Remote Virtual Companion *via* Tactile Codes and Voices for The People With Visual Impairment*

GE Song[1)], HUANG Xuan-Tuo[2)], LIN Yan-Ni[1)], LI Yan-Cheng[2)], DONG Wen-Tian[3)],
DANG Wei-Min[3)], XU Jing-Jing[4)], YI Ming[5)**], XU Sheng-Yong[1)**]

([1)]*Key Laboratory for the Physics & Chemistry of Nanodevices, School of Electronics, Peking University, Beijing* 100871, *China*;

[2)]*School of Electronics Engineering and Computer Science, Peking University, Beijing* 100871, *China*;

[3)]*Peking University Sixth Hospital, Peking University Institute of Mental Health, NHC Key Laboratory of Mental Health* (*Peking University*),

National Clinical Research Center for Mental Disorders (*Peking University Sixth Hospital*), *Beijing* 100191, *China*;

[4)]*School of Microelectronics, Shandong University, Jinan* 250100, *China*;

[5)]*Key Laboratory for Neuroscience, School of Basic Medical Sciences, Neuroscience Research Institute, Department of Neurobiology,*

School of Public Health, Peking University, Beijing 100191, *China*)

Abstract **Objective** Existing artificial vision devices can be divided into two types: implanted devices and extracorporeal devices, both of which have some disadvantages. The former requires surgical implantation, which may lead to irreversible trauma, while the latter has some defects such as relatively simple instructions, limited application scenarios and relying too much on the judgment of artificial intelligence (AI) to provide enough security. Here we propose a system that has voice interaction and can convert surrounding environment information into tactile commands on head and neck. Compared with existing extracorporeal devices, our device can provide a larger capacity of information and has advantages such as lower cost, lower risk, suitable for a variety of life and work scenarios. **Methods** With the latest remote wireless communication and chip technologies, microelectronic devices, cameras and sensors worn by the user, as well as the huge database and computing power in the cloud, the backend staff can get a full insight into the scenario, environmental parameters and status of the user remotely (for example, across the city) in real time. In the meanwhile, by comparing the cloud database and in-memory database and with the help of AI-assisted recognition and manual analysis, they can quickly develop the most reasonable action plan and send instructions to the user. In addition, the backend staff can provide humanistic care and emotional sustenance through voice dialogs. **Results** This study originally proposes the concept of "remote virtual companion" and demonstrates the related hardware and software as well as test results. The system can not only achieve basic guide functions, for example, helping a person with visual impairment to shop in supermarkets, find seats at cafes, walk on the streets, construct complex puzzles, and play cards, but also can meet the demand for fast-paced daily tasks such as cycling. **Conclusion** Experimental results show that this "remote virtual companion" is applicable for various scenarios and demands. It can help blind people with their travels, shopping and entertainment, or accompany the elderlies with their trips, wilderness explorations, and travels.

Key words artificial visual aid, remote virtual companion, tactile code, visually impaired users, navigation
DOI：10.16476/j.pibb.2023.0053

According to the World Health Organization, there are nearly 40 million blind people worldwide[1]. The blindness of some people is congenital, whereas that of others is acquired. Visual impairment affects the quality of patients' life, most of whom are isolated and unable to integrate themselves into the daily life.

Currently, there are main two types of devices, namely, implanted and extracorporeal devices, that can help blind people regain some of their independence[2]. Implanted devices attempt to restore

a part of the patient's visual function. Such a system uses external cameras to capture images, process their data, compress and encode them, and send stimulation instructions to the microelectrode arrays implanted in the patient's retina[3], choroid[4], and visual cortex[5] using wireless transmission technology to stimulate nerve cells or axons, thus stimulating the patient's brain to create discrete white halos and a blurred black-and-white image, for example, Argus II[6], IRIS II[7], Epi-Ret 3[8], IMI[9], Alpha-IMS[10]. Because these visual prosthetics provide images with few pixels, they cannot provide an accurate and reliable basis for the user's decisions and movements, and therefore, they are mostly in the experimental stage and have not yet been widely used. Extracorporeal devices[11] transmit visual information to the patient by other means such as touching or hearing, such as the Kinect-based navigation device[12], tongue-based vision device BrainPort[13], smartphone-based navigation systems[14], ultrasound-based obstacle detection sensors[15], 3D obstacle detection smartphones[16], smart guide canes[17], voice navigation systems[18], and pedestrian and lane detection systems[19]. These devices can help blind people perform specific tasks such as walking in residential areas. Naturally, a blind person, accompanied by a guide dog or sighted person, will be able to participate in most of the daily tasks of ordinary people, such as shopping, parties, or even sporting events.

Existing studies have extensively focused on artificial intelligence, including AI logical cognitive computing[20], social interactive computing[21], and the construction of virtual companion systems[22]. Applications include virtual companions for the insomnia of the general population[23], virtual companions for the elderly people[24], elderly care[25], and conversational agents for mental health care[26-28]. With the rapid development of artificial intelligence and remote wireless communication technologies, this paper proposes the concept of "remote virtual companion" to help blind people in their daily life more effectively. The "remote virtual companion" refers to the extension of human perceptions through various devices such as cameras, radars, infrared sensors, and other optoelectronic devices, which collect information of the surrounding environment, aggregate and identify the information through a local chip, and transmit the clas-sified information data to the remote cloud processor and backend staff for remote companion. Based on the experience and judgment of the backend staff and features such as AI image recognition and cloud computing, the system makes multiple judgments and transmits the final judgment results to the remote user in the form of "instructions", which are received by the user through a tactile or auditory device. It is a complex working model that integrates local information collection, remote communication, backend services, cloud computing, and AI technology. The wearable sensors can collect various surrounding environmental parameters, including temperature, humidity, and altitude as well as surrounding vision information, speed and vitals such as heartbeat or pulse. The precise information collected by these sensors greatly extends the original biological information such as sight, hearing, smell, taste, and touch. This allows for more comprehensive information to be gathered about the surrounding environment, which can help detect dangers in advance and ensure the safety of the user. The dual protection provided by the backend staff and local AI as well as the real-time voice feedback provide security for the user, and the virtual companionship from a real person in the backend provides some emotional support.

Virtual companions can be used in many scenarios, such as remote assistance to help the elderly people when they travel alone, or warning environmental hazards during wilderness explorations.

1 Experimental setup

1.1 System modules and application scenarios

Figure 1 outlines the components of the "remote virtual companion" system and several application scenarios. The system mainly comprises the local information sensing, local instruction output, remote information transmission, cloud data and computing, and backend devices and staff modules. Among them, the local information sensing module may include wearable optoelectronic devices and sensors such as optical camera, infrared camera, laser radar (LIDAR), ultrasonic radar, temperature-pressure sensor, body condition detector. The local instruction output module may consist of various devices such as tactile, auditory, and voice communication devices. The cloud module and the complexity of backend can be changed as the application scenarios change. The system integrates technologies from different

industries, *e. g.* local information collection, remote communication, backend services, cloud computing, and AI technology. For people such as the blind, lonely elderlies, children left behind, firefighters, field crews, or friends and relatives whom we are temporarily unable to accompany, the device can replace the companionship and assistance of a real person by acting as a "remote virtual companion" which can not only help the local user to complete the complex tasks that they would otherwise be unable to do independently (*e. g.* traveling a long distance alone), but also provide emotional support through real-time communications between the real person in the backend and the user.
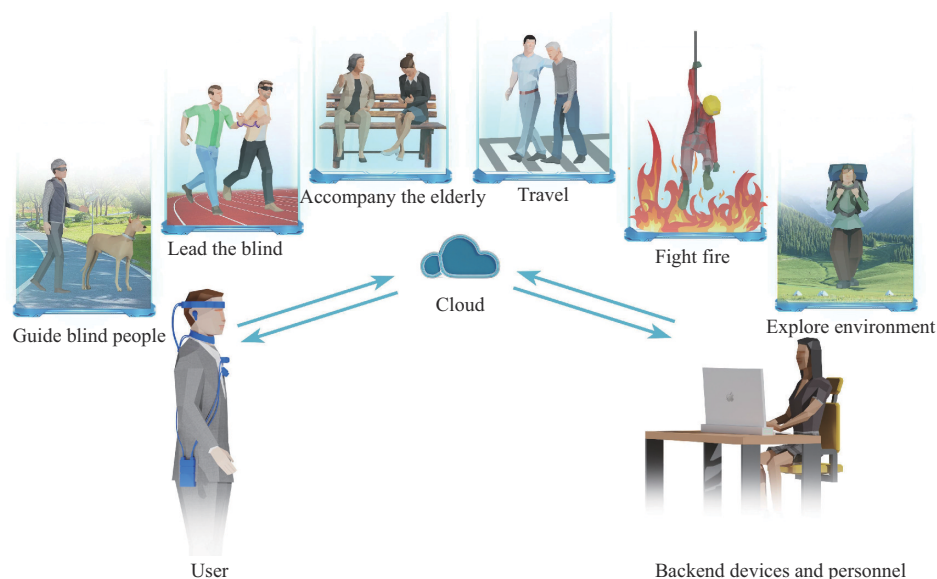
Fig. 1 Composition and several application scenarios of the "remote virtual companion" system

The system consists of four parts: local wearable devices, cloud, backend devices and personnel. It can be applied in scenarios such as replacing guide dogs to guide the blind, replacing track and field race leaders, remotely accompanying the elderly people, accompanied travels, fire-fighting assistance, and field exploration assistance.

1.2 System hardware

To help the people with visual impairment complete complex tasks, we built a hardware system which mainly consists of the local information sensing module, local instruction output module, remote information transmission module, cloud data and computing module, and backend devices & staff module.

