IASDR 2019
International Association of Societies of Design Research Conference 2019
DESIGN REVOLUTIONS
Manchester School of Art
Manchester Metropolitan University
02-05 September 2019

Research on remote learning in multimodal interaction

Zhiyong Fu[a]*; Yuyao Zhou[b]

[a b] Department of Information Art and Design, Tsinghua University, Beijing
* fuzhiyong@tsinghua.edu.cn

With the rapid development of the Internet, many new human-computer interaction methods have emerged in the education field to meet the increasing learning needs of people. This paper studies the teaching interaction of the learning maker course in the remote learning scenario, mainly observes the effectiveness of multimodal interaction in the process of distance learning and the changes in body characteristics (language, gestures and feedback). The experimental research allows the teachers and students to complete the hardware programming learning task. We set up three kinds of learning scenes that face-to-face and remote learning, and gradually increase the interactive mode to conduct experimental comparisons. According to the video, we explore the teacher-student interactions that occur in the real-time scenes, and investigate the learning effects through user interviews. The results show that multimodal interaction in the remote scene is beneficial to improve the efficiency of learning, and the picture presented is more three-dimensional, which promotes students' understanding of hands-on operation, and the sense of interaction between the two sides is more active. Finally, we analyse and summarize the multimodal interaction model of remote learning and its implications for design.

Keywords: Remote learning; Multimodal interaction; Maker Education

1 Introduction

Education innovation is the top priority of national development. The 2017 Horizon report [1] states that the educational technologies for mainstream applications in the next year are: maker space and robots; analytics technology and virtual reality in the next 2-3 years; artificial intelligence and the Internet of Things in the next 4-5 years. It shows that the country is paying more and more attention to the diversification of educational forms, the technology of educational equipment and students' practical ability. Different from the traditional single learning mode, the Maker classroom enables students to experience curriculum design and engineering more intuitively, integrate theory with practice, cultivate students' innovative consciousness and innovation ability, and stimulate learning enthusiasm.

Most people will choose offline institutions to learn Maker education. Face-to-face teaching is a traditional teaching mode, which is based on the teacher's lecture. The teacher demonstrates a point of view through explanations and examples until the students understand it. Remote learning is a cross-time, cross-regional, real-time or non-real-time interactive teaching format that is conducted online between teachers and students. With the increasingly close connection between people and computers, the human-computer

interaction approach faces new challenges. From the traditional mouse and keyboard interaction methods to the current touch screen, gesture, voice and other interactive modes, the multimodal human-computer interaction technology that combines various interaction modes has been widely developed. At present, the medium of remote learning is to use mobile phones or computers for mobile learning to meet the needs of learners to obtain learning anytime and anywhere. It is characterized by openness, flexibility, sharing and management. However, there are still many problems in Online Learning, for example:

1. Students have a single learning model and do not have a variety of forms such as group collaboration, inquiry learning or heuristic learning.
2. Students have very few ways to get knowledge, only to hear each other's voice and get information from the screen.
3. The interaction is simple. The two sides mainly interact through sound, without using other non-verbal movements of the body, and cannot interact and collaborate.
4. The indication is not clear. Online Learning is difficult to clearly point out your problems with your fingers like offline, and it takes time to explain the problem.
5. After learning, students don't know if they have achieved the effect of learning.
6. Teachers tend to ignore the emotional changes of students

Multimodal knowledge transmission refers to the transmission and learning of knowledge through various channels of communication, mobilizing the senses of learners in many aspects. This paper focuses on the impact of remote learning quality in multimodal mode and exploring the effectiveness of body dynamics (language, gestures and feedback) for knowledge transfer and learning, and further considers the help of multimodal interaction to remote learning, so as to provide inspiration for the product design of remote learning. We selected 8 groups of teachers and students to complete the hardware programming teaching tasks, set face-to-face and distance learning, and gradually increase the three learning scenarios for experimental comparison. According to the video shooting, we explore the teacher-student interactions that occur in the real-life scene of remote learning, and investigate the learning effect of the course through user interviews. We have found that the time and learning efficiency of using multimodal remote learning can reach a level consistent with face-to-face learning. Thus, we observed the knowledge transfer of eight groups of users through body language in three learning scenarios, focused on the analysis of the teaching interaction between teachers and students, and finally we summarized the multimodal interaction model of remote learning and its implications for design.

