HUMAN-COMPUTER CHINESE SIGN LANGUAGE INTERACTION SYSTEM

Xu Lin
Harbin Institute of Technology, P.R. China

Gao Wen
Harbin Institute of Technology and Chinese
Academy of Science, P.R. China

ABSTRACT

The generation and recognition of body language is a key technologies of VR. Sign Language is a visual-gestural language mainly used by hearing-impaired people. In this paper, gesture and facial expression models are created using computer graphics and used to synthesize *Chinese Sign Language* (CSL), and from it a human-computer CSL interaction system is implemented. Using a system combining CSL synthesis and CSL recognition subsystem, hearing-impaired people with data-gloves can pantomime CSL, which can then be displayed on the computer screen in real time and translated into Chinese text. Hearing people can also use the system by entering Chinese text, which is translated into CSL and displayed on the computer screen. In this way hearing-impaired people and hearing people can communicate with each other conveniently.

1. Introduction

VR is a technology that explores how to realize an ideal interaction between a computer and a human being. When a computer possessing real intelligence interacts with a human being, it must understand and use human language, i.e., natural language. Sign language is mainly used by hearing-impaired people to communicate with each other. We thus say that it is the natural language of hearing-impaired people. Currently computers are so popular that we must take the hearing-impaired users into account, making computers capable of understanding and expressing output in sign language.

The study of sign language synthesis and recognition is not only to give hearing-impaired people an intelligent human-computer interface, but it also serves the need of computer-aided education. When a human interpreter is absent or the conversation is private, a sign language interaction system can serve as a translator between a hearing-impaired person and a hearing person.

The first study of Sign language synthesis can be traced back to 1982 [1]. Because of the limitations of computers at that time, sign language synthesis was limited to describing posture. Then in the 1990s many research groups such as those associated with Hitachi Company, Trinity College, and MIT began developing systems and programs [2-4]. However, the human body model in the system created by Hitachi is not lifelike. The software at Trinity College analyzes the linguistic features of ASL, but it is still in the development stage. As for the study on sign language recognition, more is being done, but most operate at the word recognition level; not the sentence level. The aim of this paper lies in the creating of text-driven CSL synthesis and data-glove-based CSL recognition, which are the main components of a human-computer CSL interaction system.

Text-driven CSL synthesis is driven by text sentence, and the output is that of a virtual human pantomiming the corresponding CSL. In data-glove-based CSL recognition subsystem, a person with a data-glove pantomimes CSL with the computer recognizing the meaning and displaying the corresponding Chinese text on computer screen.

2. Chinese Sign Language

CSL is a visual-gestural language that transmits the message by means of hand movements, facial expression changes, and body movement. Gesture and facial expressions are important grammatical components of CSL. Sign language is characteristic of area. Although CSL and Chinese originated from the same national culture, CSL is different from Chinese and other country's Sign language.

CSL has four major formational attributes: (1) configuration of the hands, (2) location of the hands relative to the body, (3) movement of the hands and arms, and (4) facial expression. Each attribute comprises a large inventory of discrete representatives.

3. CSL Model

3.1 GESTURE MODEL

Physiological research shows that the hand is composed of skin, ligaments, tendons, muscles, and bones. Bones are linked at joints. Muscles attach to bones, and their contraction and expansion force bones to move about joints. Movements of fingers affect and constrain each other, i.e. the movement of some fingers restrict the movement of other fingers.

As a result of the anatomy of the hand and arm, we define three joints for each finger and four degrees of freedom (DOF) for the thumb, three DOF for the other finger, two DOF for the wrist joint, two DOF for the elbow joint, and three DOF for the shoulder joint. Thus we get a total of 18 joints and 23 DOF for one hand and arm, as shown in Figure 1.

