CREATION OF INFORMATION TECHNOLOGIES FOR SIGN LANGUAGE MODELING AND LEARNING

I. Krak, I. Kryvonos

V. M. Glushkov Institute of Cybernetics National Akademy of Sciences
Kyiv, Ukraine
E-mail: yuri.krak@gmail.com

In the article a conception is developed and new information technology is proposed for sign language modelling on the base of human spatial model. For the transmission of movements of the real human-informant of sign language on a spatial model technology of motion capture is used. The algorithm of synchronization of vocal information with its visualization on the human face is developed. Efficiency of the developed information technology is shown on realization of Ukrainian sign language. The proposed approach is carried by universal character and can be used for the modelling of other sign languages**.**

Keywords: deaf people, design, sign language, information technology

Introduction

The development of science, widespread use of computers, multimedia and Internet technologies make up satisfactory conditions for creation of computer-aided systems for communication with deaf people, which would use sings and language(s) comprehensible for everyone [1].

The primary mean for communication between deaf people is a sign language. The sign language has national traits and there are two sign languages, different in grammar and sets of signs, used by the deaf:

- a common sign language, which is typically used in an everyday communication; its grammar differs largely from the spoken language;
- a calculative sign language, which is used for official communication; it has some aspects of the common one as well as fingerspelling; it does not have its own grammar and follows the natural spoken language.

Comprehension of a natural language by watching lips, “lips-reading”, is an important skill for a deaf, because people generally do not know the sign language. With regard to this fact, the solution capable of reading the text being pronounced by watching lips can be regarded as an alternative technology for facilitation of communication with people having difficulties in hearing or seeing. Besides, the visual-based speech recognition is an additional independent source of information for the problem of speech recognition itself, and can be used for improvement. The synthesized equivalent of the text received by lips reading will facilitate the communication for people having difficulties seeing.

From the practical point of view, teaching systems of sign language that use 3D human model are very promising. Since the information is transmitted by the mean of arms movements, mimics and articulation it is necessary to research the process of construction of a sign language sentence as well as the synthesis of the elements in order to get a good understanding of the subject. The problem of description of human movements is complex enough with a high percentage of fuzzy knowledge about human body and its physiology [2, 3]. Thus it is important to conduct a research of the process of how the gestures are formed from the viewpoint of formalization for the problem of modeling using 3D model and for the problems of gestures analysis and synthesis.

In the paper a complex informational technology aimed to facilitate the non-verbal communication between common people and deaf people, people having hard difficulties in hearing or having damaged hearing is created. Will note that on the basis of the offered technology can be realized other sign language.

Therefore, the following problems statement has been formed: to create of computer-aided informational technologies for communication with deaf people, with intent for implementation, that would provide the following features: 1) creation of a system for sign language fingerspelling units (dactyls) modeling based on 3D model of human palm; 2) gesture synthesis for the common sign language and calculative sign language on a 3D human model; 3) representation on a spatial (3D) human model the pronunciations process with regard to emotional components; 4) lipsreading analysis and modeling.

3D model of a human hand and fingerspelling process animation

To teach fingerspelling, technology that uses a three-dimensional model of a hand based on a informational-parametric model has been developed. The technology allows observing hand from different viewpoint during the learning process, show sequence of letters etc. The main window is shown on a *Fig. 1*.

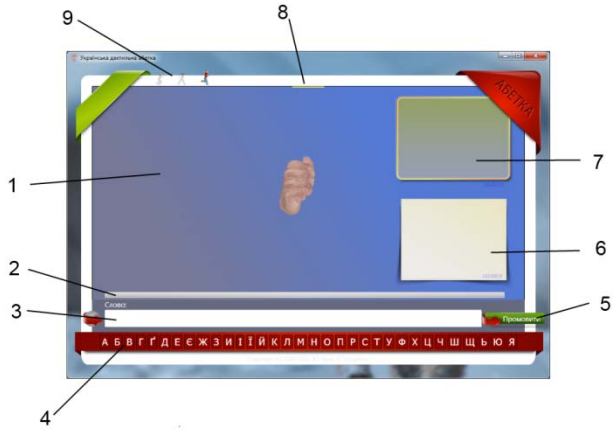

Fig. 1. The main window of a program for fingerspelling alphabet modelling

At this point the numbers mean: *1* – area of displaying fingerspelling alphabet; *2* – panel of displaying playback progress of letters or words; *3* – input panel for words; *4* – list of letters; *5* – button «spell», the process of fingerspelling of input word begins when the button is clicked; *6* – panel to display the verbal description of a hand configuration that correspond to the current displayed letter; *7* – panel to display written letter and a picture that correspond to the current displayed letter; *8* – indicator of a location of a hand rotation; *9* – define the pace of fingerspelling.