We combined the local information sensing module, local instruction output module and remote information transmission module into a wearable device. Figure 2 shows the device; Figure 2a is its overall picture, while Figures 2b, c are the front and back views of the inside of the wearable device box, respectively. The device was developed and assembled based on an Upboard development board. It is connected to the vibrating motor headband made of a flexible material (which integrates eight vibrating

motors), neck ring (which integrates one vibrating motor), and three USB 4k cameras. The cameras transmit the surrounding environment information from different viewpoints to the development board which uses 5G to transmit the visual information to the backend staff *via* AliCloud, who then uses a keyboard to transmit the instructions back to the development board after analyzing the information and making a judgment. Upon reception of the instruction, the development board controls nine vibrating motors *via* GPIO bus to vibrate for different lengths of time and at different frequencies in order to send the instruction to the user who receives the instruction and responds. In the meanwhile, voice messages are available as required by the scenario.

The wearable devices for receiving vibration instructions can be distributed around the head, neck, waist, ankle, *etc*., depending on the user's ability to

perceive and receive vibrations. If the devices are temporarily worn on the head and neck, the user can hide them by wearing a peak cap and scarf for privacy protection.

Each vibrating motor is 10 mm in diameter and 2.7 mm in thickness, and operates between 3.0 and 5.0 V. As the operating voltage increases, the vibration increases significantly. Based on test results, we selected 3.8 V as the operating voltage for the nine vibrating motors, taking into account the human perceptibility to vibrations and tolerance to shocks. After testing the motor spatial resolutions of the head and neck, we distribute the eight vibrating motors around the head, namely, in the front and rear, left front and left rear, right front and right rear of the head: this distribution is in line with the human body's natural ability to distinguish between front, rear, left and right, and can help the user to identify and quickly respond to the instructions. The neck motor is directly in front of the neck to receive stop instructions, which are highly distinguishable from the head vibrations. In addition, to enhance the vibrations, we use hot melt adhesive to make a hemispherical bump on the surface of the motor to concentrate its vibration pressure, thus making it easier to perceive.
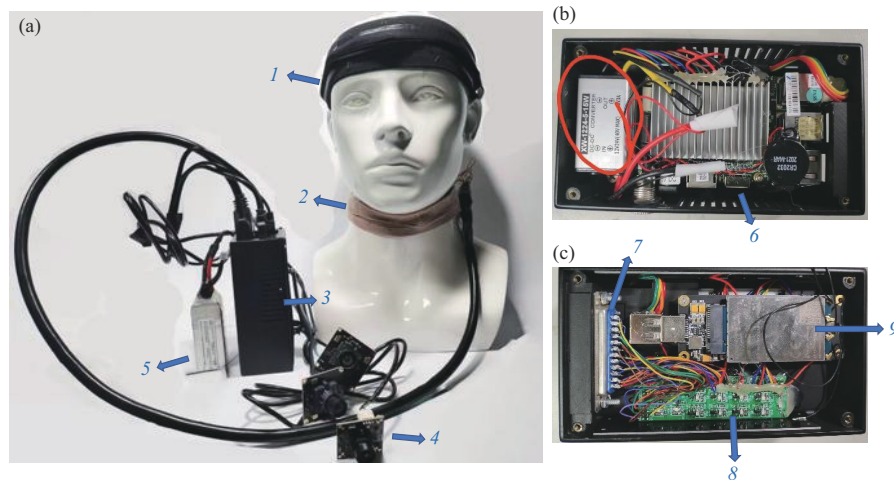
Fig. 2 Composition of the device and internal hardware composition of the device box

(a) Composition of the device: the device consists of five parts: *1*, vibrating motor headband made of a flexible material, which integrates eight vibrating motors; *2*, neck ring made of a flexible material, which integrates one vibrating motor; *3*, wearable device box for remote communication connections with the backend; *4*, three 4k cameras connected to the device box *via* USB; *5*, battery. (b) and (c) are the front and back of the device box: *6*, up board; *7*, interface with the nine vibrating motors; *8*, PCB board that amplifies IO interface signals to drive motor vibrations; *9*, 5G module.

The device can remotely receive peripheral visual information captured by the three cameras and remotely control motor vibrations to send instructions to users. By connecting the local device to the backend *via* the cloud, the device can be used to remotely assist users with complex daily tasks such as arm movements, walking, and cycling. Its workflow is as follows: the backend acquires the peripheral visual information, such as street boundaries, steps, potholes, pedestrians, bicycles and garbage bins, as well as indoor items such as tables, chairs, cups and books, and transmits tactile instructions after analyzing and judging the visual information. The instructions are transmitted *via* 5G to the device worn by the subject to drive the vibrating motors at specified locations, thereby acting as tactile instructions, *e. g.* instructions for navigation or arm movements, and descriptions of road conditions, pedestrian appearances, names of stores, locations of tables, description of the menu or other items.

1.3 Instruction set for tactile vibration devices

One of the advantages of this system is its rich instruction sets: the instruction sets are easy to understand and can be mastered by the user after a short training period. Different instruction sets can be applied for different scenarios, *e. g.* walking and cycling navigation instruction set, arm movement instruction set, and instruction set for numerical

patterns used in entertainment activities such as cards playing. After negotiating with the subject, the backend staff can apply a specified instruction set for the scenario to simplify the code and thereby make it easier for the subject to receive the instructions faster. The vibration instructions of the experimental setup are transmitted through the vibrating motors in the headband and neck ring. In the following tests, we applied appropriate instruction sets after conducting the spatial resolution, vibration duration, and vibration interval tests on the head and neck. Take the walking navigation instruction set as an example. Table 1 shows the navigation instruction set, where the vibrating motors in the front, rear, left, right, left front, left rear, right front, and right rear of the headband are coded G, B, L, R, GL, BL, GR, and BR, respectively, and the vibrating motor in the neck ring is coded S.

Table 1 Walking navigation instruction set

Action	Instruction	Action	Instruction
Background	Long B	Stop	Long S
Turn back	Triple short B	Warning stop	Triple short S
Forward	Long G	Turn right	Triple short R
Fast forward	Triple short G	Oblique right turn	Long GR
Turn left	Triple short L	Step forward	Combined G+B
Oblique left turn	Long GL	Step back	Combined B+G
Step right	Long R	Step up	Long BR
Little turn right	Triple short GR	Little turn left	Triple short GL
Step left	Long L	Step down	Long BL

1.4 Field scenario tests

To test the functionality, practicality, accuracy, and safety of the developed device, we designed three types of tests: arm movement tests (*e. g.* block building and puzzle construction), walking navigation tests (*e. g.* avoiding obstacles, supermarket shopping and getting coffees at cafes) and fast navigation tests (cycling tests on two road conditions). The arm movement tests simulate the tasks that require grasping and sorting actions in daily life, while the walking navigation tests simulate the tasks that require interactions with the environment and pedestrians. The above two types of tests are designed to test the functionality and practicality of the device, while the cycling tests simulate the tasks that require the instructions to be sent and received at a higher speed, which are designed to test the accuracy and safety of the device. In the meanwhile, we could

adjust the number and locations of the cameras as required during the tests. During walking navigation tests, the cameras could be installed on the chest forward or tilted down. During cycling tests, the cameras could be installed on the chest forward or on the bike basket tilted down. During arm movement tests, the cameras could be installed on the chest or neck. The backend instructions can also be defined based on scenarios. For example, in navigation scenarios such as walking and cycling, the main instructions are navigation instructions for such movements as moving forward, moving backward, and turning. During the puzzle construction tests, the main instructions are to move the arms backward, forward, left and right, and rotate the puzzle, continue, *etc*. During card playing tests, the instructions are the patterns and numbers of the cards.

Ten volunteers between the ages of 20 and 30 years participated as subjects in all the tests in the present study. All subjects gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of Beijing Municipal Science & Technology Commission (Grant No. Z201100005520092). All subjects were in good health, had no tactile perception problems, and had not previously participated in similar tests. All subjects were of medium build, between 150 cm and 180 cm in height, and none of them were extremely thin or obese, which ensures that each subject can wear the test devices appropriately, thus making it easier to adjust the devices to the most comfortable level for each subject. In the practical scenario tests, except for the comparison cycling tests, in order to temporarily deprive the subject of his judgment ability, each subject wore a blindfold before the test to temporarily deprive him/her of the visual interaction ability, so that he or she could only rely on the tactile and auditory instructions given by the device to perform actions.