Each gesture is composed of a sequence of actions, described by the following data structure:

```
typedef  STRUCT tagFinger
{ //right hand parameters
    short RGZ11_z RGZ12_z RGZ13_z RGZ13_x
    short RGZ21_z RGZ22_z RGZ23_z RGZ23_x
    short RGZ31_z RGZ32_z RGZ33_z RGZ33_x
    short RGZ41_z RGZ42_z RGZ43_z RGZ43_x
    short RGZ51_ z RGZ52_z RGZ53_z RGZ53_x
    short RGW_z RGW_x
    short RGZ_x RGZ_y
    short RGJ_z RGJ_x RGJ_y
//left hand parameters
    short LGZ11_z LGZ12_z LGZ13_z LGZ13_x
    short LGZ21_z LGZ22_z LGZ23_z LGZ23_x
    short LGZ31_z LGZ32_z LGZ33_z LGZ33_x
    short LGZ41_z LGZ42_z   LGZ43_z LGZ43_x
    short LGZ51_z LGZ52_z LGZ53_z LGZ53_x
    short LGW_z LGW_x
    short LGZ_x LGZ_y
    short LGJ_z LGJ_x LGJ_y
} Finger;
```

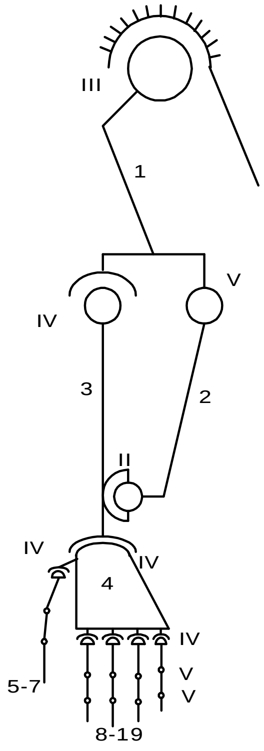


Figure 1. Machine Model of Human Hand.

Each parameter in the data structure corresponds to the DOF of each joint, which is described by various angles. Each action in CSL corresponds to a single state. Each sign word in CSL corresponds to a gesture and a facial expression. One CSL sentence is composed of a sign word sequence according to CSL grammar.

We use key state method to control the movement of the hand. The key state method involves selecting the states reflecting the essential features of actions to describe gestures. We can illustrate key state method as follows.

Suppose the movement trajectory is shown in Figure 2. The idea of key state method is that we can select states A, B, C to describe the whole movement of the hand. Each sign word corresponds to a number of states. We represent each sign word in sign database by an *N*-tuple $(N, J_1, J_2, \cdots, J_N)$, where *N* indicates the number of states corresponding to the sign word corresponds, and J_i indicates the i th of *N* states.

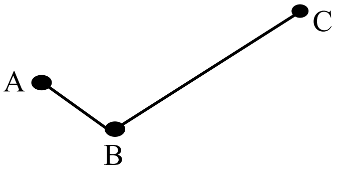


Figure 2. Illustration of the State-Selecting.

3.2 PHYSICS-BASED FACIAL EXPRESSION MODEL

Muscle pulling produces facial expression. Ekman and Friesen present famous *FACS* [5], which describes 44 independent movement "AUs".. Muscles and AUs are closely related. Ekman and Friesen also studied and identified six types of basic facial expressions, representing anger, abhorrence, fear, happiness, sadness and surprise respectively.

In this paper we take a simplified physics-based human face model. The AU is used to define the change of facial expression. An uneven 3D mesh model is created on the skin, the points of which concentrate on the areas where there are great changes in the facial expression. Less changing areas are covered with large polygons.

Muscles are the source of facial action while facial expression is being produced. In the model each relatively independent skin deformation created by one or several bundles of muscles is described by an AU. Each muscle can be simplified to a muscle vector, which has its own contract direction and action range. In addition, we define a group of characteristic points on the skin that consist of mesh points fully reflecting expression characteristics when an expression alters. The 3D physics-based human face model was created after we developed a skin model, a muscle model, and characteristic points on the skin as shown in Figure 3.

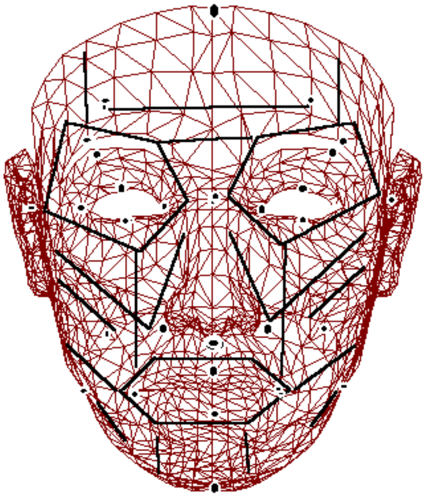


Figure 3. Model of facial expression.