2 Related work

2.1 Multimodal interaction

Bill defines Interaction Design（Figure.1）as a threefold question: How do you do? How do you feel? How do you know? [2] Physics says that interaction is an encounter of two elements that transform each other mutually. Verplank's definition shows that the human feels the world, reflect upon it from what they know and act on the world accordingly. After the interaction, both human and world are not the same, so interaction can stimulate more collisions of inspiration and bring about self-improvement. The early human-computer interaction interface is Command-Line Interface (CLI), which can only support users to input information through the keyboard. Its operation form is simple, but it requires users to remember a large number of operation commands, which is very unfriendly to users. The

Graphical User Interface (GUI) has been improved both in aesthetics and operation. On the one hand, the GUI does not require users to memorize a lot of commands, on the other hand, it can show users more abundant visual and auditory information, and add mouse as a new interactive channel, but the user's hand is too heavy to operate. With the development of information technology, Multimodal systems can offer a flexible, efficient and usable environment allowing users to interact through input modalities, such as speech, handwriting, hand gesture and gaze, and to receive information by the system through output modalities, such as speech synthesis, smart graphics and other modalities, opportunely combined [3].

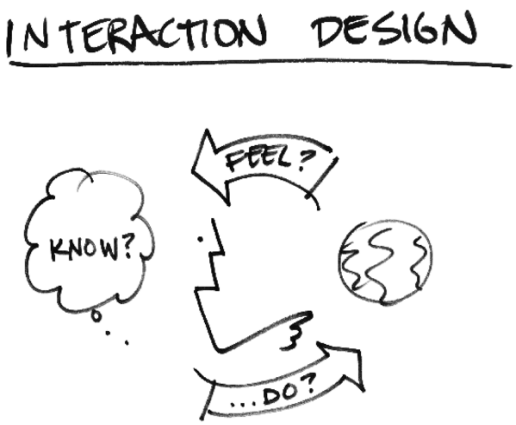


Figure 1. Bill Verplank's drawings about defining Interaction Design

2.2 Remote interactive learning

The traditional classroom learning is teacher-centered, and the teacher is the indoctrinator in the teaching process. The teaching medium mainly includes chalk, blackboard, model, etc. The students are mainly in a state of passive acceptance, and the teacher-to-student is a one-way interaction state [4]. The subjective status of students is neglected, which is not conducive to the creativity and divergent thinking of students. The increasing popularity of multimedia and network technology has begun to give students a leading position, no longer limited to traditional teaching materials, various resources are integrated and utilized, teaching content is enlarged, and learning mode appears mobile informal learning [5]. Therefore, interaction is an important factor in distance learning. The interaction in teaching can be understood as two-way communication between teachers and students. In 1989, Moore identified three types of interactions in distance education: interactions between learners and content, learners and teachers, learners and learners. Later, in 1994, Hillman et al. proposed the fourth type—the interaction between learners and interfaces [6]. The objects that learners interact with in these four kinds of interactions are content, teachers, other learners, and interfaces. So how to design the interaction between students and different objects is very important. In the process of remote learning, all kinds of knowledge and information are presented in a visual or auditory way. People receive and produce information mainly in five sensory spaces, namely vision, hearing, touch, smell and taste, of which the first three items account for 95.5% of the information. According to the research of educational psychology, the more senses involved in learning, the more neural connections between the outside world and the brain, and the better the effect of perception, understanding and memory [5].

Through literature research, we found that using Telepresence Robot [7,8] for remote learning, the participants' response is more efficient, the interaction is more positive, and the

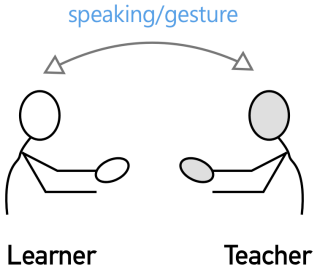
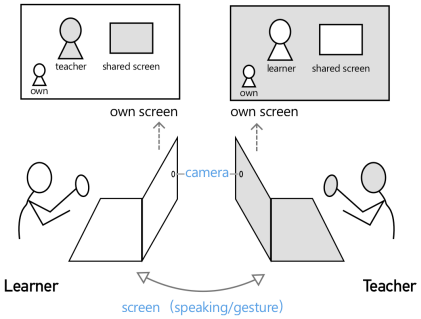
efficiency of the teacher's active teaching is correspondingly improved. The interaction of the Q&A forms increases the interaction of movement, gestures, object processing and so on. For the remote guidance of AR using head-wearing equipment, not only can the content be clearly positioned, but also can be immersed in teaching [9]. There are also studies that indicate that non-verbal behave is important, and visualization of multiple eye gaze can improve the interactive experience in multi-person remote learning. Depending on the learning task, eye movements and changes are very effective information for the teacher to observe the student's learning situation [10]. This is a form of instant feedback that allows the teacher to infer what the student is doing and understand their thinking process through the student's gaze path. If the student is looking for the wrong place, the teacher can also remind him in time. Group clustering makes it easy to manage multiple visualizations, and teachers can monitor the entire class while focusing on individuals [11].