2 Results

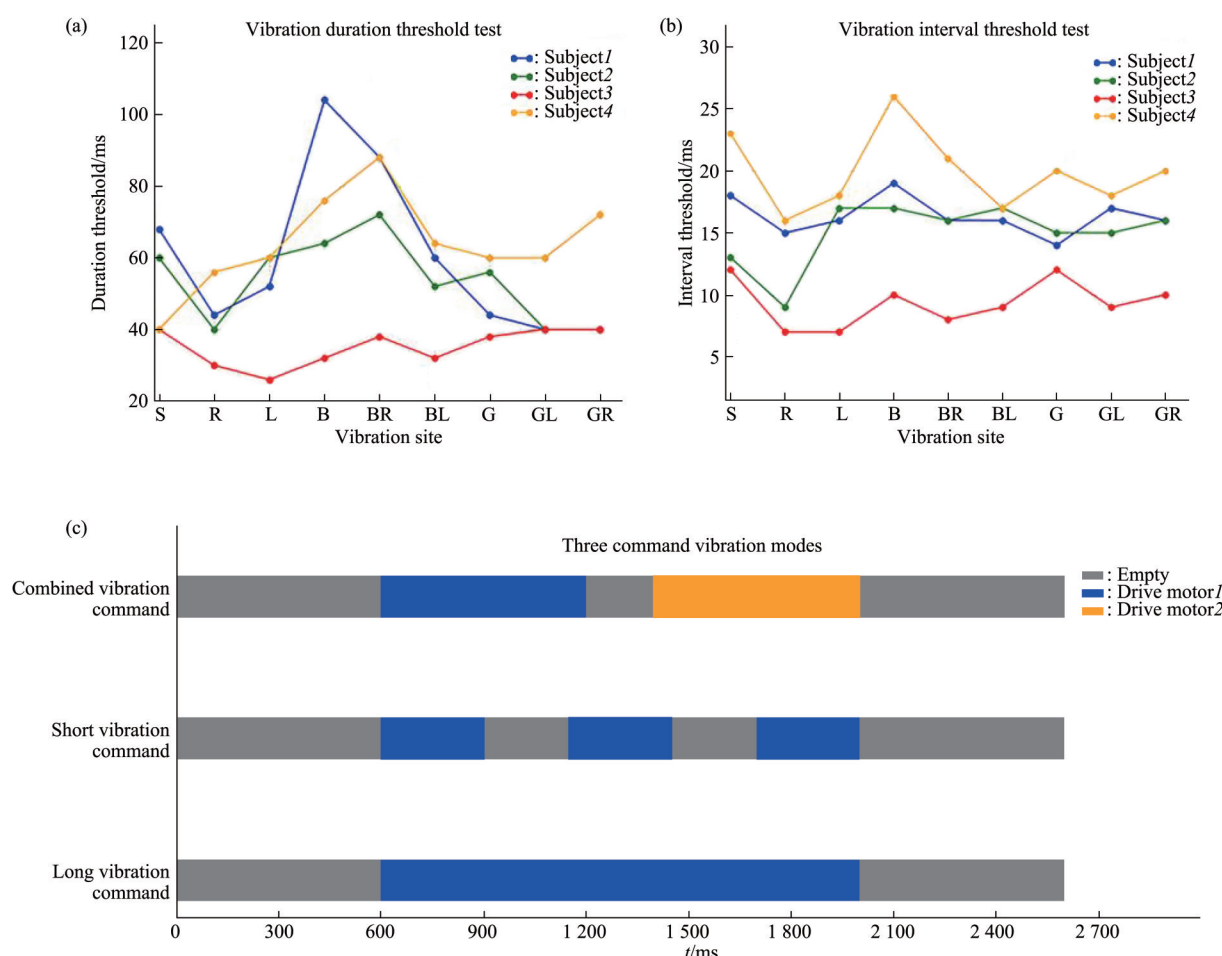### 2.1 Results of tactile vibration threshold tests

Since the vibration instructions of the experimental setup are transmitted through the vibrating motors in the headband and neck ring, before selecting the instruction set, we first conducted

spatial resolutions tests of motor vibrations on the head and neck: spatial perception tests were conducted on 5 subjects, and the results showed that the spatial resolution threshold of the head was mainly affected by the length of hair; the motor spatial resolution of the scalp was between 1.5 cm and 3.0 cm; in other words, when the distance between two motors is above 3.0 cm, the subject can clearly perceive the vibrations of different motors; the spatial resolution threshold of the neck is within 2.0 cm. During the test, a total of 8 motors were installed in the front, rear, left, right, left front, left rear, right front and right rear of the headband. The head circumference of an adult is usually between 54 and 58 cm, and therefore, the distance between two adjacent motors is more than 6 cm, which meets the spatial resolution requirement of the motors. The neck-ring motor is directly in front of the neck to receive stop instructions, which are highly distinguishable from the above-mentioned 8 motors so that the subject can perceive the stop warnings more quickly and sensitively.

Secondly, we conducted threshold tests on the shortest vibration durations and the shortest vibration intervals at the specified positions (head and neck): by using the Arduino platform to control the duration and interval of the motor vibration, and then using the least variable method of psychology to obtain the tipping point data. Ten tests were performed on each vibration point on the four subjects, and the average of the resulting tipping values was the threshold value for each vibration point.

The test results are shown in Figure 3. Figure 3a shows the data plot comparing the vibration duration thresholds for the nine points on the four subjects, and Figure 3b shows the data plot comparing the vibration interval thresholds for the nine points on the 4

Fig. 3　Results of vibration duration and interval threshold tests and schematic diagram of the three vibration instructions
(a) Data plot of vibration duration thresholds, the results showed that the vibration duration thresholds fell between 25−105 ms; (b) data plot of vibration interval thresholds, the results showed that the vibration interval thresholds fell between 8−26 ms. (c) The 3 vibration instructions. From bottom to top in the figure are the instruction for a long vibration, for three short vibrations, and for combined vibrations of two motors. The total vibration duration for all 3 instructions is 1 400 ms.

subjects. The data show that the vibration duration thresholds and interval thresholds vary across the 9 points on the same subject, and that the levels of sensitivity of different subjects to vibrations vary, and so do the thresholds. In general, the vibration duration thresholds range from 25−105 ms, with the highest being 105 ms, and the vibration interval thresholds ranged 8−26 ms, with the highest being 26 ms. In other words, when the motor vibration duration exceeds 105 ms and the vibration interval exceeds 26 ms, the subject can clearly perceive the vibration frequency. According to the above data, we defined 3 instructions for the vibration instruction set: single long vibration of one motor, three short vibrations of the same motor and combined vibrations of two motors. The 3 vibration instructions are shown in Figure 3c. Test results show that our scalp and neck can clearly distinguish between these 3 instructions, where the combined vibrations are asynchronous vibrations rather than synchronous vibrations, because it is easier to perceive and distinguish between asynchronous vibrations than synchronous ones. In addition, in order for the subjects to clearly distinguish between instructions, we defined the total vibration durations of the three instructions to be the same. The total vibration duration for all three instructions is 1 400 ms, which means the single long vibration of one motor lasts for 1 400 ms, the 3 short vibrations of the same motor last for 300 ms with 250 ms interval between each vibration and the combined vibrations of two motors last for 600 ms with 200 ms interval between each vibration. This design solution adequately meets the threshold requirements for the neck and head sites tested earlier without delaying the subject's actions for too long an instruction.

2.2 Results of arm movement tests

Five tests were designed to investigate how well the device can control arm movements, including simple classification of color blocks, 3×3 puzzle, building of 2 pyramids of blocks, and finally card playing test with entertainment interactions. The levels of complexity of the first 4 tests gradually increased, with their average instruction execution times ranging from 3.7 s to 6.0 s. The response speeds meet the demands of most arm movements in daily life, including grasping, sorting, organizing, and storing of items. Each subject wore a blindfold to temporarily remove visual information and rely solely on the tactile and auditory instructions provided by the device to make movements. The first 4 arm movement tests were conducted indoors without considering the safety issues caused by unexpected situations, so tactile instructions were used for all of them; the fifth test, namely card playing, also used tactile instructions to prevent the leakage of card information.

Figure 4a is the video screenshot of the color-block test: the color-block test sorted 60 color blocks into blue, green, and red piles, and used 5 instructions for start, end, and color-block classification. Subjects could accurately sort different color blocks without incorrect movements. The test lasted 233 s and used 62 instructions with an average execution time of 3.8 s per instruction. Figure 4b is the video screenshot of the puzzle test: the puzzle test placed nine puzzle pieces into a wooden frame in correct order and ensured that each piece is oriented correctly. The 10 instructions used are: start, end, move forward, backward, left, and right by one puzzle piece, select the piece, rotate the piece 90° counterclockwise after placing its, flip the piece after placing it, and proceed to the next step. The subjects could accurately place the different puzzle pieces into the puzzle frame in correct order and orientation without incorrect movements. The test lasted 191 s and used a total of 51 instructions with an average execution time of 3.7 s per instruction.

Figure 4c is the video screenshot of the first block building test: the block building test arranged 28 color blocks into a seven-level pyramid (with each level consisting of 1−7 blocks). The eight instructions used are: start, end, move forward, backward, left, and right by one block, select the block, and move up by one level. The subjects could accurately select the color blocks of the specified colors and place them in correct positions without tipping or misplacing them. The test lasted 299 s and used a total of 60 instructions with an average execution time of 5.0 s per instruction. Figure 4d is the video screenshot of the second block building test: the test arranged 30 color blocks into a four-level pyramid (with each level consisting of 1−16 blocks). The eight instructions used are: start, end, move forward, backward, left and right by one block, select the block, and move up by one level. The subjects could accurately select the color blocks of the specified colors and place them in correct positions without tipping or misplacing them.
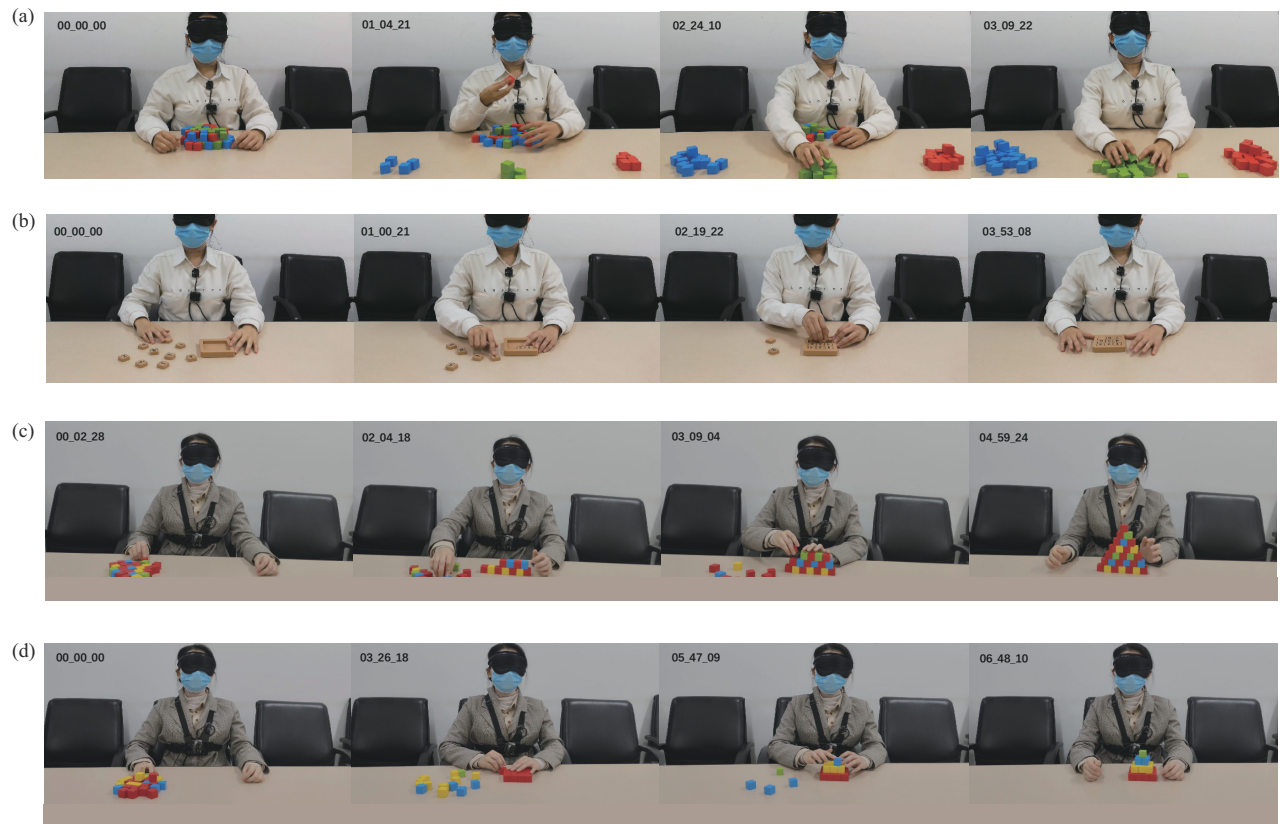
Fig. 4 Screenshots of the four arm movement tests