The AU is used to define the change of facial expression, as mentioned above, while each AU is created by contraction of one or several bundles of muscles. Each basic expression can be composed of several AUs combinations.

Anger	=	AU25 +AU4 +AU9
Abhorrence	=	AU12 +AU4 +AU9
Fear	=	AU12 +AU1
Happiness	=	AU6 +AU12
Sadness	=	AU1 +AU15
Surprise	=	AU26 +AU1

Here we consider facial skin as an ideal elastic object, which produces no inelastic external deformation. Expansion and contraction of muscles can be simulated by adjusting the coordination of human face characteristic points. Based on these six types of basic expressions, compound expressions can be produced.

Any expression related to every sign word can be obtained by weighting and merging seven basic expressions including the neutral expression. For example:

<neutral×Happy×Surprise×Anger×Horror×Disgust×Sorrow

4. Human-Computer CSL Interaction System

The human-computer CSL interaction system is composed of CSL synthesis subsystem and CSL recognition subsystem, by which hearing-impaired people with data-gloves pantomime CSL. That CSL is displayed on the computer screen in realtime and translated into Chinese text. Hearing people may enter a Chinese sentence, which is translated into CSL and also displayed on the computer screen. In this way hearing-impaired people and hearing people can communicate with each other conveniently.

5. Text-Driven CSL Synthesis Subsystem

5.1 ARCHITECTURE OF TEXT-DRIVEN CSL SYNTHESIS SUBSYSTEM

The CSL synthesis subsystem mainly consists of an input module, a Chinese language analyzer, a transformer, a CSL synthesis module, and a CSL display module. Figure 4 illustrates.

The text input uses any system that can enter a Chinese sentence or paragraph. The Chinese analyzer analyzes the syntax and morphology of the entered Chinese sentence. The transformer translates each word and punctuation into the form of internal code of CSL word. The CSL synthesis module synthesizes gestures and corresponding facial expression. The CSL display module outputs a sequence of 3D animated graphics, driven by a set of parameters.

5.2 ACQUISITION OF GESTURE PARAMETER

In the synthesis subsystem a special-purpose visual interactive tool is developed to obtain the gesture parameters. It can display the configuration and position of the hands controlled by the parameters entered. The state of the hands can be obtained by continuously adjusting the parameters so as to change the hands configuration and position.

There are a total of 3300 CSL sign words [6]. We selected 2200 frequently used words according to the *Current Chinese Frequency Dictionary* [7] as well as 30 Chinese phonetic letters in the *Sign Word Dictionary*. We invited a special CSL teacher to evaluate and possibly modify all the gestures.

6. Sentence-level CSL Recognition

We use two Cyberglove data-gloves produced by Virtual Technologies as the input device. The parameters and principles can be found in Ref. [8]. The CSL recognition subsystem consists of the anticipation module, the *Semi-Continuous Hidden Markov Model* (SCHMM), the recognition module, the CSL display module, and the output module as shown in Figure 5.

To reduce the influence between two adjacent words in the SCHMM recognition module, we propose the following improved SCHMM recognition algorithm:

1. There is no “prefix” in front of the first word in a sentence, so an isolated sign word recognition method can be used, and the Viterbi [9] decode result is recorded.
2. After recognizing a sign word in front of some divided word, the point v^* and the Viterbi decode result is recorded, we first learn the state q corresponding to the last smooth section of gesture data from the adjacent sign word in front of v^*. Then attach the observed probability density of the state q to the lead candidate sample as the initial state and revise the state transition probability matrix A. Finally, recognize the next adjacent sign word of the divided word point v^* using an HMM sample with “prefix”, recording the Viterbi decode result.
3. Repeat Steps 1 and 2 until the recognition of the whole sentence is complete.