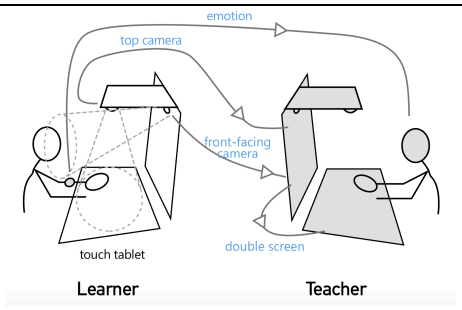
3 Experience

3.1 Overview

This paper focuses on the impact of remote learning quality in multimodal mode and exploring the effectiveness of body dynamics (language, gestures and emotions) for knowledge transfer and learning. We conducted two progressive analyses of the experiment. The first analysis explores the learning efficiency of remote multimodal interactive learning. The interactive channel modes of the three scenarios are gradually increasing, mainly including face-to-face learning, computer video remote learning and multimodal remote learning, as shown in table 1.

Table 1 System composition of three scenarios

Learning scenario	System composition	Interactive channels
Face-to-face learning		Talking, Gesture
Computer video remote learning (Dell)		Screen (voice, gesture), Keyboard, Mouse

Multimodal remote learning (HP Sprout + Empatica's E4 wristbands)		Double screen (voice, gesture, touch screen), Double camera (different viewing angles, 3d recognition), Keyboard, Mouse, Emotion recognition

Face-to-face learning means that teachers and students are close to each other in the same space for teaching. Computer video distance learning is the way of remote learning for most people. Dell computer is selected for the experiment to complete the remote learning. Multimodal remote learning uses HP Sprout and Empatica's E4 wristbands. HP Sprout reinvents learning, authoring, collaboration and sharing by integrating PCs, projectors, 2D and 3D scanners into a cost-effective, all-in-one solution. It has a camera at the top that can project the contents of the desktop, and a second screen on the desktop touch pad. During the experiment, students wear Empatica's E4 wristbands to monitor the skin's electrocortical activity during the learning process. The skin electrical signal is an important physiological signal that has been proven to contain reliable emotional information. According to Rich Voithofer, an associate professor of Education Science and technology at Ohio State University, wearable devices are a way of interacting with us and the world. Computers and phones are what we choose to interact with, while wearable devices automatically collect our data[12].In the experiment, we added the expression recognition channel to pass the students' learning emotions to the teacher, then we observed the teacher's reaction after understanding the students' emotions. The second analysis explored the body dynamics in the learning process, comparing the interaction characteristics of multiple channels under three learning scenarios. We observed the differences, learning efficiency, clarity and interaction of students in different scenarios, and we analysed the body dynamics at each stage.

3.2 Participants and design

We selected 16 participants to participate in the experiment, 10 women and 6 men. 16 people are divided into 8 groups, one group has teachers and students. The teacher's age is 28 years old and has 3-4 years of programming teaching experience. The average age of students (undergraduates and graduates) is 20 years, and they all have certain Arduino learning background. Teachers and students are randomly assigned to a group. Teachers can teach according to their own style, but before the experiment, we will remind the teacher to pay attention to the students' emotions during the teaching process.

The study was conducted using a between-subjects independent-measures design [13]. We designed one independent variable, three different learning scenarios. Each group of teachers and students completes three learning scenarios. The teaching contents are different under the three learning scenarios, but the difficulty is the same, ensuring that the learning difficulty does not affect the experiment duration. In the three scenarios, the teacher and the student each have a computer and a microduino kit. In the distance learning scenario, it is carried out in two different rooms through networking. During the experiment, we use the video recorder to record the experiment process. We first analyse the students'

learning efficiency and emotional changes through video observation, skin electricity data and questionnaires. Then we use video observation and user interviews to analyse the dynamic changes of teachers and students in three different learning scenarios.