(a) Color-block sorting test, the test sorted 60 color blocks into blue, green, red piles; (b) puzzle test, the test placed nine puzzle pieces into the puzzle box; (c) the first block building test, the test arranged 28 color blocks into a seven-level pyramid shape; (d) the second block building test, the test arranged 30 color blocks into a four-level pyramid shape. The top left corner of each picture shows the moment of the current test video screenshot.

The test lasted 408 s and used a total of 68 instructions with an average execution time of 6.0 s per instruction. The average instruction execution times of the two block building tests are significantly longer than those of the previous two tests, namely, color block and puzzle tests, mainly because the subjects needed to pay attention to the placement and balance of the blocks during the block building test.

Figure 5d is the video screenshot of the card playing test with entertainment interactions: the blackjack rules were followed. The instructions used are: start and point instructions (A−K, 13 types). The game in the video took 32 s, with 20 points for the subject and 18 points for the opponent, the subject won. During the card game, the subject could receive information of cards in her hand through vibration instructions, make judgments, and finally, after the cards were revealed, receive information of the opponent's cards through instructions to determine the outcome of the game.

2.3 Test results of several life scenarios

We conducted three navigation tests for blind people: walking and obstacle avoidance, shopping in a supermarket, and buying coffee at a cafe and being seated. The walking and obstacle avoidance test was a simple navigation instruction test, while the supermarket shopping and coffee tests were closer to our daily tasks, and involved dodging passersby and the combination of navigation instructions with arm instructions. Each subject wore a blindfold to temporarily remove visual information and rely solely on the tactile and auditory instructions provided by the device to make movements. In the three walking tests, the main navigation instructions were also sent in the form of tactile codes, and auditory instructions were added mainly to allow for timely voice communications with the subjects in case of unexpected situations such as difficult-to-avoid pedestrians.

Figure 5a is the video screenshot of the walking and obstacle avoidance test: there were intentionally deployed obstacles along the corridor, and the navigation instructions used in the test are simple, including forward, backward, stop, left, and right. It took the subject in the screenshot 42 s to walk through a 1.5 m wide and 15 m long corridor with 6 obstacles without touching the walls or obstacles. Figure 5b is the video screenshot of the coffee test at a cafe: the subject got up from the seat, went to the counter to pick up his coffee and returned to the seat. In addition to the navigation instructions used in the walking and obstacle avoidance test, other instructions used in this test are: turn left, turn right, pick up, and move hand

forward, backward, left, and right. It took the subject in the screenshot 72 s to successfully pick up the coffee and return to the seat without colliding with passersby, tables, or chairs. Figure 5c is the video screenshot of the supermarket shopping test: the subject entered the entrance of the supermarket and selected three predefined items. In addition to the navigation instructions used in the walking and obstacle avoidance test, other instructions used in this test are: turn left, turn right, squat down, start choosing items, move fingers up, down, left, right, select, and pick up. It took the subject in the screenshot 228 s to successfully select three items without colliding with passersby or shelves.
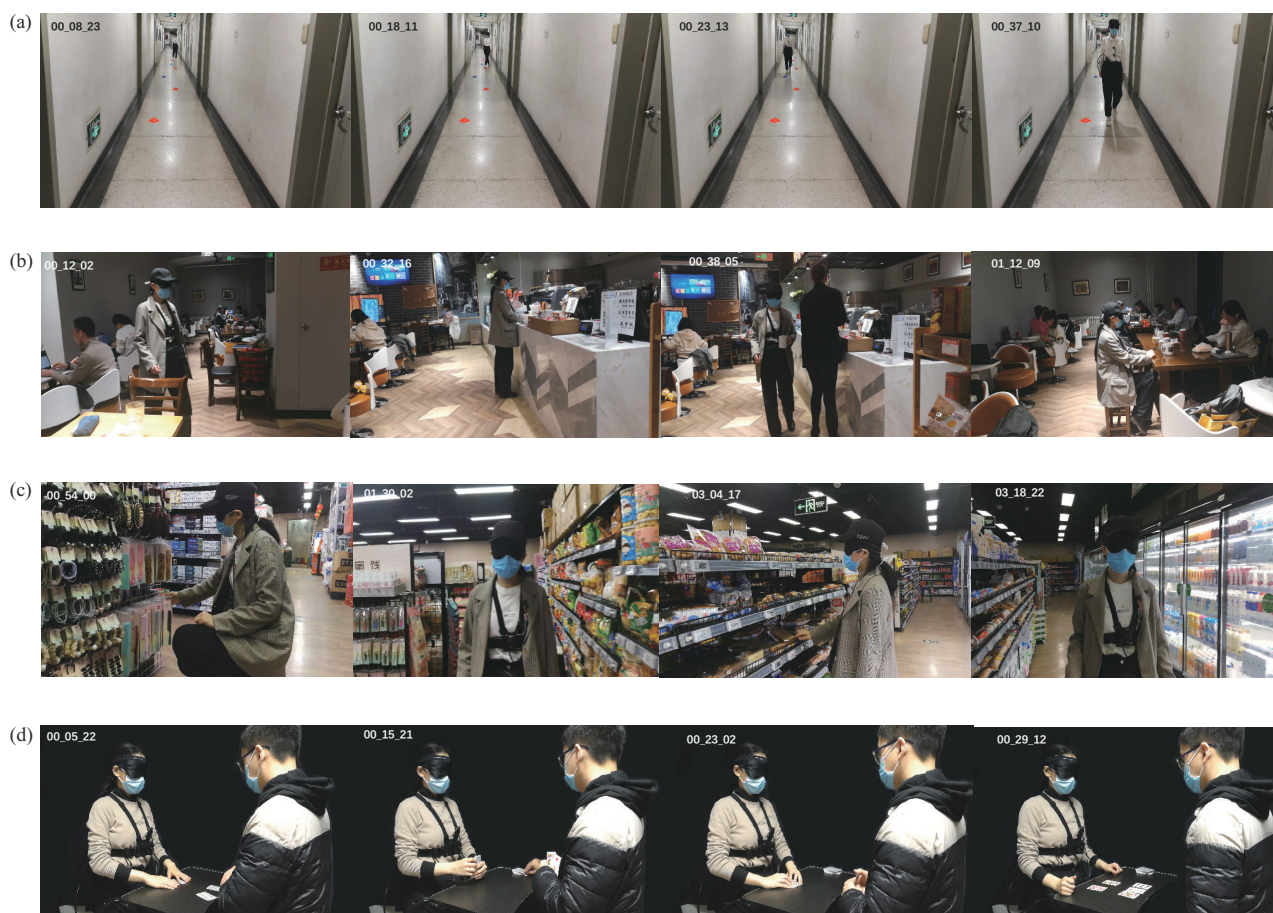
Fig. 5　Screenshots of the three walking tests + card playing test

(a) Walking and obstacle avoidance test, the subject walked through it without colliding with the walls on either sides or obstacles on the floor; (b) coffee test at a cafe: in a crowded cafe, the subjected avoided tables, chairs, and customers, went to the front desk to pick up his coffee, and returned to the original seat; (c) supermarket shopping test: the subject dodged the shelves and customers in the supermarket, found the 3 predefined items in 3 shelves and checked out; (d) card playing test: the test completed a blackjack game and announced the winner. The top left corner of each picture shows the moment of the current test video screenshot.

The above two types of tests (arm movement tests and walking navigation tests) simulated the tasks that require grasping and sorting actions in daily life, and the tasks that require interactions with the environment and pedestrians, respectively. The test results showed that after being temporarily deprived of the visual interaction ability, the subjects wearing the device could still complete the tasks perfectly by relying solely on the tactile and auditory instructions given by the device, which proved that the device has the functions that can assist patients in their daily life.

2.4　Results of cycling tests

The previous two types of tests mainly simulated slow daily tasks. To meet the demand for fast-paced daily tasks, we also conducted cycling tests. The tests were conducted on two road conditions, namely, a road with three right-angle turns, and a narrow s-curve park trail.

The road with right-angle turns is about 27 m long and 1.9−3.4 m wide, while the s-curve park trail is about 24 m long and 1.2 m wide. Figures 7c, d

show the details of the two road conditions. We performed the cycling tests on the 4 subjects in the following manner: firstly, the subjects cycled on the two roads twice to collect the offsets from the centerlines and their cycling speeds under normal cycling conditions; then, the subjects wore blindfolds to remove visual information, and cycled 5 times only under control of vibration and voice instructions given by the device to collect the offsets from the centerlines and their cycling speeds. The main navigation instructions were also sent in the form of tactile codes, and auditory instructions were added mainly to allow for timely voice communications with the subjects in case of unexpected situations such as difficult-to-avoid pedestrians, or collision with a surrounding bridge pillar or bush.

