

Research Article

“Avatar to Person” (ATP) Virtual Human Social Ability Enhanced System for Disabled People

Zichun Guo ¹, Zihao Wang ², and Xueguang Jin ³

¹College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China

²Edinburgh College of Art, The University of Edinburgh, Edinburgh EH17, UK

³College of Cyberspace Security, Changzhou College of Information Technology, Changzhou 213164, China

Correspondence should be addressed to Xueguang Jin; jinxueguang@ccit.js.cn

Received 24 September 2021; Revised 29 October 2021; Accepted 13 November 2021; Published 3 December 2021

Academic Editor: Ming Yan

Copyright © 2021 Zichun Guo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to make communication more effective has been underlined unprecedentedly in the artificial intelligence (AI) era. Nowadays, with the improvement of affective computing and big data, people have generally adapted to construct social networks relying on social robots and smartphones. Although the technologies above have been widely discussed and used, researches on disabled people in the social field are still very limited. In particular, facial disabled people, deaf-mutes, and autistic patients are still meeting great difficulty when interacting with strangers using online video technology. This project creates a virtual human social system called “Avatar to Person” (ATP) based on artificial intelligence and three-dimensional (3D) simulation technology, with which disabled people can complete tasks such as “virtual face repair” and “simulated voice generation,” in order to conduct face-to-face video communication freely and confidently. The system has been proven effective in the enhancement of the sense of online social participation for people with disabilities through user tests. ATP is certain to be a unique area of inquiry and design for disabled people that is categorically different from other types of human-robot interaction.

1. Introduction

Communication is the main component of interpersonal relationships, which defines the existence of human beings and serves as a vital connection between individuals. It is also one of the most important bases for promoting the development of human civilization. Communication can be divided into verbal and nonverbal forms, and its essence is to exchange information between the receiver and the sender. As the components of communication, verbal and nonverbal forms are separately regarded as spontaneous and disguised spontaneous communication, the former manifested as the intentional communication from the motivation-emotional state, and the latter as the instinctive intentional strategic operation [1]. In general, people focus more on language but lack attention to nonverbal ways of communication which psychologists believe can effectively convey instinctive emotions and that can be an indirect result of the main and noncommunicative functions in the

communication process [2]. Verbal forms in this study refer to oral and written language, while nonverbal forms refer to facial expressions, body postures, and other bodily motions used to communicate with others.

Communication is a natural ability for most people. Nevertheless, congenital or acquired physical deformity people with impairments suffered from exacerbation of their communicative difficulties [3]. Psychological as well as physical problems make it more difficult for them to participate in societies, and so they are viewed as a distinct group [4]. Normalizing communication between impaired and healthy people is a tough process. Face-to-face communication allows people to have more intimate expressions during social interactions, but for most people with facial disabilities, this form of communication is unachievable [5].

Most of the previous studies focus on two aspects, the use of “social robots” and “smartphones” for communication. Robots were used to improve social skills and were generally used in education [6], autistic children [7], and elderly

care [8]. These researches mainly focus on the use of robots to complete the dialogue and interaction between humans and machines and then gradually improve the technical iterations of emotional computing and perception [9]. Although the interaction between humans and robots has been meaningfully explored with children, adults, and the elderly, the research on the interaction between people with disabilities and robots is still very limited [10]. This kind of interaction is not the same as normal cases when all the social interaction between humans and robots is conducted through emotional computing.

Apart from social robots, the popularity of smartphones has also greatly benefited people with disabilities, such as reducing their reliance on others, increasing their job competitiveness, encouraging them to participate more actively in public debates, and also building previously impossible communication channels for them and their peers [11]. In particular, we-media social media platforms have had a significant impact on the social and professional lives of individuals with disabilities since the product launch. The rise of blogs also provided disabled people with a new venue to share news, criticism, opinions, and personal experiences [12]. Disability culture on the internet grew in fruitful ways with the development of blogs. Moreover, disability is linked to both the built environment and the physical body [13]. Written language, devoid of emotional components, serves only as a regular way for exchanging information, while face-to-face video connection has become more emotive as we-media platforms have grown into varieties of social forms like live video streaming.

However, none of the social robots for social simulation nor self-media for text or voice communication could be accepted by certain groups of disabled people. Those with disfigurements and autism, for example, are less inclined to use the live video streaming platform to showcase their abilities like storytelling and singing. To hide their flaws, they prefer typing chat with strangers to face-to-face communication. Similarly, for deaf and dumb people, voice communication is also not available. In general, these specific groups include three categories of people: (1) individuals with facial injuries or peripheral facial palsy who are reluctant or afraid to reveal their handicapped faces in front of the camera [14]. The face is an essential aspect of humans' personality and body image. The psychological factor is one of the main factors that influence the healing and rehabilitation of face trauma [15]; (2) deaf-mutes lack sound judgment due to their congenital or acquired deafness and cannot pronounce correctly either; (3) people with cerebral palsy and autism suffering from brain lesions and psychological trauma which prevent their facial muscles from moving as they want are unable to convey information properly through facial expressions. The three categories of people are illustrated in Figure 1.

Based on the specific context, the model of virtual avatar social system called the Avatar to Person (ATP) is proposed as a video social means to bridge the gap between virtual people and actual people. Powered by artificial intelligence, the system is aimed at improving the experience for people with disabilities in online video meetings. All the communi-

cation ways, including text, voice, facial expressions, and body language, are supported by the avatar, while three main functions, including facial expression augmentation, text-to-facial expression, and facial repairment, are supported by the virtual avatar social system.