3.3 Procedure

Before the experiment, according to the level of the students, the course contents of three learning scenarios were established: buzzer, steering gear and dot matrix screen. The learning difficulty is the same, and the learning time is about 15 minutes on average. Each group of teachers and students has the same course content and completes three learning scenarios. The learning process is divided into two parts: hardware teaching and software teaching. After experiment, the teachers and students need write the questionnaire. During the experiment, the camera records the time required to complete the hardware assembly and software operation. For the first analysis of the results, it is necessary to pre-process the electrical signals. In the multimodal remote learning, a total of 8 students' effective skin electrical signal data were collected. First, the original signal is sampled. The Nyquist sampling theorem proposes that when the sampling frequency is more than twice the maximum frequency of the original signal, the sampled signal can retain the effective information in the original signal without distortion, so the original signal at 1000 Hz is sampled down to 100 Hz. Subsequently, the sampled signal is denoised by wavelet transform threshold noise reduction. Finally, due to the large difference in the individual's basal skin electrical signals, the skin electrical signals between the different subjects were standardized to be comparable [14].

4 Results

4.1 The learning efficiency

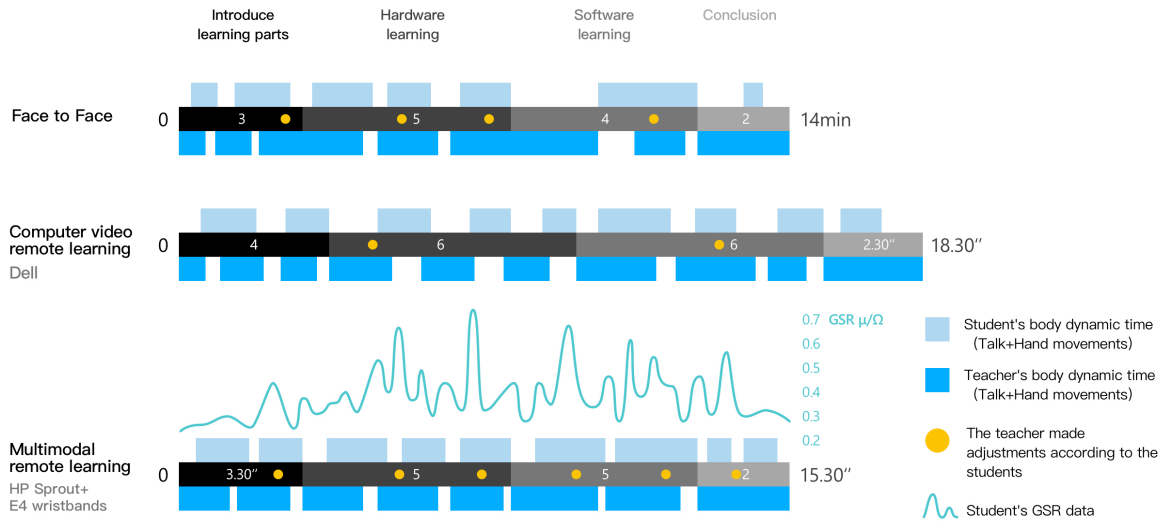


Figure 2. The average timeline of teacher-student learning in three scenarios

We use three indicators (time, learning gains, emotions) to judge learning efficiency. We calculated the average of the learning durations of the three groups of teachers and students, and Figure 2 shows the typical average timeline comparison. From the time point of view, face-to-face learning time is the shortest, multimodal learning is in the middle, and computer

video remote learning time is the longest. We choose the length of interaction between the instructional language and the gestures. The face-to-face interaction between teachers and students is 3:1. Most teachers are hands-on and speaking, and students are prone to dependence. The computer video remote learning teacher-student interaction time is 2:1. When multimodal learning, the teacher-student interaction time ratio is 1:1. The student will actively ask questions and find solutions to solve problems, so that the content of self-harvesting and understanding will be more. According to the questionnaire, students also have a higher understanding of learning when using multiple screens. From the perspective of emotion, face-to-face teachers observed students' learning problems on average 4 times and found that they did not understand the teaching knowledge. Computer video teaching, it is easy to ignore the changes in students' emotions. In multimodal teaching, since students wear E4 wristbands, we will give tips to teachers whenever students have emotional changes (up to 6 times), and teachers will take the initiative to care about students' problems. Students are more likely to have doubts during distance learning, and their tension can also be seen from the skin electricity data. Because remote learning is prone to problems that are difficult to express, it will increase the vigilance and concentration of students. Therefore, from the analysis of learning time, self-harvesting and learning emotions, multimodal interactive learning is more efficient, and it can actively improve the concentration in a quantitative time, and also help teachers understand the emotional changes of students. The effective task assignment in multimodal interactive learning can further improve the efficiency, fun and interaction of distance learning. Therefore, we further analyse the characteristics of multimodal interaction and learn the dynamics of the three scenes.