The recognition algorithm can eliminate, to a high degree, the influence on the recognition rate by the transition section of the two adjacent words.

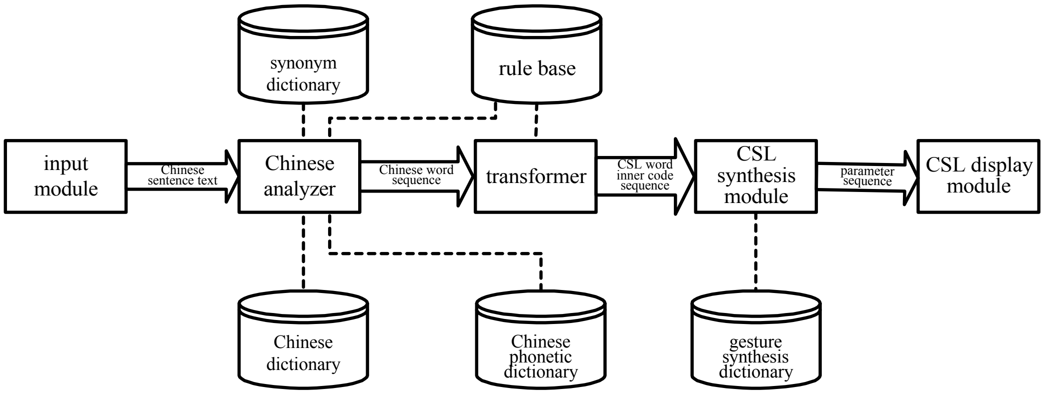


Figure 4. Configuration of the CSL Synthesis Subsystem.
]

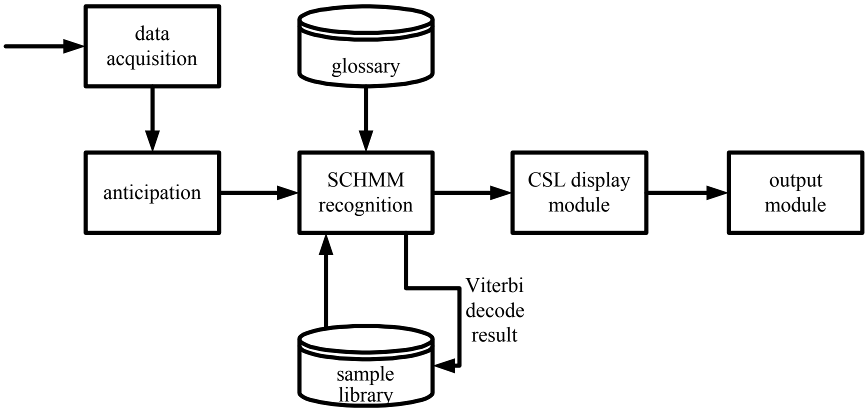


Figure 5. SCHMM Based Sentence-Level Gesture Recognition Architecture.

7. Experiment and Conclusions

We have implemented a human computer CSL interaction system. When we input Chinese, the synthesized virtual 3D human successfully pantomimes the corresponding CSL. When we input CSL by the data-gloves, the corresponding Chinese text is displayed on the screen. The system operates in realtime.

We pantomimed 120 CSL sentences with data-glove, and the recognition rate was 81.7%. We input 2200 Chinese words, and the virtual human pantomimed the corresponding sign words correctly. Here we give some running examples.

Figure 6 shows a sequence of images that represents the sign word “hello. It consists of two actions and the expressions are neutral. The “hello” is described as follows: The index of one hand points to the other person; then make a fist with the thumb stretched upward. Figure 7 shows the sequence for the sign word “guard”. It consists of four actions: The sign word “guard” is described as follows: stretch the index finger of one hand and point at the temple, show the alert appearance; then spread both palms to express defense. Figure 8 illustrates the entire sentence ‘Welcome to Harbin Institute of Technology “.

REFERENCES

[1] M Shantz and H Poizner (1982), “A Computer Program to Synthesize American Sign language”, Behavior Research Methods and Instrumentation, Vol.14, No.5, 1982, pp. 467-474.