This paper will firstly describe present impediments to communication for disabled people; secondly discuss the technological possibility, user psychology, and user experience; and thirdly evaluate the system's possible future of real-time usage in ATP applications.

2. Materials and Methods

2.1. Research and Design across Disciplines. Communication, artificial intelligence, and animation design are integrated in the process of the model creation and development of the ATP model. For example, owing to the rapid development of deep learning technology and its outstanding performance in the field of artificial intelligence, facial restoration technology has made tremendous progress in feature recognition, mapping, and modelling of pain intensity from face pictures [16]; affective computing is a collection of approaches for extracting effect from data in many modalities and at various granularity scales, aimed at developing a system that can recognize, interpret, process, and simulate human emotions and provides the possibility of emotion design of virtual avatars [17]; and animation technology could be utilized to build virtual avatars with different styles to satisfy various requirements of people with disabilities.

In this project, the model design of ATP is based on the theoretical context of the traditional Person to Person (PTP) transmission mechanism put forward by Sudweeks et al., which can be divided into three ways: one-way communication, two-way communication, and interactive communication (as shown in Figure 2) [18].

The ATP model, as shown in Figure 3, refers to the social interaction between a real human and a virtual avatar, in which users may transform their physical appearance into a virtualized avatar.

2.2. Content Design. The project focuses on the three categories of handicapped individuals with difficulty in face-to-face online communication. They are people with autism, deaf and dumb people, and people with facial disabilities.

2.2.1. Disabled People with Autism. Autism often occurs at a young age, and the symptoms mainly include communication disorders, language communication disorders, and stereotyped repetitive behaviours [19]. Compared with normal children, autistic children are always more flat or neutral in their expression, which implies the difficulties they have in sharing their affection and making their affective signals to be understood [20]. The current *Diagnostic and Statistical Manual of Mental Disorders (DSM)* diagnostic criteria for autism have included items concerned with problems in perceiving and processing emotions: "marked impairments in the use of multiple nonverbal behaviours, such as ... facial expression ..." and "lack social or emotional reciprocity" [21].



FIGURE 1: “Certain handicapped people” may not be willing to use a social robot to perform a social simulation or use self-media for voice communication.

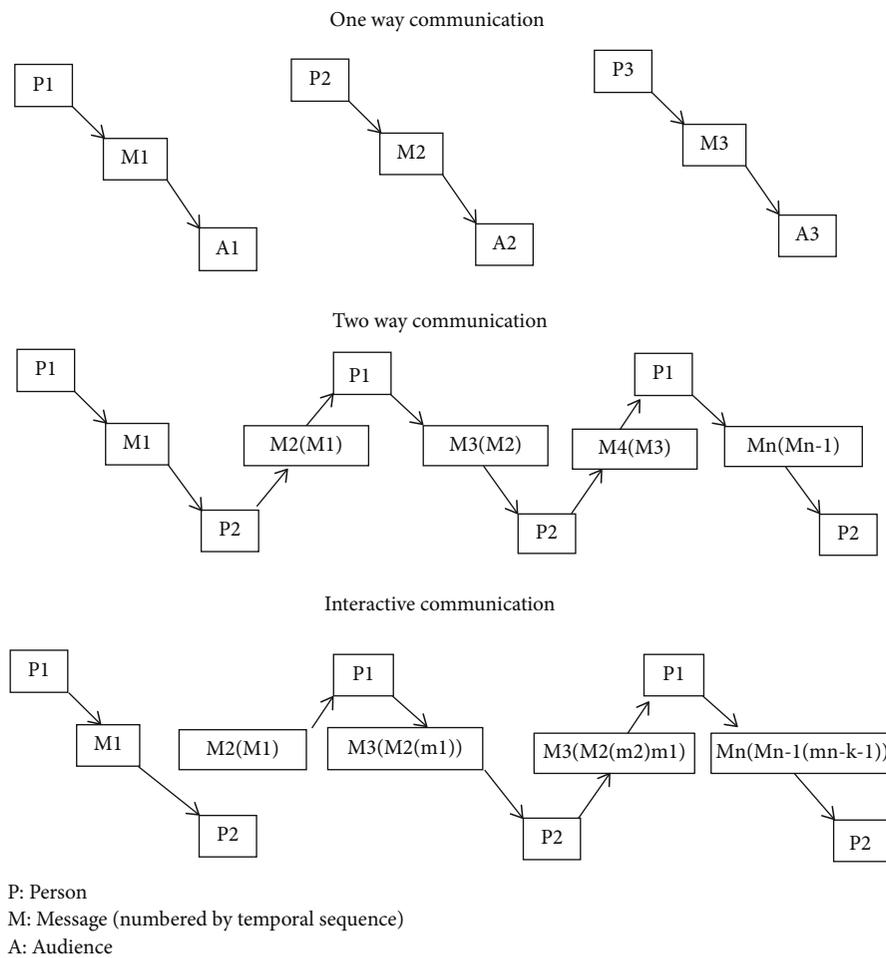


FIGURE 2: The Person to Person (PTP) transmission mechanism.