4.2 The body characteristics of embodied cognition

Table 2 The experience of multimodal interaction in three scenarios

Learning scenario	Interactive features	Interactive experience
Face-to-face learning	**Speaking:** instruction, command, inquiry **Gestures:** parallel, imitation	Students are more **imitating** the teacher's operation, although the teaching is very clear, but more **dependent on** the teacher.
Computer video remote learning （Dell）	**Speaking:** instruction, command, inquiry, descriptive words (e.g. colour, character, direction) **Gestures:** vertical, autonomous emphatic gestures **Screen:** a display	It is **inconvenient** to explain the operation. Teacher need to hold the object up to the camera. It is also **difficult** for teachers to observe students' emotions
Multimodal remote learning （HP Sprout + Empatica's E4 wristbands）	**Speaking:** instruction, command, inquiry, descriptive words (e.g. colour, character, direction), concern **Gestures:** parallel+ vertical, autonomous emphatic gestures, indicative gesture **Screen:** two display	Students can see objects in more **three-dimensional** way and ask questions more **actively**. In teaching, gestures are more **focused on** the explanation of knowledge.

Body dynamics is an important concept in embodied cognition. Embodied cognition means that the human body plays a key role in cognitive processing. Cognition is mainly formed through the interactive experience of the various senses in the environment and active forms [15]. Embodied cognition points out that cognition, body and environment are nested and inseparable from each other; cognition exists in the brain, and the brain exists in the body [16]. I summarized the physical characteristics of multimodal interaction in three scenarios, as shown in Table 2. Face-to-face learning conveys knowledge to students through the interaction of speech and gestures. As can be seen from the blue block of the teacher-student interaction in Figure 3, face-to-face learning is more inclined to synchronous teaching, while the teacher is teaching and the students are following the imitation. Single-screen remote learning tends to be a step-by-step teaching. The teacher needs to pick up the teaching parts and align them with the camera to explain the situation. A single-screen camera with only one front-view camera often causes the item to be picked up and put down, which is a state of vertical gesture, as shown in Figure 3. The teacher can't pick up a lot of parts to explain in the hand, so the teacher takes a step and the student follows one step, which leads to slower efficiency. The amount of information that a screen needs to carry is large, which results in a small learning window and it is not convenient to use it.

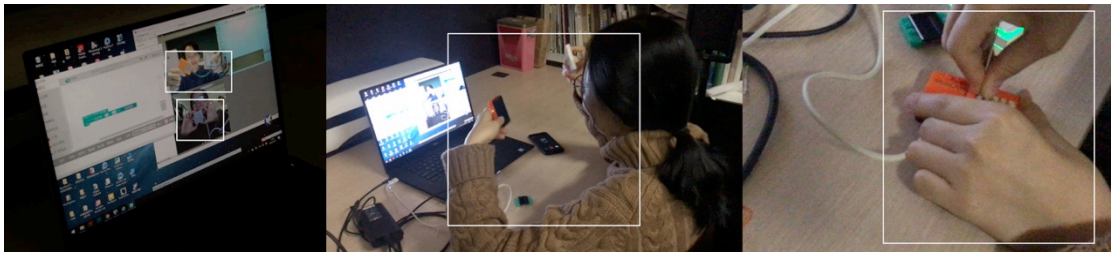
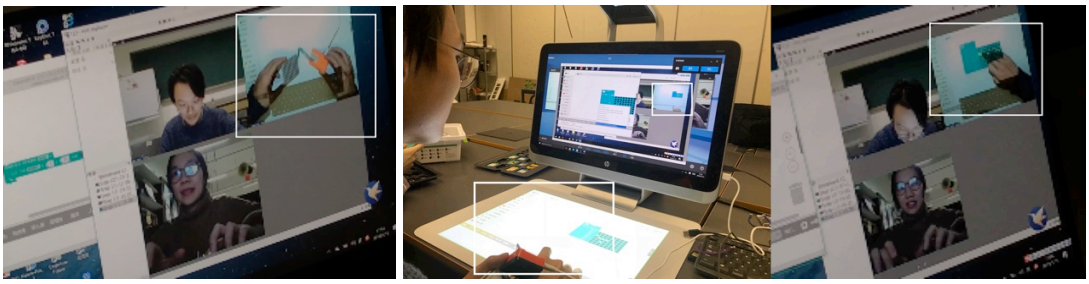


Figure 3. Gesture style when computer video remote learning (when the teacher explains the part, the part should be picked up, and the splicing should be placed on the desktop again, the gesture is a vertical up and down state)