[2] Justine Cassell, Catherine Pelachaud, Norman Badler and Mark Steedman, Animated Conversation, Ruled-Based Generation of Facial Expression, Gesture & Spoken Intonation for Multiple Conversational Agents SIGGRAPH 94, Orlando, Florida, July 24-29, 1994 pp. 413-420.

[3] J Loomis, H Poizner and U Bellugi, Computer Graphics Modeling of American Sign Language , Computer Graphics, Vol.17, No.3, July, 1983, pp. 105-114.

[4] Masanu Ohki, et. al “Sign Language Translation System Using Pattern Recognition and Synthesis , Hitachi Review, Vol. 44, No. 4, 1995, pp. 251-254.

[5] P. Ekman and W. V. Friesen, “Facial Action Coding System , Consulting Psychologists Press Inc.,

577 College Avenue, Palo Alto, California94306.

[6] *Chinese Sign Language*, Hua Xia Press, Bei Jing, 1995 (in Chinese).

[7] *Current Chinese Frequency Dictionary*, (in Chinese).

[8] B. Pang. Data-glove Based Chinese Sign Language Recognition. Master Thesis. Harbin Institute of Technology. 1999.7

[9] L. R. Rabiner, B. H. Juang, “An Introduction to Hidden Markov Models , IEEE ASSP Mag. 1986, Vol.3, No.1, pp. 4-16.

BIOGRAPHIES

XU Lin is a Ph.D. candidate at Harbin Institute of Technology. Her main research interest is intelligent human computer interface technology.

Contact information.
XU Lin
Vilab lab, 321#
Computer Science and Engineering Department
Harbin Institute of Technology
Harbin, 150001, China
Phone +86-451-6416485

GAO Wen is the head of the Institute of Computing Technology, Chinese Academy of Sciences, and chief expert of the intelligent computer portion of the National 863 Project Committee. He is also professor and doctorate supervisor of Harbin Institute of Technology, honors professor of Hong Kong City University. He is joint professor of Tsinghua University, University of Science and Technology of China, University of DaLian Science and Technology, and Xi’an Jiao Tong University as well. He is also the Editor-in-Chief of the Chinese Journal of Computers, committee member of Journal of Software, committee member of Journal of Computer Aided Design and Graphics. He earned his Ph.D. degree in computer application from Harbin Institute of Technology in 1988, and earned a Ph.D. degree in electronics from Tokyo University in 1991. From 1991 to 1995 he was a visiting professor at Tokyo University, University of Carnegie Mellon, and MIT. His major research fields are artificial intelligence and multimedia technology. He has been working for research and exploration in the areas of image data compress, medical image database, bilocomotion model computer vision, multimedia parallel acceleration system, multifunction perception system, and so on. He has published more than 116 scientific research papers and 7 books.

Contact information
GAO Wen
Institute of Computing Technology, CAS
Beijing, 100080, China
Phone +86-10-62565533
E-mail: mailto:wgao@cti.com.cn

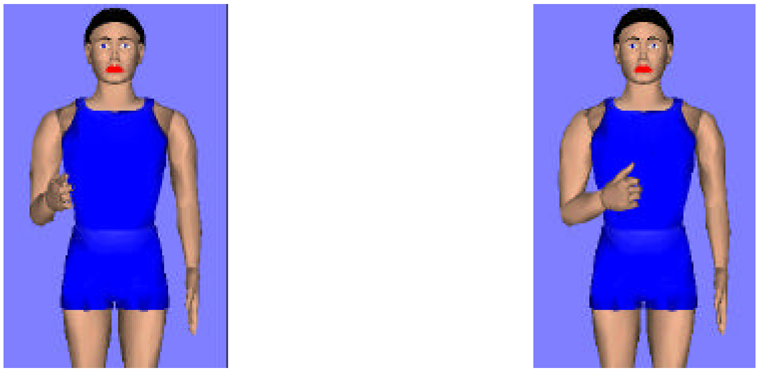


Figure 6. Gesture Pictures of “hello!”

