Design and Application of Virtual Avatar Framework Based on E-commerce Live Streaming

Xing-hong Fu[1], Qing-sheng Li[2], Na-ye Ji[2], Han Xu[3], Zhen-hui Chai[1], Hong Yang[1]
1.School of Communication and Journalism, Communication University of Zhejiang, Hangzhou, China
2. School of Media and Information, Engineering, Communication University of Zhejiang, Hangzhou, China
3. Department of Computer Sciences, University of Exeter, Exeter, EX4 4QF, UK
Email: fuxinghong9923@163.com

***Abstract*—With the rapid development of live streaming services, enterprises, as well as online celebrities, are increasingly in demand for 24-hour live streaming support to engage a broader audience. However, no person can broadcast twenty-four hours a day for a long period of time. Therefore, the virtual avatar is proposed to replace the real avatar for streaming. Based on the new thinking on the use scene of digital human, this paper designs a virtual avatar called Miebo. Through a series of experiments, it is found that the parameters that can produce the most effect on Miebo's emotion of "Sadness", "Happiness", "Surprise" and "Anger" is "Upset", "Grin mouth", "Open mouth big" and "Eyes closed" respectively, and the presentation results of Miebo's facial expression are optimized according to the experimental analysis. We also present a set of highly feasible virtual avatar application schemes and develop a special platform for virtual avatar design called "Virtualfull", which allows ordinary users could create their own virtual avatars.**

***Keywords*—e-commerce live streaming, virtual avatar, artificial intelligence**

I. INTRODUCTION

With the continuous upgrading of Internet technology and ubiquitous usage of mobile devices, the e-commerce live streaming has become one of the most popular commodity sales models. Meanwhile, the outbreak of Coronavirus has affected the way people live, which has additionally led to the rapid development of live streaming services. In the circle of e-commerce avatars, there is a popular saying that the first live broadcast is not important, but the 100th live broadcast is important. The truth is, no person can broadcast twenty-four hours a day for a long period of time, which even the Taobao studio "lipstick brother" Li Jiaqi also made the same exclamation. In the post-epidemic era, people from all walks of life pay more and more attention to e-commerce live streaming. Enterprises, as well as online celebrities, are increasingly in demand for 24-hour live streaming support. They want to keep the live stream running and engaging consumers, reducing the loss of potential consumers due to the end of the stream. In addition, excluding the limited time and space of real avatars, the recruitment cost and training cost of a real avatar are also very high. Based on the pain points existing in real avatars in e-commerce live streaming, Virtual avatars have naturally been proposed.

A set of data showed that there were 6,000 virtual avatars from all over the world in 2019. In March 2020, Taobao introduced the first virtual avatar called "Brother Wolf". Through the statistical result of Taobao live broadcast account, the number of viewers of the first official live broadcast once rose to 1.42 million. During April of the same year, virtual avatar Luo Tianyi appeared in the popular cargo avatar Li Jiaqi's studio, which attracted the number of online viewers on Taobao was once as high as 6 million [1]. On the eve of Singles 'Day, virtual avatars, Ali AI, appeared in business studios such as Givenchy, L' Oreal and Nestle alone [2].

To generate the animation of the virtual avatar, real-time data is collected from voice, face, and body movement. The advent of artificial intelligence-generated virtual avatars marked the era of virtual avatars shifted from virtualization of real avatars to automatically generated avatars.

At present, the group of Virtual You Tuber (Vtuber) is higher than that of Virtual Uploader (Vup) in terms of fans, income, and popularity of VUP. For example, among the number of active virtual avatars on Station B follows more than 100,000, Japanese virtual avatars still constitute the majority [3]. This data just shows that English speaking or Japanese speaking consumers and audiences have more acceptance and preferences for virtual avatars. Presently, both at home and abroad, virtual avatar technology is being applied, but from the research literature on virtual avatar in China, there is a lack of relevant research in its research field., so how virtual avatars can play their greatest utility in e-commerce live broadcast is also the problem what we need to think about and solve.

In this work, we aim to combine the benefits of voice drive and real-time control technology by proposing a new client application architecture for vivid virtual avatar models called Miebo. Our goal is to break the barrier of the real avatar between twenty-four-hour continuous live broadcasting and live broadcasting at anytime and anywhere, for providing higher quality electricity live service. In addition, Alibaba, Jing Dong, or the virtual avatars in the big avatar broadcast room are behind the support of capital. While online streamers seek to share themselves through virtual avatars and express themselves on social platforms, the high cost makes this impractical for them. Therefore, based on the virtual digital human technology and mature cloud side construction technology, this paper built a platform specially providing virtual avatar services -- "Virtualfull", letting the technology of virtual avatar enter the homes of ordinary people.