The multimodal learning efficiency can be reached almost in line with face-to-face learning, and the top camera plays a big role. The top camera can capture the hand movements on the desktop, so it is no longer necessary to lift the hand and use the horizontal gesture to guide, as shown in Figure 4 (1). When there is an incomprehensible problem, the teacher will explain through the emphasis gestures and descriptive language, such as lifting a certain part to the camera and saying: "The red block is inserted into the second hole from the right side", then point his finger at that location. There are three components when guiding the physical task: identification of target objects, description of actions to be performed on targets, and confirmation that the targets have been performed successfully. The guiding task includes the elements of pointing object and extracting information from objects (identify objects, location, direction, and orientation). The display can only indicate something by cursor. The touch pad on the desktop is equivalent to the expansion of the host screen. You can drag the software operation into the touch pad, then point your finger to the position of the touch pad. The contents of the desktop are captured by the top camera and can be displayed on each other's screens., as shown in Figure 4 (2). Hand representation is richer than a cursor pointer in terms of degree of freedom. The pointer does not support representational hand gestures which showing orientation and movement [17].

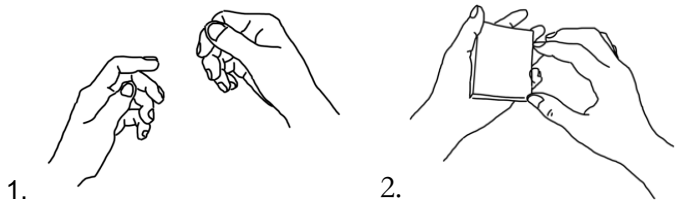
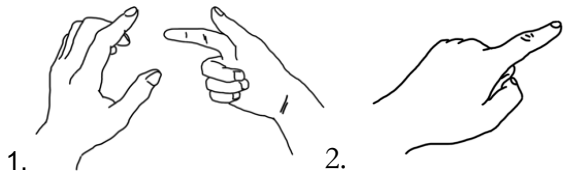
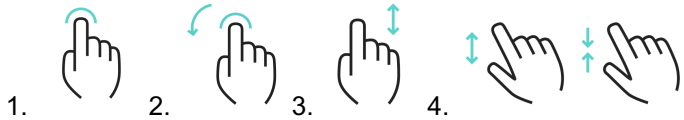


(1) Hands on the table splicing *(2) The teacher points to the touch pad*

Figure 4. Gesture style in multimodal remote learning

Double screens can effectively assign learning tasks to different screens, and double cameras also give teachers and students more perspective. I sorted out the gesture classification under multimodal interaction. As shown in table 3, the splicing gesture, the emphasis gesture and the directional gesture can be used to clearly teach. In the process, we will remind the teacher of the students' emotional changes, and the teacher can also adjust the teaching schedule and content for the students.

Table 3 Gesture of Multimodal Remote Learning

Teaching gesture 1. Insert a cable 2. Stitching module	
Explanatory gestures: 1. Emphasize (hold up and aim at the camera) 2. Point at a position	
Touch screen gestures: 1. Click,2. Drag,3. Slide,4. Zoom in and zoom out	