Figure 6 shows the video screenshots of the cycling tests on the two road conditions. Figure 6a is the right-angle cycling test, while Figure 6b is the s-curve cycling test. All 4 subjects successfully completed the cycling tests, making turns at the correct positions without falling down or colliding.
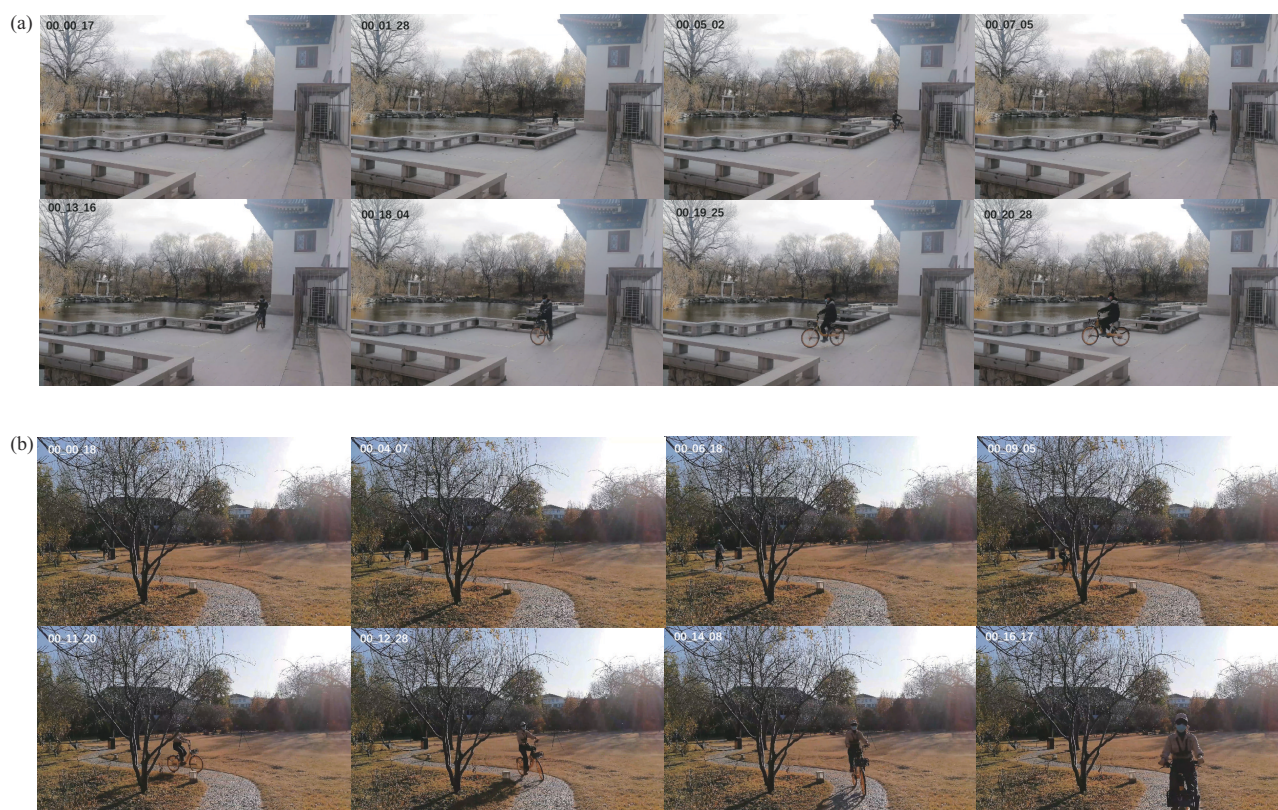
Fig. 6　Video screenshots of the cycling test on the road with right−angle turns+cycling test on the s−curve trail

(a) Right-angle cycling test, the test lasted 20 s; (b) s-curve cycling test, the test lasted 18 s. The top left corner of each picture shows the moment of the current test video screenshot.

The above cycling tests simulated the tasks that requires the instructions to be sent and received rapidly. The test results showed that after being temporarily deprived of the visual interaction ability, the subjects wearing the device could still complete the cycling tasks perfectly by relying solely on the tactile and auditory instructions given by the device. Being able to cycle safely in straight lines, s-curves, and right angles proves that the device can send and receive instructions rapidly, accurately and safely.

3 Discussion

3.1 Delays and predictions

In the tests such as walking & obstacle avoidance and cycling tests, the delay of the device itself and the delay of the human body's response to the vibration instruction shall be taken into consideration. In other words, a predictive analysis is required when the backend sends an instruction.

Figures 7a, b are two cycling tests. As shown in Figure 7a, in the tests, we found that we could use the angle θ between the camera and the horizon to offset the prediction time. The picture captured by the camera is s ahead of the subject's footstep or position of the front wheel of the bicycle . And we find that the tilt angle of the camera can compensate the delay of the system. Futhermore, if we suppose the walking or cycling speed is v, the total delay is t and $s=v\times t$, the tilt angle of the camera titled in Figur 7a can be expressed as $\theta=\arctan(h/s)$. In this case, the effect of the delay of the system can be exactly eliminated. In the walking & obstacle avoidance and cycling tests, due to the delays of response to vibration instructions, long vibration instructions are usually: including forward, stop, turn left, turn right, step left, step right, turn left, and turn right. Long vibration instructions can significantly reduce the human response time. The average response time in the tests is 180 ms. Combined with the system delay of 330 ms, the total delay is $t=510$ ms. The cycling test speed is between 1 and 2 m/s, and the walking speed is between 0.5 and 1 m/s. The lead, *i.e.* distance s, created by tilting the camera is between 26 and 102 cm. In Figure 7a, the subject's $s=70$ cm, $h=70$ cm, and $\theta=45°$. When the camera is tilted 45°, the image in the video window on the backend computer's corresponding screen is the real-time location of the subject. In practice,
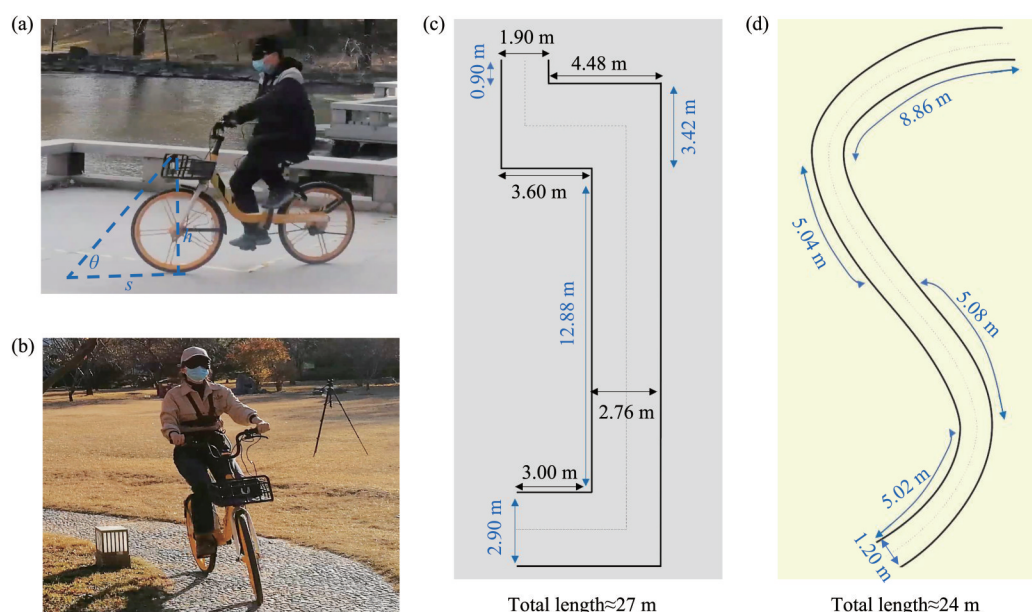
**Fig. 7 Cycling tests+two road conditions**

(a) Right-angle cycling test: θ is the angle between the camera and horizon; h is the height of the camera from the ground; and s is the distance between the road surface illuminated by the camera and the bicycle. (b) S-curve cycling test. (c) Right-angle road condition: the total length is 27 m, and the narrowest width is 1.9 m. (d) S-curve road condition: the total length is 24 m, and the width is 1.2 m.

although we cannot always set θ very precisely, we can still considerably offset the effect of the system delay by even setting an appropriate θ.

3.2　Analysis of cycling test results−accuracy and safety

Four subjects were given comparative riding tests under the two road conditions. The first road is a road with three right-angle turns, and the second road is a narrow s-curve park trail. Figure 8a shows the trajectories of the four subjects who cycled normally on the right-angle road twice, and Figure 8b shows the trajectories of the 4 subjects who wore blindfolds to remove visual information, and cycled 5 times on the right-angle road only under control of vibration and voice instructions given by the device. As shown in the figures, there are more turning points in the trajectories of the subjects wearing the device.