The performance of virtual avatars is superior to that of real avatars when their professional skills of streaming are the same as real avatars [4]. Although the current research on virtual avatars is on the rise year by year, there is still a dearth of research on the application of virtual avatars in e-commerce platforms in the academic industry and live streaming industry. Furthermore, the "silent singing" live accident of the virtual singer " Tian-yi Luo " in the broadcast room of Li Jiaqi has revealed that the technique and application of virtual avatars require significant improvement and innovation.

In addition, research shows that the effective design and application of human-computer interaction technology can promote consumers' positive emotional experience, and the

978-1-6654-6667-7/21/$31.00 ©2021 IEEE
DOI 10.1109/IUCC-CIT-DSCI-SmartCNS55181.2021.00077

improvement of human-computer interaction technology is conducive to consumers' ability of independent choice and enhance consumers' shopping desire [5]. On October 22, 2021, the State Administration of Radio, Film and Television recently released the "14th Five-Year development plan on radio, television and network audio-visual science and technology ". The plan calls for the wide use of virtual avatars and animated sign language in news broadcast, weather forecast, variety science and education. At the same time, explore the application of short video avatars, such as digital web celebrities, live streaming, and other virtual images in the interactive links of the program, to increase personalization and interest.

The rest of this paper is organized as follows. Section 2, we mainly elaborate the virtual avatar and related technical theories, including the general framework of the operation process of virtual digital human and its technical composition. Section 3 the new framework for the virtual avatar Miebo, including its Client design and construction of background data, and briefly records our experimental process and the results of data analysis. Section 4 mainly designs two feasible schemes for the virtual character, service platform--"Virtualfull". Section 5 shows the conclusion of our work and the future and direction. Section 6 express our sincere gratitude to the persons who supported the completion of the paper.

II. Related Work

A. Digital human

A Digital human has two concepts: narrow and broad. In a narrow sense, it is the product of the integration of information science and life science. It is the virtual simulation of the shape and function of the human body at different levels by using the method of information science. It includes four overlapping development stages, visual human, physical human, physiological human, and intelligent human. Finally, it establishes a multi-disciplinary and multi-level digital model and achieves the accurate simulation of the human body from micro to macro. In a broad sense, digital man refers to the penetration of digital technology at all levels and stages of human anatomy, physics, physiology, and intelligence. It should be noted that digital man is a general term for relevant fields in the development stage. At present, the application of virtual digital human in Pan Entertainment (film and television, media, games, etc.), finance, culture and tourism, education, education, medical treatment, retail, and other key scenes is becoming more and more popular [6].

B. System framework construction of Digital Human

At present, as a new generation of technology to realize virtual interaction, digital human is still in the developmental stage, and there is no unified universal standard for the design of digital human system framework. Most of the digital human services and products currently available on the market are based on a universal system framework of digital humans. As shown in Fig. 1. In general, the virtual digital human system is composed of five modules, including: (1) character image design, (2) voice generation, (3) animation generation, (4) audio and video synthesis and display, (5) and human-computer interaction.

In character image module, the character image can be divided into 2D and 3D, and be converted to into cartoon, anthropomorphic, realistic, or super realistic, etc. Then, in the speech generation module and action generation module, the corresponding character voice and matching character animation can be generated respectively based on text. The audio and video synthesis display module combines video, voice, and animation to create a visual display for the users. Finally, the human computer interaction module enables digital human to have the interactive function, that is, to identify the user's intention through intelligent technologies such as voice semantic recognition. In addition, the human-computer interaction module can be used to determine the subsequent actions and voice of the digital human based on the user's current intent. Then, the iteration of using the human-computer interaction module culminates in the movement of the digital human.

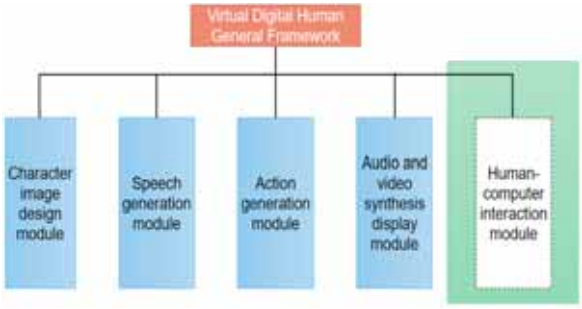

Fig. 1. The virtual digital human universal system framework

Based on the two core pain points of e-commerce live streaming identified in the introduction, this paper examines interactive digital humans and builds a virtual avatar platform through 3D cartoons and live 2D digital humen-driven technology.