The capacity to identify emotions from facial expressions and then to give proper feedback is such an essential tool in successful social interaction [22]. The inability to recognize facial expressions hinders the social interaction of people with autism. Therefore, it is necessary to take some measures to interfere in autistic children’s facial expression recognition, and the way to augment patients’ poor facial expres-

sions has become one of the most urgent problems that need to be solved by virtual avatars.

Owing to the emergence of artificial intelligence, autistic individuals may get a helping hand. Face recognition and sentiment analysis powered by artificial intelligence can augment users’ facial expressions in online face-to-face meetings, helping autistic individuals recognize and

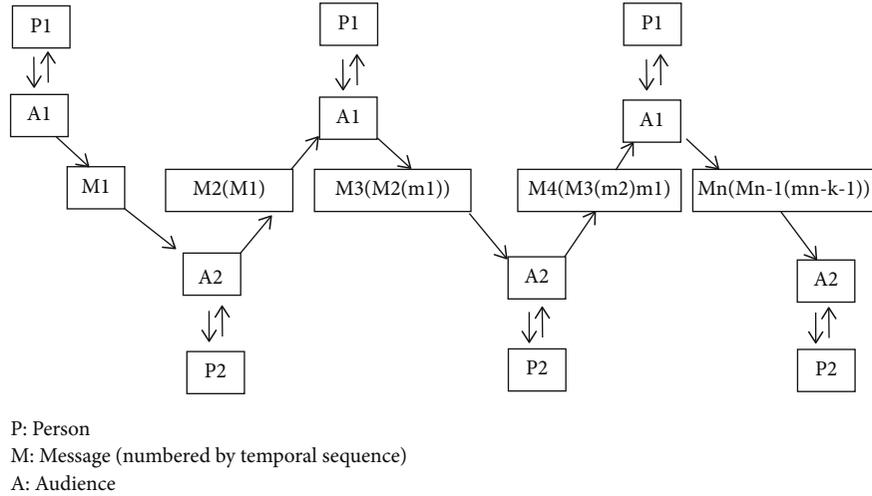


FIGURE 3: The PTP communication pattern.

comprehend the facial expression of the person speaking to them. The desired result can be accomplished in two ways. One is facial expression recognition augmentation, which applies AI technology to the process of facial expression recognition of autistic people and then uses affective computing to extract expression labels to distinguish their facial expressions, while the other is expression intervention augmentation, which refers to performing emotion recognition by extracting autistic patients' weak physical expressions and then augmenting them in the virtual humanization process.

2.2.2. Disabled People with Deafness. Although deaf and dumb people can communicate with writing, they cannot still mingle with the social world due to the lack of face-to-face communication abilities. There are more than 450 million deaf and dumb people all over the world, which takes up 5% of the world's population. Despite the great number, there is quite little research to bridge the communication barrier between deaf and dumb people and the social world [23]. Therefore, an efficient system must be set up to help them communicate and interact freely.

As for this project, virtual avatars are introduced to help deaf and dumb people overcome the difficulty of pronunciation through text-to-speech and facial expression animation. The mouthpiece translation technology is applied to help deaf and dumb individuals improve their capacity to express themselves clearly. The text input is converted into the information in voice and mouth type, and the user's facial expressions are complemented with the avatar. Furthermore, for congenital deaf-mutes, the system can provide AI voices with various voice styles, while for acquired deaf-mutes, the system will conduct AI training on users' previous voice samples to simulate and generate their inborn voice.

2.2.3. People with Facial Disabilities. Facial disability results in deficits that are necessary for both physical functions like eating and drinking and social functions and mental health like conveying conversational signals and intimate human information [24]. Macgregor emphasized that social interac-

tion during brief encounters in daily life causes people with facial disabilities continuous stress, affliction, and anxiety [25]. Facial disability makes them feel strange and fearful when showing their faces and leads to their inner feelings of inferiority, thus causing problems like social isolation and addiction [26].

Given the psychological barrier that people with facial impairments experience, we propose virtual avatars to restore their faces during online video communications, in which way their willingness to engage themselves in online meetings with strangers would be increased. In order to create virtual avatars, a facial repairing system, an intervention technique for facially disabled people, could be used. The 3D avatar is built by AI based on a database of millions of thousands of samples of predisabled pictures. Users can then use the real-time face capture technology to synchronize their facial expressions with those of the 3D virtual avatars in the video communication process, in which way the barriers between people with facial disabilities and the social world can be bridged.

2.3. System Design

2.3.1. System Overview. Figure 4 depicts the framework of the proposed system. Firstly, the system creates high-fidelity animation to reflect the emotion and look of handicapped individuals properly. Secondly, the system retargets personal characters through various speech styles and animation synced with the vocal input, in which way the AI voice output is shown in accordance with the emotional expression. Finally, the system renders quickly, with the goal of realizing real-time animation.