However, the subjects wearing the device could cycle along the road without colliding with the surrounding walls or pillars, and make three turns correctly and in time. Figure 8c shows the trajectories of four subjects who cycled normally on the s-curve trail twice. Figure 8d shows the trajectories of the four subjects who wore blindfolds to remove visual information, and cycled 5 times on the s-curve trail only under control of vibration and voice instructions given by the device. As shown in the figures, there are more turning points in the trajectories of the subjects wearing the device. However, the subjects wearing the device could cycle along the s-curve trail without colliding with the neighboring bush, except for that the subjects cycled off the trail at some points. The subjects could make two big turns correctly and in time, and cycle along s-curves.
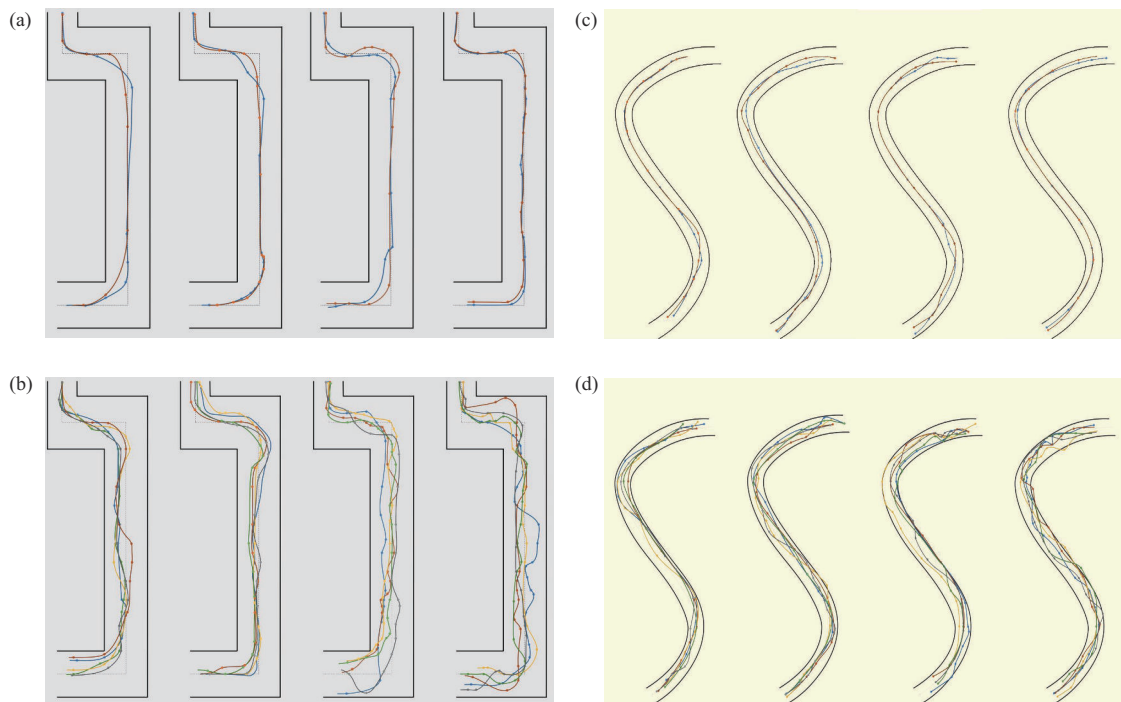
Fig. 8　Trajectories of right−angle cycling+s−curve cycling

(a) Trajectories of the four subjects who cycled normally on the right-angle road twice; (b) trajectories of the 4 subjects who cycled on the right-angle road 5 times only under control of vibration and voice instructions given by the device; (c) trajectories of the 4 subjects who cycled normally on the s-curve trail twice; (d) trajectories of the 4 subjects who cycled 5 times on the s-curve trail only under control of vibration and voice instructions given by the device.

Taking 20 cm as the unit, we calculated the percentages of the offset of the cycling trajectories from the centerline (0−20 cm, 20−40 cm, 40−60 cm, 60−80 cm, *etc*.) in the overall road. The offset

percentages of the cycling trajectories from the centerlines of the right-angle road and s-curve trail are shown in Figure S1−S4: Figure S1 shows the offset percentages of the cycling trajectories from the

centerline of the right-angle road where the 4 subjects cycled twice normally: the results show that the offsets of all the subjects are within 0−100 cm, and within 40 cm on over 85% of the road; the offsets of the subject *4* from the centerline are within 20 cm on over 90% of the road. Figure S2 shows the offset percentages of the cycling trajectories from the centerline of the right-angle road where the 4 subjects cycled 5 times with the device: the results show that the offsets of all the subjects wearing the device are within 0−140 cm, which is 40 cm more than the normal cycling condition, and within 100 cm on over 85% of the road; the offsets of the subject *2* from the centerline are within 60 cm on over 90% of the road. Figure 9a shows the comparison of offset on over 85% of the right-angle road where the four subjects cycled normally or with the device. In other words, when the road width is 280 cm or more, the subjects can successfully pass the sections with right-angle turns with the assistance of the device.
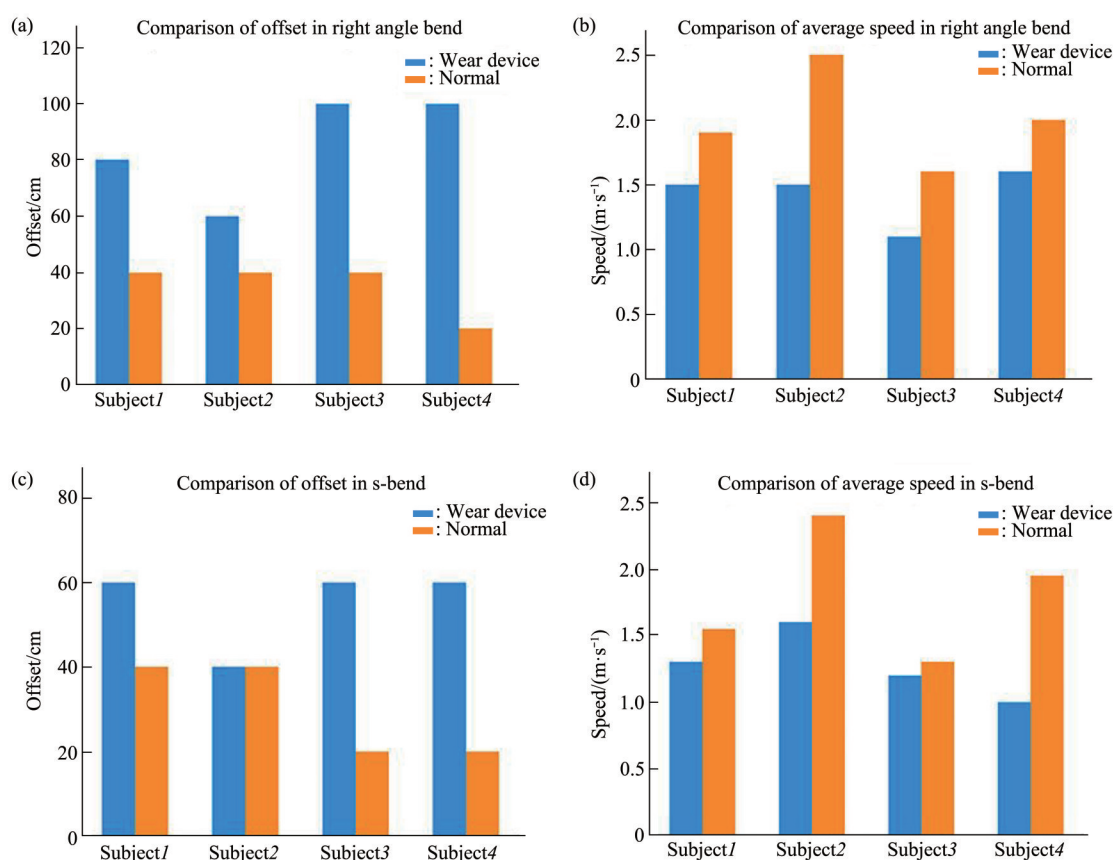
Fig. 9 Diagrams of offset and cycling speeds in the comparison tests under the two road conditions
(a) Offset of the trajectories from the centerline on over 85% of the right-angle road where the four subjects cycled normally or with the device; (b) average speeds of the 4 subjects who cycled on the right-angle road normally or with the device; (c) offset of the trajectories from the centerline on over 85% of the s-curve trail where the four subjects cycled on the s-curve trail normally or with the device; (d) average speeds of the four subjects who cycled on the s-curve trail normally or with the device. We also compared the average cycling speeds: Figure 9b shows that the average speeds of the 4 subjects cycling normally are within 1.6−2.5 m/s, while the average speeds of them with the device are within 1.1−1.6 m/s. The subjects with the device cycled slower than they did without it mainly because they feared the unknown with the blindfold. Their speeds could be improved by training.

Figure S3 shows the offset percentages of the cycling trajectories from the centerline of the s-curve trail where the 4 subjects cycled twice normally: the results show that the offsets of all the subjects are within 0−40 cm, and within 40 cm on over 85% of the trail; the offsets of the subject *3* and subject *4* from the centerline are within 20 cm on over 90% of the trail. Figure S4 shows the offset percentages of the cycling

trajectories from the centerline of the s-curve trail where the four subjects cycled five times with the device: the offsets of the subjects wearing the device are within 0−80 cm, which is 40 cm more than the normal cycling condition, and within 60 cm on over 85% of the trail; the offsets of the subject 2 from the centerline are within 40 cm on over 90% of the trail. Figure 9c shows the comparison of offset on over 85% of the s-curve trail where the 4 subjects cycled normally or with the device. In other words, when the trail width is 160 cm or more, the subjects can successfully pass the s-curves with the assistance of the device. Most of park trails are 150 cm wide or more, which means the device can help people cycle on most of them. We also compared the average cycling speeds: Figure 9d shows that the average speeds of the 4 subjects cycling normally are within 1.3−2.4 m/s, while the average speeds of them with the device are within 1.0−1.6 m/s. The subjects with the device cycled slower than they did without it mainly because they feared the unknown with the blindfold. Their speeds could be improved by training.

3.3 Innovation and advantages

Among the existing two types of devices that can help people with visual impairments restore partial visual function, implanted devices generally face technical challenges, such as the need for surgical implantation, unrecoverable trauma, difficulty in transocular operation, the need for biocompatibility of device materials, high cost, and limited number of pixels leading to the result that it can only restore weak vision, and so on. Extracorporeal devices are safer and more portable, without causing damage to the subjects. However, the existing devices have relatively simple instructions and limited application scenarios, making it difficult to adapt to the complex environment and road conditions of modern cities. Moreover, they also rely heavily on AI's judgment which can't provide enough security, so they are not suitable for large-scale application and promotion.