Interactive digital human can be divided into intelligent driven and real-life driven according to the different driving methods. An intelligent driven digital human can automatically read and analyze the external input information through the intelligent system and determine the subsequent text output of the digital human according to the analytical results, and then drive the character model to generate the corresponding voice and action to allow the digital human could interact with the users. The operation process is as shown in Fig. 2. The character model is pre-trained through AI technology to generate speech and corresponding animations through text drive, which is referred to in the industry as the TTSA character model in the industry. The digital human is driven by the real person. The main principle is that the real person can communicate with the user in real time according to the user video transmitted from the video monitoring system, and user real-time voice, at the same time, the expression, action and voice collected through action capture collection system is transferred to the audio and video synthesis display system to generate the virtual digital human image, so as to interact with the user, the operation process is shown in Fig. 3.

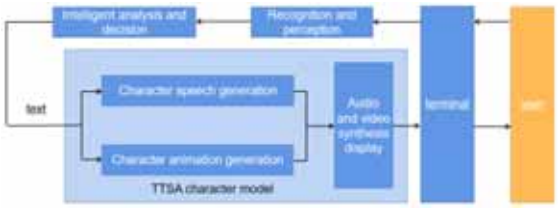

Fig. 2. Operation process of intelligent-driven virtual digital human

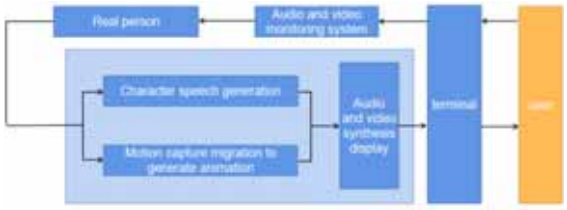

Fig. 3. Live-action-driven virtual digital human operation process

C. Technical architecture of virtual digital human

The digital human is currently related to many technical fields, and the production mode of the digital human has not yet been determined. Through the research on the general technologies involved in the production of existing virtual digital human, the following technical structures are extracted based on the framework of virtual digital human general system, as shown in Fig. 4. Green module refers to the five technical modules used for digital human production and interaction, namely, character generation, character expression, composite display, recognition, and perception, analysis, and decision-making and other modules. Among them, character expression includes voice generation and animation generation. Animation generation is composed of two major parts: driver (action generation) and rendering. The orange module refers to 2D and 3D digital human module. Compared with 2D digital human model, 3D digital human model needs to use additional 3D modeling technology to generate a digital image, the information dimension increases, and the larger computational amount is required.

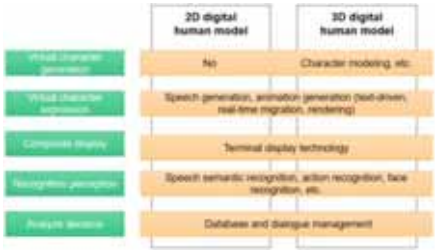

Fig. 4. Virtual digital human technical architecture

III. The Architecture for Virtual Avatar Miebo

The design of a virtual avatar is composed of two parts Client design and server design, and both have been considered in this paper. The front end is image display and discourse output, including voice text output and virtual avatar image design; The back end is responsible for the automatic generation of content. The module that used for the connect the interaction between of Client and server is called text. Its flow is shown in Fig. 5.

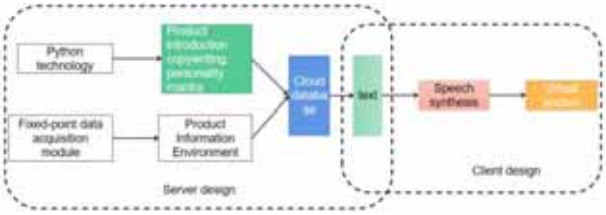

Fig. 5. The virtual avatar design process

The following part will introduce the design ideas of the virtual avatar Miebo from the Client design and the server design.