5 Discussion

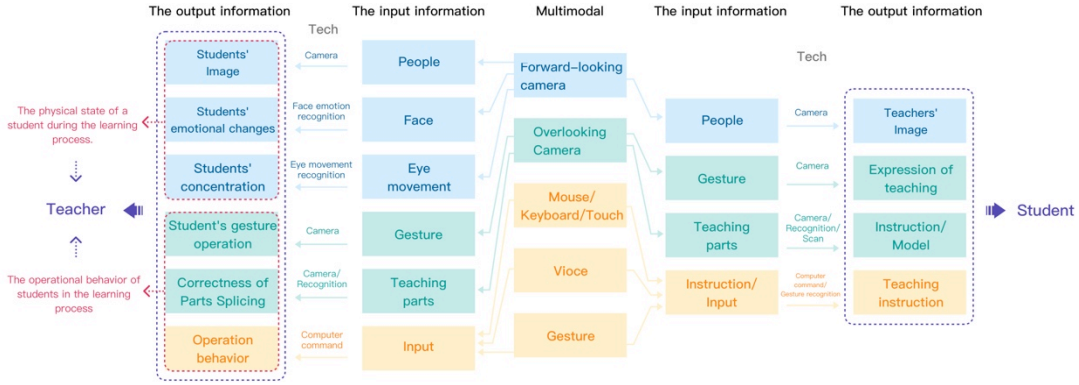


Figure5. Model of Multimodal Interactive based on Remote Learning（MIRL model）

Based on the above analysis, we have established a multimodal interaction model for remote learning between teachers and students, referred to as the MIRL model (Figure 5). In the remote state, the teachers and students interact with each other and then establish a connection between them. After relying on the camera to establish the image of both sides, the ordinary talking and chatting is very smooth, and there will be many communication obstacles when the teaching of the physical object is added. However, the development of multi-screen, AI recognition and data analysis has brought great learning advantages to both teachers and students. Multimodal interaction is to make teaching and learning clearer, and the interaction rate between teachers and students is higher, which enhances students' initiative and learning. However, too many channels may increase the burden of learning, and effective allocation can optimize learning efficiency. How to allocate needs to consider where the individual's concerns are. Teachers focus on the one hand is to express clearly, on the other hand is to observe students, so multi-screen can help teachers to have a clear direction of teaching, emotional recognition and eye movement recognition can convey students' learning status to teachers. Students only pay attention to learning. Remote learning students need to receive a large amount of learning information. The information carried on a single screen is very limited, and many windows are easy to overlap. Multi-screen is good for assigning tasks, and the learning ideas are clearer, not easily interfered by other things, and also beneficial for observing teaching gestures. The identification and scanning of teaching parts can promote students to explore the learning content and promote the initiative of learning.

In real life, when people interact with each other, they often use multiple channels to complete a task [18]. For example, a person likes to explain something in a dance, explaining that speaking is the main channel, and dancing is the auxiliary channel. The same strategy can be used in remote learning (Figure 6). The language gesture expression of the students in the learning process is the main channel of the task operation, and the individual emotion or eye movement change is the auxiliary channel, and the picture and data are transmitted to the teacher. When the main task is completed, the channel with high flexibility and high recognition accuracy is selected as the main channel, and the auxiliary channel is also used to perfect the user experience. This process is parallel. Then the teacher can judge the right or wrong of the student's task according to the main channel, adjust the

teaching task through the auxiliary channel, and continuously update in the process of teaching and learning.

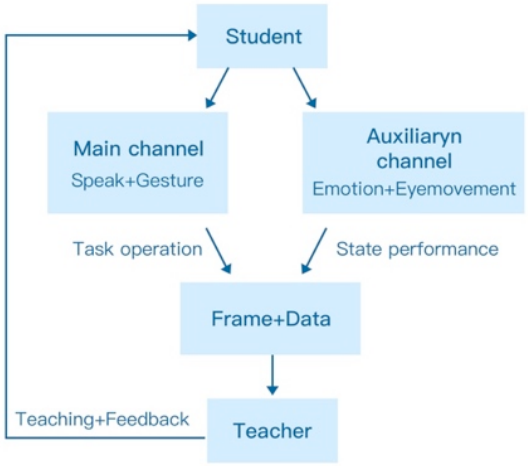


Figure6. Main and auxiliary channel for students

We also summarize several inspirations for the product design of remote learning:

1. It is necessary to present students with a three-dimensional teaching perspective and an immersive teaching atmosphere, which is conducive to students' understanding of physical teaching.
2. In the future, we should consider effectively combining more channels to improve the natural efficiency of human-computer interaction and make human-computer interaction more and more close to human-to-human interaction. Paying attention to the interaction mode of auxiliary channels will be beneficial to observe the details of students and understand students more comprehensively. For example, data analysis of students' emotions, eye movements and hand operations.
3. And further consider how to enhance the user role and personalization in the teaching process. In the future, the teacher's teaching responsibilities should be weakened. Teachers are no longer the masters, but the facilitators and helpers of student learning. They provide guidance and feedback according to the needs of students when necessary.
4. Although face-to-face learning is the current common teaching method, there is a great prospect for remote learning in the future. Remote learning brings a lot of space and time, suitable for synchronous teaching between teachers and students, and also suitable for fragmented interactions. It is worth studying how to use multimodal interaction to help students conduct personalized self-learning and let teachers provide effective guidance in the process. After that, we will also design an interactive kit to help students with remote collaborative learning according to the multimodal interaction model, providing low-cost, immersive, interesting, clear and anytime and anywhere learning.

6 Conclusion

This paper studies the impact of multimodal mode on remote learning quality and explores the effectiveness of body characteristics for knowledge transfer and learning. According to this goal, the research has proved that the multimodal interaction is conducive to improving

the efficiency of remote learning, and further explores the reasons from the physical interaction between teachers and students, and studies the multimodal interaction characteristics of multi-screen, multi-view, language gestures and emotions. Thus, we propose a multimodal interaction model of remote learning and a parallel and complementary cooperation mode of main and auxiliary channels. We hope that our research content will provide a theoretical basis for the remote interactive learning. In the future, we will further explore multimodal personalized learning in the distance, meet the learner's demand for learning efficiency, and open the active learning of the whole people.