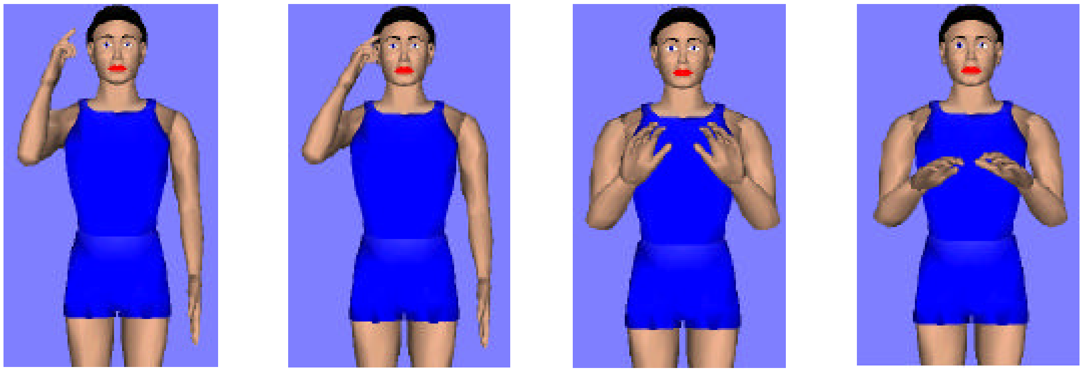
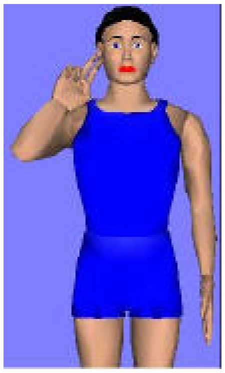
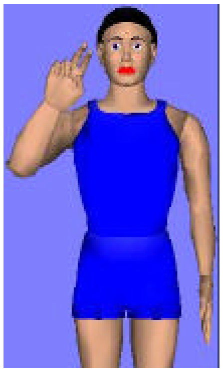


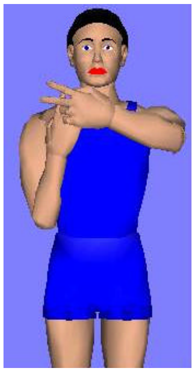
Figure 7. Gesture Pictures of “guard”.



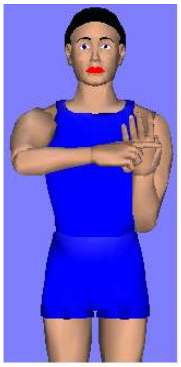
Harbin(1)



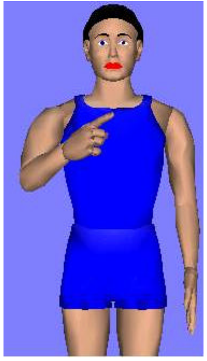
Harbin(2)



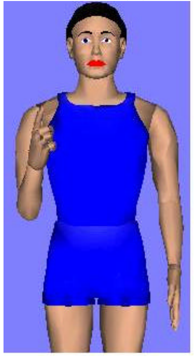
Technology(1)



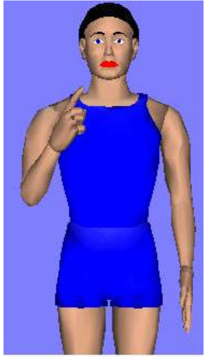
Technology(2)



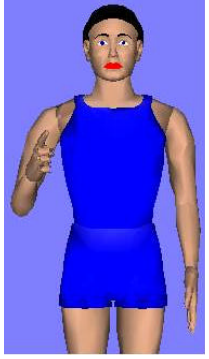
Institute(1)