A. Client architecture

1). Model design

The virtual avatar Miebo, as the avatar with goods, can take the image of any real avatar such as star, online celebrity, KOL and KOL as the prototype, and can also be made into any lovely cartoon image that is not real. Where the 3D character is similar to the real characters, but the animation of 3D is much more difficult. On the contrary, character images in 2D are easy to create, but are not sufficiently vivid. This paper designs a three-dimensional image for virtual avatar Miebo through 3D Max technology. The Miebo character skeleton is modeled in the design software. After the character skeleton modeling is completed (as shown in Fig. 6), adjust the parameters such as "animation" "IK" "Outline" "physics" "show IK bones" "show rigid" to synthesize a virtual avatar Miebo dance animation video. About Miebo's image design, firstly, from the perspective of modern aesthetic vision, "White & Naïve & Charming" and "the First Love Face" can attract more attractive attention and desire for protection [7-8], so the image of the characters in this paper selects the image of lovely and clever animation young women who are more in line with the public aesthetic style. Secondly, the clothing mainly adopts the modern improved Hanfu combined with Chinese culture, which integrates the elements of the school emblem logo of Communication University of Zhejiang and presents the sense of belonging of the works and publicizes Chinese culture and media culture at the same time. Finally, the image of virtual avatar Miebo is presented, as shown in Fig. 7.

Fig. 6. Skeleton modeling of virtual avatar Miebo

Fig. 7. The 3D image of the virtual avatar Miebo

2). Speech construction

Following the completion of the design of the Miebo's appearance. Designing a voice for the avatar poses an additional challenge, as a beautiful voice could leave a lasting impression on the audience.

As shown in Fig. 8, we adopt text conversion technology (Text-To-Speech, TTS) to let the text output by voice. The TTS technology can transform any text information into format and process natural voice and read it out in real time. It is an interdisciplinary subject of linguistics, voice signal processing, multimedia communication, electronic technology, and other disciplines. TTS technology can effectively solve the problem of transforming text information into speech information, and with the combination of virtual reality technology, the virtual avatar could realize the action of speaking.

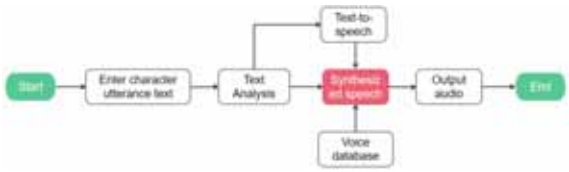

Fig. 8. TTS text transformation process

3). Voice-driven animation

The mouth shape and face movements are also the key component of a virtual avatar. That is, the Miebo's mouth shape needs to follow the pronunciation rhythm and trembling, this technology is called voice-driven animation (Speech-Driven Animation, STA) technology, namely the technology through voice-driven virtual image speech and feedback emotions and movements. The technology combines deep learning neural network and computer graphics, allowing the computer to understand the content of voice and finely adjust the avatar lip action, facial expression, and body posture, generating a very realistic avatar animation. The STA method is mainly divided into text-driven animation and speech-driven animation. Text drive is mapping text information to face actions and expressions, which transforms text information into a phoneme sequence. Database is used for mapping. Text information can be directly converted into animation view, or text information can be converted into animation view based on rules. Speech-driven face animation, like text-driven, is a process of synthesizing the corresponding facial expressions and lip mouth movement with the input voice based on the virtual human face model. Voice information is converted to face animation. The specific process is as shown in Fig. 9.

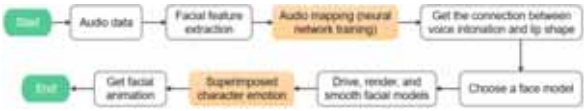

Fig. 9. Voice-driven animation process

4). Perceptual design of human-computer interaction

Real-time expression migration. There are two main methods to realize real-time expression migration: one is to input the real-time expression change of the user through the interaction on mobile terminal interaction, and restore the image 1:1 in the virtual digital environment, as shown in Fig. 10; another is to input a performance video or multiple expression pictures of the mobile phone interaction, map the virtual digital human and the user face key points using 2D RGB camera in real time and use GAN to further estimate face posture [9-10]. The reconstruction function of 3DS MAX transfer a 2-dimensional image to a three-dimensional face model, rendering the two-dimensional facial expression of the current posture in real time. Then, the performance face expression is fused with the target face expression to realize the automatic capture of face, the estimation of target face, the reconstruction and automatic rendering of three-dimensional face, and finally realize the real-time expression migration.