For example, the tongue-based vision device BrainPort uses an array of tactile electrodes on the tongue to help wearers perceive weak visual information around the periphery so that they can identify objects in front. However, the number of electrodes is limited by the small area of the tongue, and the use will affect diet and speech; ultrasound-based obstacle detection sensor can only detect close-range obstacles, which means it can only help users

judge the distance from obstacles and complete independent walking on specific roads; smart guide cane mainly uses visual servo technology and walking obstacle avoidance algorithm, and uses ground motion tactile feedback and voice to provide users prompts. Similarly, the instructions are relatively simple and the application scenarios are few, which is not suitable for large-scale promotion. Compared with system mentioned above, our system converts surrounding environmental information into tactile instructions for the head and neck, and assists in voice interaction. It can covertly, effectively, and safely provide subject action instructions, remote network information support, and realize the function of guide. Its effectiveness, safety, and information volume are superior to existing in extracorporeal devices. Bedsides, the system is low-cost, low-risk, and suitable for various life and work scenarios. The innovation and superiority of this system are reflected in the following aspects:

Firstly, we use a combination of AI and backend staff to provide sufficient security for users through dual judgment: if the user relies solely on local AI for performing daily tasks and entertainment interactions, several sensors and computation are required, including the recognition and detection of various possible situations, and backing up. With the assistance of the backend staff, the system can provide humanistic care, emotional support and more safety. For examples, the elderly people can be accompanied by the backend staff to banks, post offices, and hospitals, thus enabling them to travel alone with more peace of mind; children can be accompanied remotely to do their homework. The backend staff can also identify more detailed situations that AI cannot accurately identify at present, *e. g.* recognizing the face of a person wearing a mask, or his/her facial expressions. Alternatively, local AI is mainly designed for the following scenarios: (1) unexpected events such as evacuations in case of dangers, or emergency stop instructions of pedestrians or vehicles that appear all of a sudden; (2) AI features can assist the user with entertainment interactions (*e. g.* card playing), or help the user to search the destination and required information when he/she walks outside. We are conducting research on image recognition. The system can already recognize 3 types of special items, *i. e.* mahjong tiles, traffic lights, and sidewalks for the blind, at a success rate of 89%, and detect items at

short distances and within 120° with a laser radar so as to minimize user's collisions with the surrounding items. We are researching how to recognize facial expressions with a binocular lens. The combination of AI and backend staff can provide more secure remote virtual companion for more groups and more demanding scenarios.

Secondly, we choose vibration on the skin of the head and neck as the output mode of tactile instructions to make the vibration instructions comfortable and recognizable: studies have shown that among various communication methods such as light, vibration, sound, poking, and temperature, vibration is the most reliable and fastest communication method[29-30]. Among wearable devices, devices that utilize visual interactions need to attract the user's visual attention, and devices that utilize auditory interactions need a quiet environment to avoid the interference of noise. In contrast, tactile interactions can make up for the deficiencies of the visual and auditory interactions without using the user's visual and auditory functions[31]. In addition, auditory instructions are not suitable for some scenarios. For example, vibration instructions are more appropriate for card games, elderly people with hearing loss and noisy environments where voice instructions are not suitable. Among the various tactile feedbacks (pressure stimulation, temperature stimulation, and electrical stimulation), the tactile feedback generated by vibration is more comfortable, efficient, and adjustable[32]. Hence, our device uses vibration tactile instructions as a long-term instruction output mode.

The densities of tactile sensor surfaces in various parts of the body are different. The tactile spatial resolutions of the fingers, palms, and other parts of the body are high, while the tactile spatial resolution of the back is low. Our tests show that the tactile spatial resolution of the scalp is within 2−3 cm, which is slightly higher than that of the back, and the tactile spatial resolution of the neck is within 2 cm. In addition, the skin of the head and neck has natural ability to distinguish between front, back, left, and right directions, sufficient space to provide multiple vibration sites, and does not affect daily life. So our vibration site selection is, the dedicated stop button is on the neck, while the other eight buttons are deployed in a circle around the head, thus enabling the user to distinguish between different instructions rapidly. Selecting head and neck skin to provide multiple vibration sites is conducive to the learning, cognition and resolution of instruction set.

Finally, we use tactile and auditory instructions to provide sufficient action instruction, environmental information and humanistic care in a variety of scenarios: we use a tactile encoding composed of eight motors on head and one motor in the neck, combined with voice instructions to provide users with surrounding environment information and action instructions. The tactile encoding formed by the nine motors can not only provide relatively simple instructions such as forward, stop, left turn, and right turn, but also provide a large amount of instruction information similar to encoding languages such as Morse code, which can be applied to various life and work scenarios. In addition, the combination of tactile and voice instructions can not only send action instructions through tactile encoding to allow subjects to quickly and accurately respond, but also provide more detailed and complex information through voice, such as overall road conditions or surrounding environment description. And through voice interaction, users can provide timely feedback and convey their needs. Through communication with backend staff, the security and reliability of the system can be further increased and personalized services can be customized to realize humanistic care and emotional sustenance. At the same time, tactile and auditory instructions can be flexibly combined in different scenarios. For example, tactile instructions are more reliable in noisy areas, and voice instructions are clearer when shopping. The combination of tactile and auditory instructions can respond to different scenarios, providing sufficient information and humanistic care.

3.4 System signal transmission delays

System delays include the video transmission delay, instruction transmission delay, and backend staff's and user's response times. The video and instruction delays can be offset by the tilt angle of the camera, and the user's response time can be minimized by training (currently to (180±30) ms; instructions for cycling and walking tasks that require fast responses are as simple as possible, and are not confusing or unclear). The number and locations of the cameras can be adjusted for different scenarios. For examples, in the cycling scenario, one camera can be installed on the bike basket and tilted to offset the

transmission delays, while the other camera can be installed on the user's chest to monitor the big picture; likewise, in the waking scenario, one camera can be tilted down to offset delays, while the other one can be installed on the chest for monitoring purposes.

In order to further reduce system delays, we should start from two aspects. There are mainly two means to further reduce system delays: firstly, to reduce the delay of human reaction speed, which is composed of two parts, the backend staff identifying and judging the scene and issuing instructions, and the users receiving instructions and making corresponding actions. For backend staff, an interactive program can be used for training in advance. Through presenting different application scenarios and situations in the UI interface, backend staff can learn to issue correct instructions to improve the accuracy of instructions and reduce the reaction time. During training, VR technology can be used to enhance the immersion of backend staff. Through generating various application scenarios and road conditions, VR can help backend staff learn instructions from multiple perspectives. After a long period of familiarity with device instructions, the subjects can also significantly reduce the duration of the reaction time for tactile and auditory perception perceptual responses.

Secondly, significantly properly increasing the proportion of AI will also reduce system delay. AI not only uses cloud electronic map data, obstacle avoidance navigation algorithms, and image recognition of objects needing attention in application scenarios to quickly issue instructions to compensate for the defects of backend staff's reaction speed or network transmission delay, but also uses AR technology to circle the objects that need requiring to be focused by backend staff in the backend interface, provide navigation route arrows, and perform directional recognition. Visualization enables backend staff to timely understand the results of AI recognition calculations and navigation, so as to facilitate timely correction and assistance in certain task scenarios where AI can't complete correctly.

4　Conclusion

This paper presents the concept of "remote virtual companion" for the first time and demonstrates

the related hardware and software as well as test results for some scenarios.

The "remote virtual companion" is based on the latest remote wireless communication and chip technologies, microelectronic devices, cameras, and sensors worn by the user, as well as the huge database and computing power in the cloud. With the support of the "remote virtual companion", the backend staff can get a full insight into the scenario, environmental parameters, and status of the user remotely in real time. In the meanwhile, by comparing the cloud database and in-memory database and with the help of AI-assisted recognition and manual analysis, the "remote virtual companion" can quickly develop the most reasonable action plan and send instructions to the user. In addition, the backend staff can maintain a real-time dialog with the user to provide humanistic care.

This "remote virtual companion" can replace a guide dog or human companion and help the user with visual impairment perform daily tasks that they would otherwise be unable to do independently, *e. g.* traveling, shopping, and entertainment. Furthermore, it helps elderly people who are not adapted to the fast pace and high technologies of the modern life, such as shopping, visiting hospital alone, or assisting them with their travels.

We built a "remote virtual companion" prototype. The device uses a combination of AI and backend staff to provide sufficient security for users through dual judgment, which leverages the experience and judgment of the backend staff. In addition to normal instructions, the backend staff can also send instructions such as "abort" and "pause" to advise the blind people to seek help from nearby pedestrians, police, or ambulance personnel. Hence, the device not only enables information exchanges between the local user, cloud, and backend staff, but also provides emotional companionship and support for the user. The local AI-assisted system, electronic map in the cloud and other databases can achieve rapid responses, making up for the deficiencies of slow responses of the backend staff or network transmission delays. Working with the local LIDAR, infrared sensors, and ultrasonic ranger for obstacle avoidance, the local and remote AI techniques can assist the user to identify dangerous items more accurately and rapidly, and to execute emergency instructions for improved safety.

We choose vibration on the skin of the head and neck as the output mode of tactile instructions to make the vibration instructions comfortable and recognizable. The skin of the head and neck has natural ability to distinguish between front, back, left, and right directions, sufficient space to provide multiple vibration sites, and does not affect daily life, which is conducive to the learning, cognition and resolution of instruction set.

The combination of tactile and auditory instructions can provide the users sufficient action instruction, environmental information and humanistic care: the system can covertly send instructions to the user to achieve basic guide functions, for example, helping a person with visual impairment to shop in supermarkets, find seats at cafes, walk on the streets, construct complex puzzles, and play cards. Also, the system could meet the demand for fast-paced daily tasks such as cycling. Tactile instructions can provide multi-channel information encoding to increase the amount of input information, conceal it, and improve its security, while voice instructions are more sensitive and direct, support more complex instructions, can be corrected in case of emergencies, and can also provide humanistic care and emotional sustenance in a timely manner. In addition, backend work can provide positions for other disabled people and help them realize their pursuit of personal value.