2.3.2. Mental Cognitive System. The mental cognitive system model used in the paper was designed by Baron-Cohen [27]. 412 emotions divided into 24 emotional categories and 6 developmental levels (from 4 to maturity) are considered. In the phase of the face detection stage, facial expression

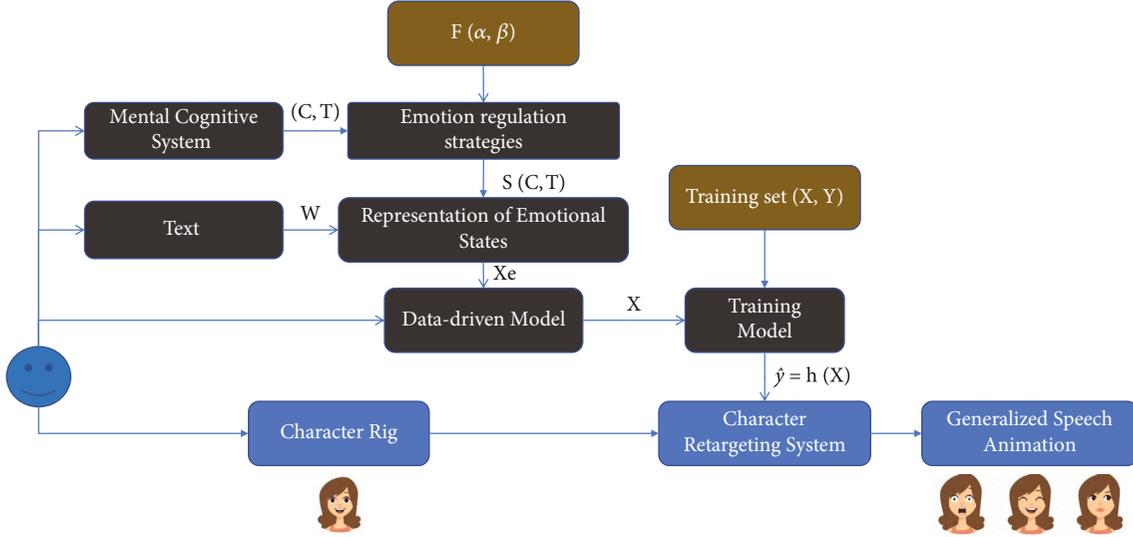


FIGURE 4: The framework of the ATP social system.

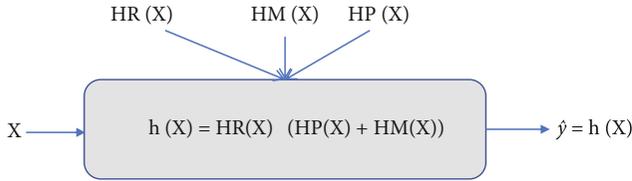


FIGURE 5: The training model.

classification C and recognition template T are taken into account.

2.3.3. Emotion Regulation Strategy. Emotion regulation is an influential factor in many problems that blind people suffer from [28]. Interventions that target emotion regulation strategies would be useful [29]. For blind people, cognitive reappraisal and expressive suppression are less applied. Emotion regulation strategy distribution function $F(\alpha, \beta)$ is defined as

$$N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho), \quad (1)$$

where μ_1 and σ_1^2 are the normal distribution factors of cognitive reappraisal and μ_2 and σ_2^2 are the normal distribution factors of expressive suppression. These parameters can be adjusted by the system.

For emotion regulation response generation, given a query $q(C, T)$, an emotion category $e(C, T)$, and emotion regulation strategy $F(\alpha, \beta)$, the goal is to generate a response $s(C, T)$ that is not only meaningful but also in accordance with the desired normal people emotion.

The Emotional Chatting Machine (ECM) addresses the emotion factor with three new mechanisms: Emotion Category Embedding, Internal Memory, and External Memory [30]. Specifically, (1) Emotion Category Embedding models the high-level abstraction of emotion expression by embedding emotion categories and concatenates the corresponding embeddings to the input at each decoding step. (2) Internal

Memory captures the change of implicit internal emotion states with reading and writing gates. (3) External Memory applies an external emotion vocabulary to express emotions explicitly and finally assigns different generation probabilities to emotion and generic words.

The training sample (q, s) is defined as

$$q(C, T) = (q(C_1, T_1, q(C_2, T_2) \cdots q(C_m, T_m)), s(C, T)) \quad (2) \\ = (s(C_1, T_1, q(C_2, T_2) \cdots s(C_m, T_m))).$$

The emotion regulation strategy distribution function $F(\alpha, \beta)$ is calculated through a training sample.

The loss function of the training sample (q, s) is defined as

$$-\sum_{t=1}^m p_t \log(o_t) - \sum_{t=1}^m q_t \log(\alpha_t) + \left\| M_{s(C,T), F(\alpha, \beta), m}^I \right\|, \quad (3)$$

where o_t and p_t are the predicted token distribution and gold distribution, α_t is the probability of choosing an emotion word or a generic word, $q_t \in \{0, 1\}$ is the true choice between them in $s(C, T)$, and $M_{s(C,T), F(\alpha, \beta), m}^I$ is the internal emotion state at the last step m with emotion regulation strategy distribution function $F(\alpha, \beta)$. The first term is the cross-entropy loss, the second is used to supervise the probability of selecting an emotion or a generic word, and the last is used to ensure that the internal emotional state has been expressed completely once the generation is finished.