Fig. 10. Digital human restores user real-time expression migration

Real-time action migration. The real-time action migration of characters is performed primarily using somatosensory recognition technology. Real time fusion of human posture and full wireless motion capture of the foreground virtual avatar, collecting the posture data of the moving limb through the sensor node and sending it to the host computer through Wi-Fi wireless transmission. The host computer calculates the posture data and space-time trajectory and drives the virtual model to follow the real human motion [11-12]. Through the multi view motion capture method of the virtual interactive system [13], the human posture and foot trajectory are combined to realize the real-time tracking of the

posture and absolute position of the human body in space by the human model, which can basically follow the real human body in a variety of motion states. Through the combination of personal perceptions and human-computer interaction, the virtual digital person has a sense of authenticity and affinity that enhances the user's experience, making it more immersive.

Fig. 11. Virtual avatar driven by real-time action migration of characters

B. Server architecture

The server design of the avatar first needs to store the information related the product and dialogue that used to introduction and recommendation to consumers. Meanwhile, the personal characteristic of the real avatar is also needed to record, such as catchphrase, product introduction characteristics and habits, such as Li Jiaqi's "oh, my god", "buy it! buy it!", "attention all girls!", etc. Crawler technology can be used on most product web page (Web Crawler, also known as Web Spider or Web information collector, is a computer program or automated script to automatically download web pages and is an important part of the search engine[9]) for crawling the product principle and other relevant information, then translate them into natural language, and deposit the crawled information as keywords into the cloud database of "Virtualfull", a created virtual avatar service platform. When asked to close Miebo's live talk to real avatars, the staff can adjust and improve the data that Miebo crawled in the server module, insert the personal characteristics of real avatars into the virtual one, making it like the personal image of real avatars.

At present, there are mainly two common virtual digital human schemes for users to choose on the "Virtualfull" virtual avatar service platform. Users can customize virtual avatars according to their specific needs and input relevant information.

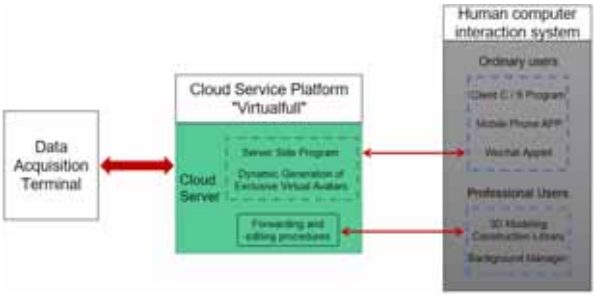

Fig. 12. Virtual avatar service platform - "Virtualfull" Architecture

Live-action driven mode. Real person driving mode refers to professional and accurate data driving, including action grasping and facial data grasping. This application scheme can meet the needs of different users, as shown in Fig. 13. Under human-computer interaction, the user's real person drives the action, expression, and expression of the virtual digital human to complete some cross space live broadcasting instead of the user. It is suitable for marketing activities or endorsement performances that are not broadcast live for a long time.

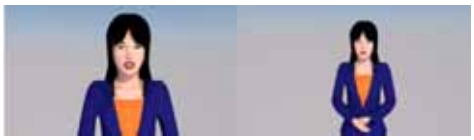

Fig. 13. Demonstration of the live-action drive mode

AI driven mode. AI drive mode refers to the design of virtual digital human image (which can be anthropomorphic) instead of real person avatars driven by artificial intelligence technology and code editing. AI texts or voice driven digital human avatar to speak, expression, movements, etc. As shown in Fig. 14. This virtual avatar is suitable for 24-hour uninterrupted live streaming, auxiliary with goods, FAQ question and answer, circular broadcasting, radio, and other scenarios.

Fig. 14. The AI drive mode is shown

IV. Results and Discussion

To make the external image of the virtual avatar and the presentation effect of facial rhythm driven by voice more vivid and natural, this paper obtained the change process of four basic emotions: Sadness, Happiness, Surprise and Anger by repeatedly changing the parameters of the facial expression of the virtual avatar Miebo and extracted 20 groups of data with obvious expression changes from multiple groups of experimental data. The experimental data are shown in the tables below.