7 References

New Media Consortium, EDUCAUSE Learning Initiative. (2017). Horizon Report Basic Education Edition

Interaction (Design) is a complex phenomenon. http://fredvanamstel.com/blog/interaction-design-is-a-complex-phenomena

Caschera M. C., Ferri F., Grifoni P. (2007). "Multimodal interaction systems: information and time features". *International Journal of Web and Grid Services (IJWGS),* Vol. 3 - Issue 1, pp 82-99.

Aleksandar Karadimce. (2013). Model for interactive, collaborative and multimedia mobile learning environment. *International Conference on Advances in Mobile,* pp 585-588.

Su li, Saiqi Pi, Lingfang Tian. (2018). Research on Adult Mobile Learning Design Based on Activity Theory. *Adult Education*,8,1-4.

Xiuqi Feng, Ying Liu. (2003). Study interactive system in distance education. *China Educational Technology*, 3,69-71.

Fumihide Tanaka, Toshimitsu Takahashi, Shizuko Matsuzoe, Nao Tazawa, Masahiko Morita. (2014). Telepresence Robot Helps Children in Communicating with Teachers who Speak a Different Language. *HRI '14 Proceedings of the 2014 ACM/IEEE international conference*, pp 399-406.

Kyoung Wan Cathy Shin, Jeonghye Han. (2016). Children's Perceptions of and Interactions with a Telepresence Robot. *HRI '16 The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, pp 521-522.

Kostia Robert, Dingyun Zhu, Weidong Huang, Leila Alem, Tom Gedeon. (2013). MobileHelper: Remote Guiding Using Smart Mobile Devices, Hand Gestures and Augmented Reality. *SA '13 SIGGRAPH Asia 2013 Symposium on Mobile Graphics and Interactive Applications*, NO.39.

Sarah D'Angelo. (2017). Contextually Relevant Gaze Representations for Remote Learning. *CHI EA '17 Proceedings of the 2017 CHI Conference*, pp 268-273.

Nancy Yao, Jeff Brewer, Sarah D'Angelo, Mike Horn, Darren Gergle. (2018). Visualizing Gaze Information from Multiple Students to Support Remote Instruction. *CHI EA '18 Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, No. LBW051.

Wearable devices enter the field of education. http://www.jingmeiti.com/archives/8996

David S. Kirk, Danaë Stanton Fraser. (2005). The Effects of Remote Gesturing on Distance Instruction. CSCL '05 Proceedings of the 2005 conference on Computer support for collaborative learning, pp 301-310.

LI Bao-lin, ZHAO Jian-chuan, LIN Wen-bin. (2013) Application of Wavelet Analysis in Signal Denoising. Electronic Design Engineering, 9: 39-42.

Wei Chen, Benyu Guo. (2014). Embodied, cognitive science. Psychological Exploration, 2,111-116.

IONESCU T, VASC D. (2014).Embodied cognition：challenges for psychology and education.Procedia-social and behavioral sciences,128:275-280.

Jane Li, Anja Wessels, Leila Alem, Cara Stitzlein. (2007).Exploring Interface with Representation of Gesture for Remote Collaboration. *OZCHI '07 Proceedings of the 19th Australasian conference on Computer-Human Interaction*, pp 179-182.

Tian Qi. (2018). Human-computer interaction intelligent interface technology based on multimodal. Nanjing University.

About the Authors:

Zhiyong Fu: Zhiyong Fu is doctoral advisor, associate Professor of Information Art and Design Department at Tsinghua University. His areas of expertise are information/interaction design, service design. He is committed to promote art & technology integration and interdisciplinary innovation education.

Yuyao Zhou: Yuyao Zhou is a master of Art and Design Department at Tsinghua University. Her major is information/interaction design. She have published two international conference papers. The main research interests are design thinking and multimodal interaction.

Acknowledgement: This research is a phased achievement of "Smart R&D Design System for Professional Technology" project and supported by the Special Project of National Key R&D Program ——“Research and Application Demonstration of Full Chain Collaborative Innovation Incubation Service Platform Based on Internet+” (Question ID: 2017YFB1402000), "Study on the Construction of Incubation Service Platforms for Professional Technology Fields" sub-project (Subject No. 2017YFB1402004).