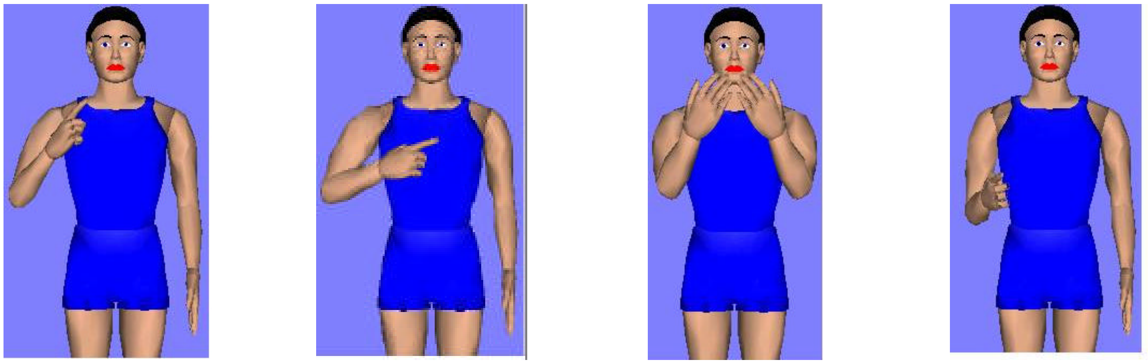
Institute(2)



Institute(3)

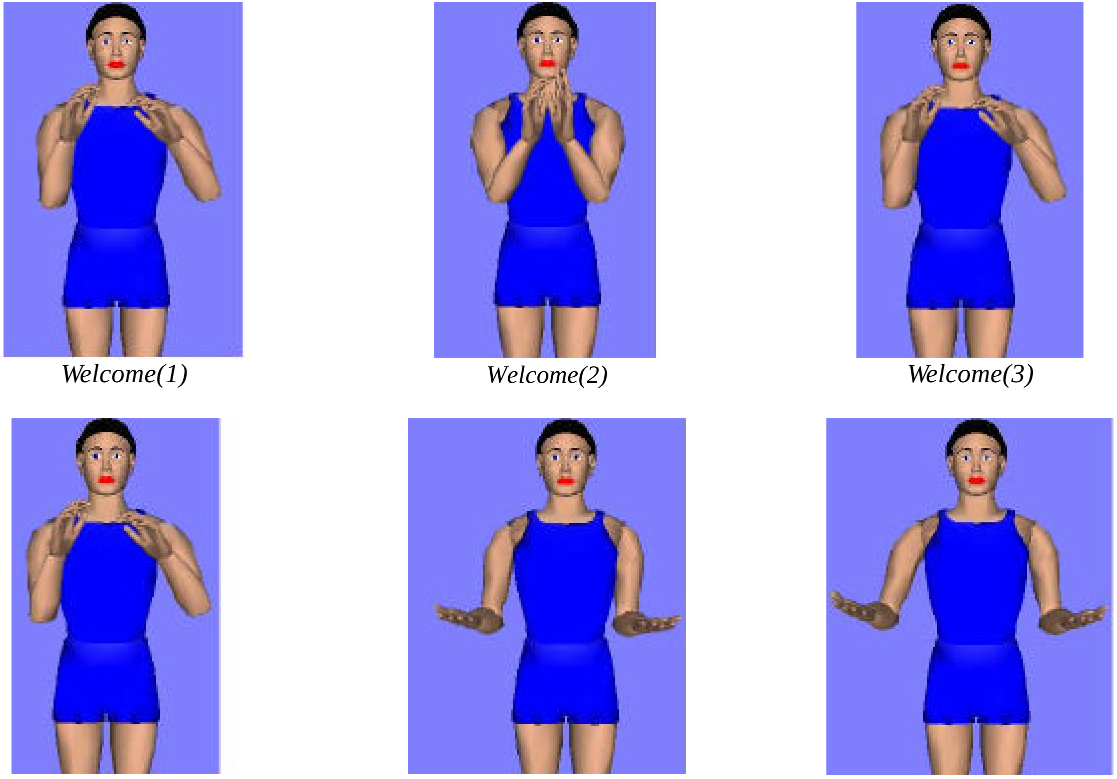


Institute(4)



Institute(5) *Institute(6)* *Institute(7)* *You*

Figure 8. Gesture Pictures of "Welcome to Harbin Institute of Technology" (cont on next page).



Welcome(1) *Welcome(2)* *Welcome(3)*

Welcome(4) *Welcome(5)* *Welcome(6)*

Figure 8. Gesture Pictures of "Welcome to Harbin Institute of Technology" (cont).