A. Presentation of experimental data

TABLE I. EXPERIMENTAL DATA OF MIEBO "SADNESS"

Morphs A	parm 1	parm 2	parm 3	parm 4	parm 5
open mouth	0.11	0.11	0.11	0.11	0.11
open mouth big	0	0	0	0	0
grin mouth	0.39	0.39	0.39	0.39	0.39
pouting mouth	0	0	0	0	0
sad	0.4	0.4	0.4	0.4	0.4
surprised	0	0	0	0	0
angry	0	0	0	0	0
upset	0	0	0	0	0.8
eyes catching	0	0	0.73	0.73	0.73
eyes closed	0	0.24	0.24	0.59	0.59

TABLE II. EXPERIMENTAL DATA OF MIEBO "HAPPINESS"

Morphs B	parm 1	parm 2	parm 3	parm 4	parm 5
open mouth	0.26	0.74	0.74	0.26	0.74
open mouth big	0.35	0.35	0.35	0.35	0.35
grin mouth	0.04	0.04	0.87	0.87	0.87
pouting mouth	0	0	0	0	0
sad	0	0	0	0	0
surprised	0	0	0	0	0
angry	0	0	0	0	0
upset	0	0	0	0	0
eyes catching	0	0	0	0	0.95
eyes closed	0	0	0	0	0

TABLE III. EXPERIMENTAL DATA OF MIEBO "SURPRISE"

Morphs C	parm 1	parm 2	parm 3	parm 4	parm 5
open mouth	0.35	0.74	0.74	0.74	0.74
open mouth big	0.34	0.34	0.88	0.88	0.88
grin mouth	0	0	0	0	0
pouting mouth	0	0	0	0	0.79
sad	0	0	0	0	0
surprised	0.22	0.22	0.22	0.76	0.76
angry	0	0	0	0	0
upset	0	0	0	0	0
eyes catching	0	0	0	0	0
eyes closed	0	0	0	0	0

TABLE IV. EXPERIMENTAL DATA OF MIEBO "ANGER"

Morphs D	parm 1	parm 2	parm 3	parm 4	parm 5
open mouth	0.35	0.35	0.35	0.35	0.35
open mouth big	0	0	0	0.57	0.57
grin mouth	0	0	0	0	0
pouting mouth	0	0	0	0	0
sad	0	0	0	0	0
surprised	0	0	0	0	0
angry	0.16	1	1	1	1
upset	0	0	0	0	0
eyes catching	0	0	0	0	0
eyes closed	0	0	0.65	0.65	0.92

B. Experimental results show

According to the above experimental data, different expression states of virtual avatar Miebo under different parameter weights are obtained, as shown in the figures below.

Fig. 15. Miebo's experimental results of "Sadness"

Fig. 16. Miebo's experimental results of "Happiness"

Fig. 17. Miebo's experimental results of "Surprise"

Fig. 18. Miebo's experimental results of "Anger"

C. Experimental data analysis

According to the data in Table I. and the effect presentation in Fig. 15. From the first group of data and the second group of data, changing the "eyes closed" parameter can aggravate Miebo's mood of "Sadness" to a certain extent. According to the first and third groups of data, Miebo's mood changes more dramatically when the "eyes catching"

parameter is altered as compared to changing the parameter of "eyes closed". The weight of Eyes Catching is greater than that of Eyes Closed. The comparison between the second group of data and the third group of data further verifies the above conclusions. From the data of the third group and the fourth group, continuously increasing the parameter of "eyes closed" can continuously aggravate Miebo's Sadness. From the fourth and fifth groups of data, adding the parameter of "Upset" can affect Miebo's Sadness to the greatest extent. All in all, Upset has the larg

est weight, followed by Eyes Catching, and finally Eyes Closed.

According to the data in Table II. and the effect presentation in Fig. 16. From the first group of data and the second group of data, changing the parameter of "open mouth" can emphasize Miebo's mood of "happiness" to a certain extent. From the first and the third groups of data, changing the parameter of "grin mouth" can affect Miebo's happiness more than changing the parameter of "open mouth". That is, the weight of Grin Mouth is greater than that of Open Mouth. The comparison between the second group of data and the third group of data, which also can verify the above conclusions. From the third and fourth groups of data, continuously increasing the "grin mouth" can continuously strengthen Miebo's happiness. The first and fourth groups of data further verify the conclusion that the weight of Grin Mouth is greater than that of Open Mouth. From the data of the fourth group and the fifth group, changing the "eye catching" can affect Miebo's happiness to a certain extent, but from the comparison between the data of the third group and the fifth group, the impact factor of "grin mouth" is still stronger than "eye catching", and the impact of "open mouth" will also be greater than "eye catching". In a word, Grin Mouth has the largest weight, followed by Open Mouth, and finally Eye Catching.