The main features of the program: a) Changing of a view angel. The use of a three-dimensional modeling enables the possibility to examine hand model from different viewpoints. That would be impossible using video materials. The range of changing an angle of viewpoint vary to 80° right/left; b) Presence of pictures which are associated with particular letter for whole alphabet (this panel can be hidden and shown back); c) Verbal description of hand configuration which is shown; d) Presentation of dactyls is performed by selecting a particular letter from a list using mouse or by pressing letter-button on a keyboard. If user wants to repeat, press space. This feature allows to implement interactive learning process, when thee right arm (trained) is in the free position and the left realize interactions with the program; e) Changing the pace of the animation. Three pace modes (slow, medium, and

fast) are implemented in the program for the different needs of the learning process (repetition after the model, recognition of the foregoing, etc.); f) Fingerspelling of a word. This feature allows entering words into input panel and observing process of fingerspelling of a word. That allows not only learn separate letters but also learn how to spell whole words; g) Verification mode. The program has a feature to "hide" the panel of the verbal description of a hand configuration and the panel with written letter and a picture. That allows conducting examination of knowledge displayed letter (hand configuration).

Based on this technology, training programs for any one-handed fingerspelling alphabet can be created. There are currently developed programs for Ukrainian, Russian, Polish, Azerbaijan and American fingerspelling alphabets.

Complex technology for sign language modeling

The complex information technology will provide the following features:

- a module for translation of the normal text into the sign language (text-to-gesture); the module will provide pronunciation animation of a common and official sign languages by presenting the output on a 3D human model;
- mimics and animation (with regard to emotional components) during the pronunciation process;
- lipsreading module for recognition of the text being pronounced.

For the implementation of the suggested concept of computer-aided non-verbal communication, a series of research works have been made and the appropriate software has been developed. For the 3D sign language animation synthesis, the geometrical classes of vector-based gestures are described. These classes were formed using motion capture technology. Motion capture is a technology for retrieving real-world 3D coordinates using multiple video streams recorded from different viewpoints. Then the coordinates are used to determine values in the mathematical model. The key frames are determined by using tracking technology [4].

For the storage of a gesture, the BVH file format was used. It allows the gesture to be applied on a virtual human (e. g. using Character Studio module in 3D studio MAX or using Poser software). The suggested implementation of motion capture technology contains the following stages: 1) A person showing a particular gesture is recorded full-face, from the left and right side views; 2) The video streams are processed: arms coordinates are detected and the motion is tracked; 3) Based on the position of arms obtained on the previous stage, the BVH is formed for further synthesis of 3D animation; 4) The BVH is applied on a virtual human for creation of the animation process (using Character Studio in 3D studio MAX or Poser).

For the input text preprocessing, the appropriate informational technology was created, which considers the stress location for each word, specifies its normalized word form; contains synonyms and idioms. The model is represented as a set of tables in a relational database along with a set of stored procedures which implement all the required functionality. For the implementation of visualization and pronunciation feature of a custom text, the appropriate synthesizer has been created. It allows creating the voice equivalent of a custom text using different voices and voice characteristics (volume, distance). The synthesized allows to visualize the process of pronunciation by showing 2-dimentional visemes as well as 3D ones.

For the complex verification of the suggested technology the appropriate software has been created (Fig. 2). It is used for translation of a custom text into the calculative sign language.

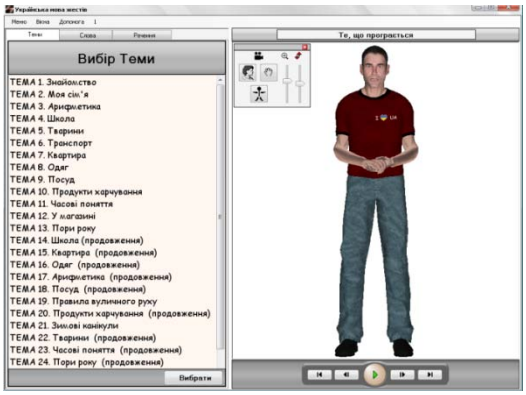

Fig. 2. Computer system for sign language modelling and learning

The software uses the following algorithm for the sign language synthesis: 1) a speech equivalent is synthesized for the input text; 2) the input text is parsed into words; 3) a speech equivalent is synthesized for the input text; 4) the input text is parsed into words; 5) for each word its normalized form (infinitive) is found by performing a look up in the database; 6) for each normalized word form a gesture is looked up (represented as a sequence of movements); 7) in case the gesture is not found, the word will be shown using dactyl alphabet.

The 3D model displays the gesture accompanied by the speech synthesized.

Conclusions

The work suggests a complex of informational technologies for non-verbal communication with deaf people using sign language. Further development will be targeted at support for the full set of signs in the Ukrainian sign language. The great attention will be consideration on the use of the created systems for modeling Russian, Polish and other sign languages.

Bibliography

1. *Kryvonos Iu. G., Krak V., Barmak O .V* . Information technology for Ukrainian sign language modeling // Artificial Intelligence. 2009. № 3. P. 186–197.
2. *Stokoe, W. C.* Sign language structure: An outline of the visual communication systems of the american deaf / William C. Stokoe. Univ. of Buffalo, 1960. P. 80.
3. *Voskresenskii A. L., Khakhalin G. K.* Multimediinyi tolkovyi slovar russkogo gestovogo yazuka // Komputernaya lindvistika I Intellektualny tehnologii: Trudy mesgdunarodnoi konf. «Dialog 2007». 2007. M. : RGGU. P. 658.
4. *Avidan S.* Support vector tracking // Proc. IEEE Conference on Computer Vision and Pattern Recognition. Kauai. Hawaii, 2001. V.1. P. 84–191.