2.3.4. Emotional State Representation. Auxiliary emotional data derived from facial, vocal, textual, or mixed data is fed into the deep learning network and training samples, along with a tiny quantity of supplementary emotional data. The deep learning network has sufficient data to infer the optimal result. All relevant aspects of the animation should be encoded following the probability distribution of the given training samples, in accordance with the additional

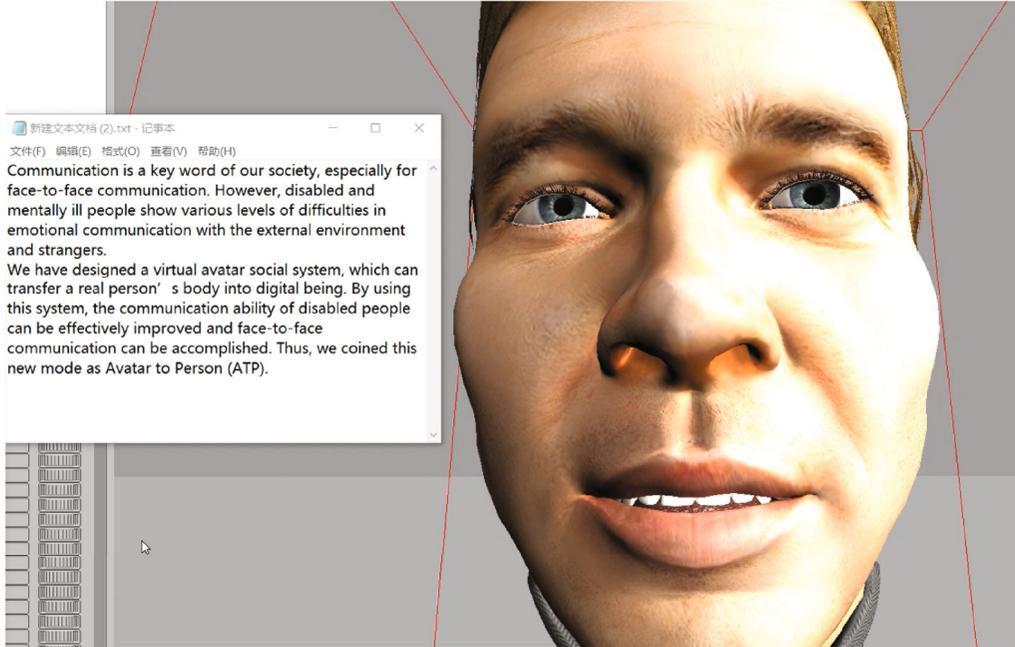


FIGURE 6: Output facial expressions corresponding to the deaf people's inputting text to the ATP system.

TABLE 1: Lessons for experiment.

Lesson	Lesson components	Reciprocity components
Having a respectful attitude	(i) Talk about feelings when upset (ii) Use request words and tone rather than demand words (iii) Compromise or accommodate others (iv) Ask permission to touch something not yours	Request Compromise
Getting to know someone new	(i) Introduce yourself (ii) Talking about some interesting thing (iii) Asking question	Ask Respond Share information Introduce oneself
Compromising	(i) Ask what another person wants (ii) Tell what you want (iii) Offer to do some of what you and another person want	Ask Share information

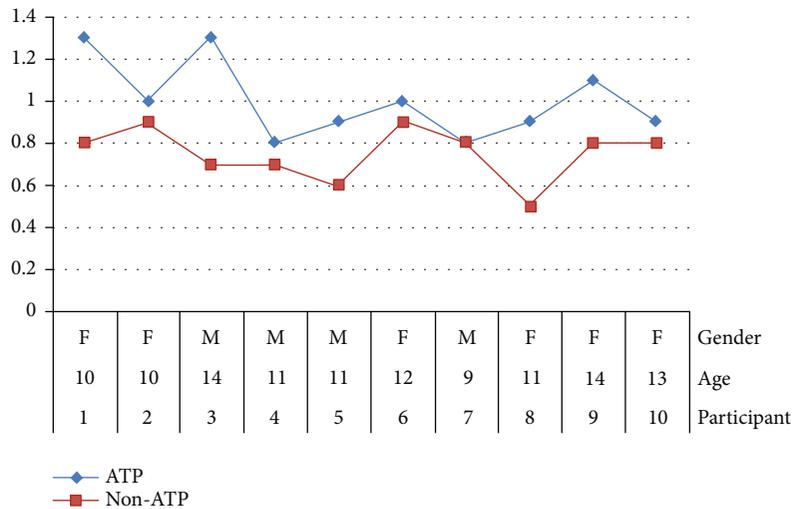


FIGURE 7: Data for ATP and non-ATP.

TABLE 2: Test result for deaf-mutes.

Participant	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
ATP	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
Non-ATP	F	S	F	S	S	F	F	F	F	S	S	F	F	F	F	F	S	F

S means information transfer completed successfully; F means information transfer failure.

emotional data. These training examples are based on a variety of variables, including facial expressions, speaking styles, and collaborative pronunciation patterns, rather than the audio itself. As a result, the emotional state of the actor is sampled. On a given vocal cord, auxiliary emotional data input is highly beneficial for reasoning to blend and match different emotional states.

Commercial software such as Baidu AI, manual tagging, and training sample classification can be applied to achieve the emotional state of disabled people [31].

2.3.5. Data-Driven Model. In the data-driven model, some deep neural networks (the grey part of the system) are modelled for learning effectively from appropriate training sets with a large variety of speech patterns. The input audio, facial, and test data are trained. It builds the formant analysis network to extract feature sequences and emotional expression labels at the same time. Thus, short-term features related to the disabled people's facial animation are extracted by the convolution layer, and the features include emphasis, specific phonemes, and intonation. And the temporal evolution of facial features is generated by another convolutional neural network. There is an audio window related to the abstract feature vectors of facial posture. Ambiguities between different facial expressions and speech styles are eliminated by the emotional expression as another input. E-dimensional vectors are used to express emotional states, which are directly connected to the output of each layer in the neural network so that the subsequent layers can change their behaviour accordingly. With the model of Active Appearance Model (AAM), the mouth features are shaped with 34 component vertices and face features are shaped with 108 vertices. Most importantly, the learning network executes the same operation on every point on the same timeline. As a result, the same training samples with different timeline offsets are generated.

