcare, remote education, safe driving, and public services [10]. However, the current limitations of emotional interaction are also significant. The related technologies of emotional interaction are still immature, and the cost of products that involve people's privacy is relatively high. The emotional information in modalities such as speech and video has sparsity, and not all moments in speech or frames in video contain emotional information. How to transform sparse emotions into an overall emotional dataset is challenging. Moreover, human emotions have a broad subjectivity, and different individuals have varying degrees of response to the same emotion, and the emotional reactions of the same person to the same scene at different times may also be different. Therefore, how to focus on emotional research in more application scenarios and personalized backgrounds under natural conditions is still a difficult problem that has not been solved yet. At present, research on emotion recognition still faces problems with the generalization ability and real-time

DOI: 10.54254/2755-2721/31/20230122

performance of models, and existing emotion databases are not the same and have small scales, with differences in language, emotion types, and a lack of unified standards.

5. Conclusion

Using sentiment analysis datasets as a base and combining human-computer interaction with cognitive psychology can effectively avoid situations where users perceive human-computer interaction as indifferent and encourage more people to accept it. And human-computer interaction with emotional analysis can bring users a better experience. The current trend is the continuous development of human-computer interaction methods, but it ignores the humanistic emotions in human-computer interaction. Emotional recognition related products have to some extent improved people's sense of life experience, alleviated the shortage of social resources, and become a powerful tool for emotional communication among special groups. However, its related technology is still immature, and can only distinguish a small number of specific emotions. In addition, it involves user privacy issues and the cost of related products is high, making it difficult to promote them on a large scale. In fact, emotion belongs to the advanced attributes of human beings, and emotion recognition is a cross discipline of physiology, psychology, Cognitive science and other fields. Even human beings who have evolved for more than ten thousand years sometimes cannot clearly recognize each other's emotions, which brings great opportunities and challenges to the current field of emotion analysis.

References

[1] Lu X, Teng Z, Fan L. Exploration of the Future Development Trends of Human Machine Interaction in the Intelligent Era [J]. Wireless Internet Technology, 2021,18 (03): 7-8

[2] Gao Z, Lu J, Peng D. Evaluation method of human-computer interaction design scheme for user perceptual needs [J]. Systems Science and Mathematics, 2023,43 (03): 610-628.

[3] Chen Y, Wang S. Analysis of visual psychological cognition and emotion design in UI Design [J]. Art and Design Research, 2021, No.94 (02): 74-79.

[4] Zhang H, Huang H, Li W. A Speech Emotion Database for Emotional Change Detection [J]. Computer Simulation, 2021,38 (09): 448-455

[5] Shi G, Fan S, Wang L. Research and application of speech and emotion recognition technology in psychological counseling [J]. Yangtze River Information and Communications, 2022,35 (09): 19-22.

[6] Zhang X, Zhang T, Sun Y, etc. Cascade classified emotional speech recognition based on PAD model [J]. Journal of Taiyuan University of Technology, 2018,49 (05): 731-735.

[7] Shen Z. Study on emotion recognition based on body movements [D]. University of Chinese Academy of Sciences (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences), 2020.

[8] Meng X, Yang W, Wang T. A Review of Emotional Analysis Research Based on Image Text Fusion [J]. Computer Applications, 2021,41 (02): 307-317.

[9] Li H, Sun X. Emotional+makes intelligent human-computer interaction possible in the future. CHINA SCIENCE DAILY[N], 2021-5-20(3).

[10] Tao J, Chen J, Li Y. A Review of Speech Emotion Recognition [J]. Signal Processing, 2023,39 (04): 571-587