In short, the "remote virtual companion" has extensive development and application prospects.

Supplementary 　Available online (http://www. pibb. ac.cn or http://www.cnki.net):
PIBB_20230053_Figure S1.pdf
PIBB_20230053_Figure S2.pdf
PIBB_20230053_Figure S3.pdf
PIBB_20230053_Figure S4.pdf

References

[1]　Bourne R, Flaxman S, Taylor H R, *et al*. Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the global burden of disease study. Lancet Glob Health, 2021, **9**(2): e130-e143

[2]　Walle H, De Runz C, Serres B, *et al*. A survey on recent advances in AI and vision-based methods for helping and guiding visually impaired people. Appl Sci, 2022, **12**(5): 2308

[3]　Villalobos J, Nayagam D X, Allen P J, *et al*. A wide-field suprachoroidal retinal prosthesis is stable and well tolerated following chronic implantation. Invest Ophth Vis Sci, 2013, **54**(5): 3751-3762

[4]　Abbott C J, Nayagam D X, Luu C D, *et al*. Safety studies for a 44-channel suprachoroidal retinal prosthesis: a chronic passive study. Invest Ophth Vis Sci, 2018, **59**(3): 1410-1424

[5]　Margalit E, Maia M, Weiland J D, *et al*. Retinal prosthesis for the blind. Surv Ophthalmol, 2002, **47**: 335-356

[6]　Humayun M S, Dorn J D, Da Cruz L M, *et al*. Interim results from the international trial of second sight's visual prosthesis. Ophthalmology, 2012, **119**(4): 779-788

[7]　Bhowmick A, Hazarika S M. An insight into assistive technology for the visually impaired and blind people: state-of-the-art and future trends, J Multimodal User In, 2017, **11**(2): 149-172

[8]　Roessler G, Laube T, Brockmann C, *et al*. Implantation and explantation of a wireless epiretinal retina implant device: observations during the epiret3 prospective clinical trial. Invest Ophth Vis Sci, 2009, **50**(6): 3003-3008

[9]　Kitiratschky V B, Stingl K, Wilhelm B, *et al*. Safety evaluation of "retina implant Alpha IMS" -a prospective clinical trial. Graef Arch Clin Exp, 2015, **253**(3): 381-387

[10]　Keserü M, Feucht M, Bornfeld N, *et al*. Acute electrical stimulation of the human retina with an epiretinal electrode array. Acta Ophthalmol, 2012, **90**(1): e1-e8

[11]　Tapu R, Mocanu B, Zaharia T. Wearable assistive devices for visually impaired: a state of the art survey. Pattern Recogn Lett, 2020, **137**: 37-52

[12]　Bhowmick A, Prakash S, Bhagat R, *et al*. IntelliNavi: navigation for blind based on kinect and machine learning//Murty M N, He X, Chillarige R R, *et al*. Proceedings of the Multi-disciplinary Trends in Artificial Intelligence: 8th International Workshop. India: Springer, 2014: 172-183

[13]　Richardson M, Lloyd E T, Petrini K, *et al*. Reading with the tongue: Individual differences affect the perception of ambiguous stimuli with the BrainPort//CHI. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. New York: CHI, 2020: 1-10

[14]　Croce D, Giarre L, La R F, *et al*. Enhancing tracking performance in a smartphone-based navigation system for visually impaired people//IEEE. Proceedings of the 2016 24th Mediterranean Conference on Control and Automation. Athens: IEEE, 2016: 1355-1360

[15]　Mocanu B C, Tapu R, Zaharia T B. When ultrasonic sensors and computer vision join forces for efficient obstacle detection and recognition. Sensors, 2016, **16**(11): 1807

[16]　Saez J M, Escolano F, Lozano M A. Aerial obstacle detection with 3-D mobile devices. IEEE J Biomed Health, 2015, **19**(1): 74-80

[17] Slade P, Tambe A, Kochenderfer M J. Multimodal sensing and intuitive steering assistance improve navigation and mobility for people with impaired vision. Sci Robot, 2021, **6**(59): eabg6594

[18] Hung Y H, Tsai K Y, Chang E, *et al*. Voice navigation created by vip improves spatial performance in people with impaired vision. Int J Environ Res, 2022, **19**(7): 4138

[19] Lei Y, Phung S L, Bouzerdoum A, *et al*. Pedestrian lane detection for assistive navigation of vision-impaired people: survey and experimental evaluation. IEEE Access, 2022, **10**: 101071-101089

[20] Abbass H A. Social integration of artificial intelligence: functions, automation allocation logic and human-autonomy trust. Cogn Comput, 2019, **11**(2): 159-171

[21] Lee S K, Kavya P, Lasser S C. Social interactions and relationships with an intelligent virtual agent. Int J Hum Comput St, 2021, **150**: 102608

[22] Dongye X, Weng D, Jiang H, *et al*. A modular haptic agent system with encountered-type active interaction. Electronics, 2023, **12**(9): 2069

[23] Philip P, Dupuy L, Sagaspe P, *et al*. Efficacy of a smartphone-based virtual companion to treat insomniac complaints in the general population: sleep diary monitoring versus an internet autonomous intervention. J Clin Med, 2022, **11**(15): 4387

[24] Arguello D, Rogers E, Denmark G H, *et al*. Companion: a pilot randomized clinical trial to test an integrated two-way communication and near-real-time sensing system for detecting and modifying daily inactivity among adults >60 years-design and protocol. Sensors, 2023, **23**(4): 2221

[25] Jegundo A L, Dantas C, Quintas J, *et al*. Perceived usefulness, satisfaction, ease of use and potential of a virtual companion to support the care provision for older adults. Technologies, 2020, **8**(3): 42

[26] Ahmad R, Siemon D, Gnewuch U, *et al*. The benefits and caveats of personality-adaptive conversational agents in mental health care. Inf Syst Front, 2022, **24**(3): 923-943

[27] Ahmad R, Siemon D, Gnewuch U, *et al*. Designing personality-adaptive conversational agents for mental health care. Inform Syst Front, 2022, **24**(3): 923-943

[28] Jegundo A L, Dantas C, Quintas J, *et al*. Ease of use and potential of a virtual companion to support the care provision for older adults. Technologies, 2020, **8**: 42

[29] Dargahi, J, Najarian, S. Human tactile perception as a standard for artificial tactile sensing-a review. Int J Med Robot, 2004, **1**: 23-35

[30] Roumen T, Perrault S, Zhao S. NotiRing: a comparative study of notification channels for wearable interactive rings//CHI. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. Seoul: CHI, 2015: 2497-2500

[31] Van D L, Schoonderwaldt E, Bird J, *et al*. MusicJacket-combining motion capture and vibrotactile feedback to teach violin bowing. IEEE T Instrum Meas, 2011, **60**(1): 104-113

[32] Park M, Bok B G, Ahn J H, *et al*. Recent advances in tactile sensing technology. Micromachines, 2018, **9**(7): 321

通过触觉编码和声音为视力障碍者提供
远程虚拟陪伴*

葛　松[1)]　黄轩拓[2)]　林衍旎[1)]　李沿橙[2)]　董问天[3)]　党卫民[3)]　徐晶晶[4)]
伊　鸣[5)**]　许胜勇[1)**]

([1)] 北京大学电子学院，教育部纳米器件物理与化学重点实验室，北京 100871；[2)] 北京大学信息科学技术学院，北京 100871；
[3)] 北京大学第六医院，北京大学精神卫生研究所，国家卫生健康委员会精神卫生学重点实验室（北京大学），
国家精神心理疾病临床医学研究中心（北京大学第六医院），北京 100191；[4)] 山东大学微电子学院，济南 250100；
[5)] 北京大学基础医学院神经生物学系神经科学研究所，神经科学重点实验室，北京大学公共卫生学院，北京 100191）

摘要　目的　现有的人工视觉装置分为植入式装置和体外辅助装置两种，但它们都有一些不足之处。植入式装置需要手术植入、会造成不可恢复创伤；体外辅助装置指令相对简单、应用场景较为单一、过于依赖人工智能（AI）的判断不能提供足够的安全性。本文提出了一种将周边环境信息转化成头颈部触觉指令、并辅助以语音交互的系统，其有效性、安全性、信息量等均优于现有体外辅助技术，同时也具有低成本、低风险、适合多种生活和工作场景等优势。**方法**　该系统借助最新的远程无线网络通讯技术、芯片技术，利用前方人员随身佩戴的微小型电子设备、摄像头和感应器，以及云端庞大的数据库和计算能力，后台人员可以实时、充分地远程（比如跨越城市）了解前方的现场景象、环境参数和人员状态等信息，通过对比云端数据库和内存数据库、AI辅助识别和人工综合分析，快速获得最合理的行动方案，并将行动指令及时传给前方人员，实现盲人导航功能。同时，也用语音互动对话提供人文关怀、情感寄托。**结果**　本文首次提出了"远程虚拟陪伴概念"，并演示了相应的硬件和软件以及多种生活场景下的测试效果。除了可以实现基础的导航功能，比如帮助视觉障碍人群完成在超市购物、咖啡厅寻座、街道行走，以及完成更加复杂的拼图、日常娱乐打牌功能，还可以满足骑行等速度相对快的移动指令要求。**结论**　实验结果表明，这种"远程虚拟陪伴"装置适用大量场景和需求，不仅可以用于视觉障碍人群出行、购物、娱乐，也可用于陪伴老人出行、辅助野外探险或旅行等，具有广泛的发展和应用前景。

关键词　人工视觉辅助，远程虚拟伴侣，触觉代码，盲人，导航
中图分类号　Q6，TN710　　　　　　　　　　　　　**DOI**：10.16476/j.pibb.2023.0053