According to the data in Table III. and the effect presentation in Fig. 17. From the first group of data and the second group of data, changing the parameter of "open mouth" can emphasize Miebo's "surprise" and shock to a certain extent. From the second and third groups of data, changing the parameter of "open mouth big" can obviously change Miebo's surprised expression, and adjusting the parameter "open mouth big" can affect Miebo's expression more than adjusting the parameter of "open mouth". That is, the weight of Open mouth big is greater than that of Open Mouth. From the data of the third group and the fourth group, adding the parameter of "surprised" can continuously deepen Miebo's surprised expression changes. However, from the comparative changes of the second and third groups or the third and fourth groups, the change of the parameter of "open mouth big" will affect Miebo's expression more than the change of the parameter of "surprised", that is, Open mouth big has the largest weight, followed by Open mouth, and finally Surprised. From the comparison of the data of the fourth group and the fifth group, the parameter of "pouting mouth" can emphasize Miebo's surprise, and the influencing factor is stronger than "surprised", that is, Open mouth big has the largest weight, followed by Pouting mouth, and finally Open mouth.

According to the data in Table IV. and the effect presentation in Fig. 18. From the first group of data and the second group of data, changing the parameters of "anger" can directly lay the emotional tone of Miebo's anger to a great extent. From the second and third groups of data, changing the "eyes closed" parameter can continuously show the outbreak of anger, and we can clearly see the emotional change, that is, the weight of Eyes closed is greater than that of Anger. From the data of the third group and the fourth group, changing the "open mouth big" parameter can affect Miebo's anger, but the influence is not obvious, that is, Eyes closed has the largest weight, followed by anger, and finally is Open mouth big. The comparison of the fourth and fifth groups of data further verifies the results of the comparison of the second and third groups of experimental data. The parameter of "eyes closed" can indeed affect Miebo's angry mood to a greater extent.

To sum up, the factors that have the greatest impact on Miebo's emotional parameters of "Sadness", "Happiness", "Surprise" and "Anger" is "Upset", "Grin mouth", "Open mouth big" and "Eyes closed" respectively.

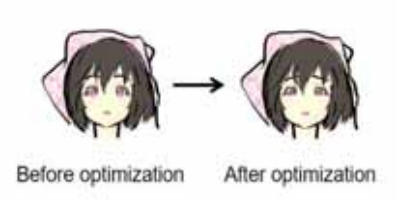

Fig. 19. The comparison of Miebo's "Sad" before and after optimization

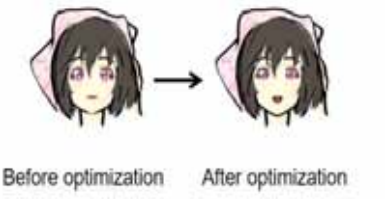

Fig. 20. The comparison of Miebo's "Happy" before and after optimization

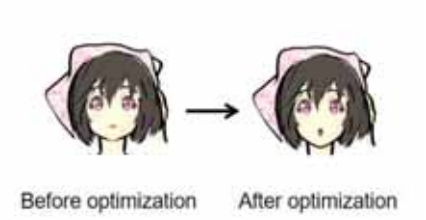

Fig. 21. The comparison of Miebo's "Surprise" before and after optimization

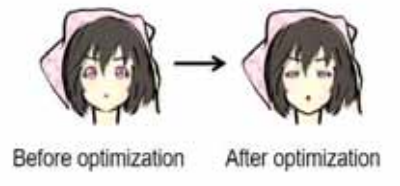

Fig. 22. The comparison of Miebo's "Angry" before and after optimization

D. Experimental optimization results

According to the data and results obtained in the experimental process, the parameters of Miebo under four emotions were adjusted, and a questionnaire was designed according to the emotional presentation before and after optimization. Teachers and students were invited to evaluate and score (10-point system). Finally, the scoring results of 55 respondents were obtained, as shown in the Fig.23 below.