With small learning faces, the specific effect of the speaker is controlled, and the details of the captured animation are focused. Cost-effective collection of the appropriate training set captures a wide range of speech patterns. This method meets the request of retarget ability and edit ability.

2.3.6. System of Character Retargeting. The speech animation with specific style settings for disabled people is generated by the character retargeting system (blue part of the system), and X denotes a sequence of inputs of the animation. The true prediction animation sequence Y is constructed by a predictor that predicts any input X . Y is corresponding to the specific reference facial model. The training set (X, Y) is collected from the reference facial data set. In general, h is nonlinear which is learned by the deep learning network

to obtain complex nonlinear mappings from X to Y . After getting the mapping h , with the retargeting function, the parameters obtained from the reference face model can be performed on any parameterization precalculation character model. Thus, automatic and quick retargeting can be done.

The operation process of the data is as follows: (1) Phoneme sequence X is generated by input audio (e.g., speech recognition software). (2) $h(X)$ is used to predict the animation parameter Y corresponding to the reference mouth and face model associated with input X . (3) Y in the reference animation model is retargeted to the target character model. Thus, the accompanying audio speech animation in real time is generated by the model. The Hidden Markov Model Toolkit (HTK) or artificial translating is used for speech-to-text translating.

2.3.7. Data Training Model. The loss function is defined to optimize and train the deep learning network as is shown in Figure 5 [32]. It has three different constraints. The position function ensures that each output vertex position is roughly correct. The motion function ensures that the vertex follows the correct motion type in the animation. The regularization function ensures that the emotion short-term changes are well controlled according to the correct animation position and direction of motion.

For a given training sample X , we define the loss function $h(X)$ as

$$\begin{aligned}
 h(X) &= HR(X)(HP(X) + HM(X)), \\
 HP(X) + HM(X) &= \frac{1}{3D} \sum_{i=1}^{3D} \left((Y^{(i)}(X) - Y^{\wedge(i)}(X))^2 \right. \\
 &\quad \left. + \left(M[Y^{(i)}(X) - M[Y^{\wedge(i)}(X) - M[Y^{\wedge(i)}(X)]] \right)^2 \right), \\
 HR(X) &= \frac{1}{E} \sum_{i=1}^E \left(M|E^{(i)}(X)| \right)^2. \quad (4)
 \end{aligned}$$

$HP(X)$ represents the position factor, and $HM(X)$ represents the motion factor. $Y(i)$ represents the i th scalar component of $Y = (Y(1), Y(2), \dots, Y(3V))$, in which V is the total number of output vertices related to the character face. Thus, for each vertex, the total number of the character's output components is $3V$ with the 3D position.

$HP(X)$ ensures that the deep learning network output is fully synchronized without promoting the quality of animations. To promote the quality of animation, it was optimized from the perspective of vertex motion. Therefore, the motion item $HM(X)$ was quoted to qualify the difference between frames in the operation M [33]. In the simulation

procedure, the data driving model network correctly reflects the emotion expected and the regularization term ensures $HR(X)$ is running at the same finite differencing operator as above.

The Adam optimization method was adopted to balance the previous three-loss terms [34]. The weight of the network is updated by Adam with the gradient of the loss function and is normalized based on the long-term estimation. Normalization applies to the overall loss function, but not to individual terms. Output facial expression corresponding to the deaf people's inputting text to the ATP system is shown in Figure 6.

3. Results and Discussion

All three groups of impaired people, disabled people with autism, deaf and facial abilities should have participated in the ATP intervention study. Unfortunately, only autistic individuals and deaf-mutes are currently willing to participate in the experiment. We met great difficulty in conducting user testing among people with facial trauma. In this case, the results are confined to evaluating the efficacy of ATP in groups of autistic individuals and deaf-mutes.

3.1. Assessment Design for Autism. The metrical model of the experiment is based on the structured learning program proposed by Baker and Myles in the book [35] *Social Skills Training for Children and Adolescents with Asperger Syndrome and Social Communication Problems*, which includes 70 specific social skills that are most likely to cause autism spectrum disorders and communication problems. The metrical model is on the basis of the three lessons in the book from the communication aspect (see Table 1).

3.1.1. Participants. Ten autistic individuals (mean age = 11.5; SD = 2.65; ranging from 9-14), composed of 6 females and 4 males, were invited to test the effectiveness of ATP on disabled people with autism. The physical age of all ten participants had been confirmed that is matched for verbal mental age.

3.1.2. Procedure. 30 questions are designed to evaluate three lessons, each of which contains ten social issues. The scores for these questions represent the level of communication skills. A full score is three points ($a = 3$), while a score of 1.8 is considered typical.

For the test, 10 participants, ranging in age from 9 to 14, are invited. Each participant was tested four times, and each time lasts 30 minutes. It took four days to complete the entire test. For each test, the participant was asked to communicate with strangers through real-time video, and all the tests were performed by the same speech-language pathologist in connection with customary medical visits to the clinic.

The experiment set both an experiment group and a no-treatment group. During the four-day testing process, 10 participants needed to complete three Baker and Myles tests without ATP in the first two days, as the no-treatment group, and then in the next two days, they were asked to complete the same test under the intervention of ATP, as

the experiment group. The result of the experiment is shown in Figure 7.

According to the result, the scores of the ATP group are significantly higher than those of the non-ATP group, which indicates that the ATP system can effectively improve the video social ability of autistic children.

3.2. Evaluation Design for Deaf-Mutes. As for the test with deaf and dumb people, 18 participants were required to convey specific information, the meaning of "panda," to normal people through video social networking software. The whole communication process was only allowed to conduct on the video software. Written language was prohibited.

The same as the form of tests on autistic individuals, two groups were set for comparison. In the ATP group, participants entered text into the ATP system, and the avatar generated a spoken animation based on the text and presented it to the recipient; in the non-ATP group, participants were free to transmit nontext information in any way they wanted, such as body language and facial expressions. Table 2 displays the outcome.

The test results show that all participants in the ATP group can successfully communicate information through online video socialization, while few participants in the non-ATP group succeeded in the transmission of the information, which can verify the effectiveness of ATP.

4. Conclusions

Most of the previous studies about social skills assisting communication always focus on children and ageing people with "social robots" and "smartphones," while studies on disabled people have just started to attract the attention of researchers. There is currently a lack of longitudinal research on developing online social tools for disabled people with ATP due to the nonexistence of flexible and effective ATP providing consistent and regular communication. On the contrary, this study showed that ATP would be beneficial for the improvement of online social skills in autistics and deaf-mutes. The ATP social system provides a platform for autistic individuals, deaf and dumb people, and people with facial disabilities, to help them realize free expression and evaluation through online video socialization, and has been verified effective in groups of autistic individuals and deaf-mutes but needs further tests on disabled groups with facial disabilities.

It is the first step for us in applying digital being research to work for impaired individuals. In the future, based on the existing ATP social system, we will continue our journey to develop frameworks including (1) avatar for people social system design strategies in the community and (2) user-generated content for handicapped individuals by ATP. Furthermore, we will go on the exploration of future trends and research areas in public health systems.

Data Availability

The data and model used to support the findings of this study are restricted by the College of Information Science

and Technology of Beijing University of Chemical Technology. Data are available from the corresponding author for researchers who meet the criteria for access to confidential data.

Acknowledgments

This research was supported by the National Science Foundation for Postdoctoral Scientists of China (Grant No. 2021M700355), by the Changzhou Sci&Tech Program (Grant No. CE20212025), by the General Project of Higher Education Reform Research in Jiangsu Province (Grant No. 2021SJG521), by the Innovation Application Centre of Changzhou College of Information Technology (Grant No. CCIT2021STIT010202), by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (Grant No. 20KJB140013), and by the Natural Science Foundation of Jiangsu Province, youth project (Grant No. BK20200190).

Disclosure

This research is based on our previous teamwork, *Avatar Social System Improve Perceptions of Disabled People's Social Ability*, presented in 2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS). Rui Hao has also made contribution to the research.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