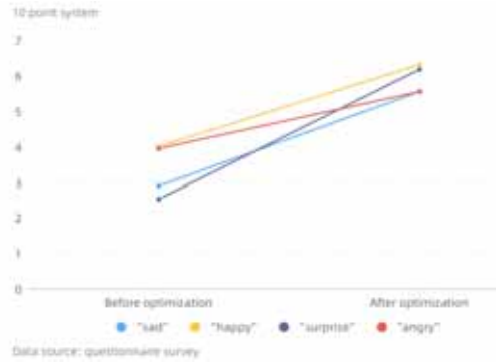

Fig. 23. The changes of evaluation scores before and after optimization experiment

It can be clearly demonstrated from Fig.23 that the emotion presented by Miebo after optimization can move the audience more than the emotion presented by forehead optimization, and the audience's praise for the optimized expression is much higher than that before optimization.

V. CONCLUSION

The White Paper on the Development of Virtual digital human in 2020 which was released by China Artificial Intelligence Alliance predicts that in the future, virtual digital human will gradually realize the integration, application and landing in multiple scenes and fields, showing a trend of changing interaction mode and less intelligence. In this paper, we develop a basic platform for the virtual avatar design, providing the Vtuber and enterprise with a high level of flexibility in designing iconic avatars. Although this platform is currently only utilized for designing virtual avatars for live streaming, it also could be utilized in the area of virtual conference communications to allow participants not only to hear voices but also to read facial expressions and body language through the online conference.

VI. ACKNOWLEDGEMENTS

This work is supported by the Fund for Key R & D Projects in Zhejiang Province: 2021C03137.

In addition, I sincerely thank Zhao Yue and Long LI of Zhejiang University of communications media engineering and information for his technical guidance for this research and Li Long of School of design and art for his image design guidance for this paper.

REFERENCES

[1] Ji-ping Li. "Research on the application of virtual anchor in Taobao live broadcast in the era of intelligent media." news communication. 11 (2021): 44-45. Doi: CNKI: Sun: ywcb. 0.2021-11-017

[2] Yu-juan Qian. "Will Weiya and Li Jiaqi be replaced when the virtual anchor comes?" Economic Observer 2020-09-07020, the company

[3] Zhen-lin Fan, and Feng Zhe-hui. "Research on the development status of China's virtual anchor industry." modern audio visual. 03 (2021): 44-49. Doi: CNKI: Sun: STSD. 0.2021-03-014

[4] Peng-yi Shen, Wan Demin, and Jinan Xu. "How human-computer interaction perception affects consumers' well-being in the context of online retail -- from the perspective of autonomy." Nankai management review. (: doi:

[5] Yi-cong Guan, and LV Xin. "Research on the representational emotional expression design of AI virtual anchor." media. 23 (2020): 35-37. Doi: CNKI: Sun: cmei. 0.2020-23-023

[6] Zi-yang Gong, Wen-jun Shen, and Fu-hao Li. "Application and related technologies of virtual anchor in various fields." electronic technology and software engineering. 08 (2021): 165-166. Doi: CNKI: Sun: dzru. 0.2021-08-068

[7] Jie Liu. "Women called in the name of gentlemen -- Female Aesthetics under male culture in Shi Shuo Xin Yu." Journal of Ludong University (PHILOSOPHY AND SOCIAL SCIENCES EDITION) 33.01 (2016): 40-45. Doi: CNKI: Sun: yszs. 0.2016-01-009

[8] "A female reporter with a" first love face. "Encyclopedia of middle school students. Z3 (2021): 4. Doi: CNKI: Sun: zxbk.0.2021-z3-004

[9] Xiang Gao, et al. "Real time facial expression migration method combining 3dmm and Gan." computer application and software 37.04 (2020): 119-126. Doi: CNKI: Sun: jyrj. 0.2020-04-021

[10] Chun-ping Liu, et al. "Expression real-time dynamic migration based on facial key points." modern computer (Professional Edition). 02 (2019): 41-45. Doi: CNKI: Sun: xdjs. 0.2019-02-011

[11] Guang-cai Song. Wireless inertial motion capture system integrating attitude and space-time trajectory. 2021. Dalian University of technology, MA thesis

[12] J.Cho.Crawling the web: Discovery and Maintenance of Large-scale Web Data [D]. L.A.: Stanford University, 2001.

[13] K.Yao Baoqi. Research on multi view motion capture method for virtual interactive system. 2021. Beijing University of Posts and telecommunications, MA thesis

[14] Haque M Imtiaz M H, Kwak S T,et al.A Lightweight Exoskeleton-Based Portable Gait Data Collection System[J].Sensors 2021 , 21(3):781-790.

[15] Fan Xia Owen. Research on the application strategy of virtual anchor in live e-commerce [J]. Modern marketing (Xueyuan Edition), 2021 (08): 62-63.