References

- [1] R. Buck and C. VanLear, "Verbal and nonverbal communication: distinguishing symbolic, spontaneous, and pseudo-spontaneous nonverbal behavior," *Journal of Communication*, vol. 52, no. 3, pp. 522–541, 2002.
- [2] R. Zajonc, S. Murphy, and M. Inglehart, "Feeling and facial effluence: implications of the vascular theory of emotion," *Psychological Review*, vol. 96, no. 3, pp. 395–416, 1989.
- [3] I. K. Zola, "Communication barriers between 'the able-bodied' and 'the handicapped'," *Archives of Physical Medicine and Rehabilitation*, vol. 62, no. 8, pp. 355–359, 1981.
- [4] A. Sheldon, *Disabled people and communication systems in the twenty first century*, [Ph.D. thesis], The University of Leeds, 2001.
- [5] M. Argyle, M. Cook, and D. Cramer, "Gaze and mutual gaze," *British Journal of Psychiatry*, vol. 165, no. 6, pp. 848–850, 1994.
- [6] A. Sharkey, "Should we welcome robot teachers?," *Ethics and Information Technology*, vol. 18, no. 4, pp. 283–297, 2016.
- [7] S. Costa, F. Soares, A. Pereira, C. Santos, and A. Hiolle, "A pilot study using imitation and storytelling scenarios as activities for labelling emotions by children with autism using a humanoid robot," in *4th International Conference on Development and Learning and on Epigenetic Robotics*, pp. 299–304, Genoa, Italy, 2014.
- [8] M. Coeckelbergh, "'How I learned to love the robot': capabilities, information technologies, and elderly care," in *The capability approach, technology and design*, pp. 77–86, Springer, Dordrecht, 2012.
- [9] J. Peter, R. Kühne, A. Barco, C. de Jong, and C. van Straten, "Asking today the crucial questions of tomorrow: social robots and the internet of toys," in *The Internet of Toys*, pp. 25–46, Palgrave Macmillan, Cham, 2019.
- [10] E. Björling, E. Rose, A. Davidson, R. Ren, and D. Wong, "Can we keep him forever? Teens' engagement and desire for emotional connection with a social robot," *International Journal of Social Robotics*, vol. 12, no. 1, pp. 65–77, 2020.
- [11] A. Martinez-Cabrera, "Web more accessible to those with disabilities," *San Francisco Chronicle*, vol. 1, 2010.
- [12] R. Anderson, B. Stenger, V. Wan, and R. Cipolla, "Expressive visual text-to-speech using active appearance models," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3382–3389, Portland, USA, 2013.
- [13] J. Hunsinger and T. Senft, *The Social Media Handbook*, Routledge, 2013.
- [14] T. Movérare, A. Lohmander, M. Hultcrantz, and L. Sjögreen, "Peripheral facial palsy: speech, communication and oral motor function," *European Annals of Otorhinolaryngology, Head and Neck Diseases*, vol. 134, no. 1, pp. 27–31, 2017.
- [15] A. De Sousa, "Psychological issues in acquired facial trauma," *Indian Journal of Plastic Surgery*, vol. 43, no. 2, p. 200, 2010.
- [16] G. Bargshady, X. Zhou, R. Deo, J. Soar, F. Whittaker, and H. Wang, "Enhanced deep learning algorithm development to detect pain intensity from facial expression images," *Expert Systems with Applications*, vol. 149, article 113305, 2020.
- [17] E. Yadegaridehkordi, N. Noor, M. Ayub, H. Affal, and N. Hussin, "Affective computing in education: a systematic review and future research," *Computers & Education*, vol. 142, article 103649, 2019.
- [18] F. Sudweeks, M. McLaughlin, and S. Rafaeli, *Network and Net-play*, AAAI Press, Menlo Park, CA, 1998.
- [19] S. Rogers, "Interventions that facilitate socialization in children with autism," *Journal of Autism and Developmental Disorders*, vol. 30, no. 5, pp. 399–409, 2000.
- [20] N. Yirmiya, C. Kasari, M. Sigman, and P. Mundy, "Facial expressions of affect in autistic, mentally retarded and normal children," *Journal of Child Psychology and Psychiatry*, vol. 30, no. 5, pp. 725–735, 1989.
- [21] American Psychiatric Association, *Diagnostic and statistical manual of mental disorders*, American Psychiatric Association, Washington, DC, 1980.
- [22] M. Harms, A. Martin, and G. Wallace, "Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies," *Neuropsychology Review*, vol. 20, no. 3, pp. 290–322, 2010.
- [23] P. Ahire, K. Tilekar, T. Jawake, and P. Warale, "Two way communicator between deaf and dumb people and normal people," in *2015 International Conference on Computing Communication Control and Automation*, pp. 641–644, Pune, India, 2015.
- [24] J. VanSwearingen and J. Brach, "The facial disability index: reliability and validity of a disability assessment instrument for disorders of the facial neuromuscular system," *Physical Therapy*, vol. 76, no. 12, pp. 1288–1298, 1996.
- [25] F. Cooke Macgregor, "Facial disfigurement: problems and management of social interaction and implications for mental health," *Aesthetic Plastic Surgery*, vol. 14, no. 1, pp. 249–257, 1990.

- [26] J. Valls-Solé, “Electrodiagnostic studies of the facial nerve in peripheral facial palsy and hemifacial spasm,” *Muscle & Nerve*, vol. 36, no. 1, pp. 14–20, 2007.
- [27] S. Baron-Cohen, H. Ring, J. Moriarty, B. Schmitz, D. Costa, and P. Ell, “Recognition of mental state terms,” *British Journal of Psychiatry*, vol. 165, no. 5, pp. 640–649, 1994.
- [28] H. Cai, H. Chen, C. Zhang et al., “Learning from easy to complex: adaptive multi-curricula learning for neural dialogue generation,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 5, pp. 7472–7479, 2020.
- [29] S. Salimi, E. Mohammadi, and A. Sadeghi, “A comparison of emotion regulation strategies of blind students with sighted students,” *Iranian Rehabilitation Journal*, vol. 14, no. 2, pp. 127–132, 2016.
- [30] X. Zhou and W. Wang, “MojTalk: generating emotional responses at scale,” 2018, <https://arxiv.org/abs/1711.04090>.
- [31] K. Ellis and G. Goggin, “Disability media participation: opportunities, obstacles and politics,” *Media International Australia*, vol. 154, no. 1, pp. 78–88, 2015.
- [32] M. Yan, S. Li, C. A. Chan, Y. Shen, and Y. Yu, “Mobility prediction using a weighted Markov model based on mobile user classification,” *Sensors*, vol. 21, no. 5, p. 1740, 2021.
- [33] M. Yan, H. Yuan, Z. Li, Q. Lin, and J. Li, “Energy savings of wireless communication networks based on mobile user environmental prediction,” *Journal of Environmental Protection and Ecology*, vol. 22, no. 1, pp. 206–217, 2021.
- [34] M. Yan, H. Yuan, J. Xu, Y. Yu, and L. Jin, “Task allocation and route planning of multiple UAVs in a marine environment based on an improved particle swarm optimization algorithm,” *EURASIP Journal on Advances in Signal Processing*, vol. 2021, Article ID 94, 2021.
- [35] J. Baker and B. Myles, *Social Skills Training for Children and Adolescents with Asperger Syndrome and Social-Communication Problems*, Autism Asperger Publishing Company, Shawnee Mission, KA, 2004.