COMPUTER TECHNOLOGY FOR SIGN LANGUAGE MODELLING

**Iurii Krak, Iurii Kryvonos, Olexander Barmak,
Anton Ternov, Bohdan Trotsenko**

Abstract: The work suggests a complex informational technology for sign language modeling (for Ukrainian sign language implementation). The technology and corresponding mathematical model for formalization of sign language primitives, dactyl alphabet and face mimics primitives have been suggested. The corresponding software applications have been created which allow gestures as well as face mimics storage and reproduction. The results have been analyzed and the roadmap for further research has been created

Keywords: Sign language modeling, Computer technology, Face mimics modeling, Ukrainian sign language

ACM Classification Keywords: I.2.8 Problem Solving, Control Methods, and Search H.1.1 Systems and Information

Conference: The paper is selected from XV^th International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.

Introduction

Significant limits of existing means for sign language reproduction give a reason to create more powerful algorithms which would allow to create computer-bases systems for sign language tutoring with aim to facilitate communication with deaf people and people having hearing disabilities. The authors suggest [1,2] a concept of informational technology of non-verbal communications with deaf people. Using a virtual 3d human model, the complex informational technology includes functionality to reproduce sign language gestures, gestures of dactyl alphabet, proper face mimics during pronunciation.

Implementation of the feature to generate pronunciation animation for a custom gesture requires having proper informational and mathematical models. Therefore, the following problem statement has been formulated:

1) To create informational and mathematical models for the formal sign language morpheme description and for the face mimics synthesis;
2) Using the created models, to create a technology along with corresponding software for the gesture capturing, storage and reproduction, with support for the face mimics.

Model of sign language morphemes fixation

Process of sign language reproduction on a 3d human model can be regarded as animation of the corresponding frequency of different skeleton states.

The skeleton human model is a simplified correspondent of the human skeleton. It can be formalized as the hierarchical structure of kinematic pairs, which reflect the basic human bones.

Modern 3d animation software (Poser, 3D Studio Max) are able to generate animation using virtual skeleton and information on angles change. Therefore, for the formal description of a gesture we can use sets which reflect the simplified human skeleton and changes in time of angles between bones: $H = \{H_i : H_i = \{k, d_i, M_i \in M\}\}$ – the simplified human skeleton (bones hierarchy) and change of angles in time, where H_i – i-th bone of the skeleton ($i = 0,\cdots,N-1$, N – amount of bones in the skeleton); k – index of the parental bone; $d_i = [x_i, y_i, z_i]^T$ –

coordinate of the bone's ending point using the coordinate system connected with the parental bone; $M = \{M_i : M_i = \{order_i, \theta_i\}\}$ – the angles and order of angle application for the each bone.

For the storage of the gesture using this formalism we suggest to use the BVH[3] file format. The format was introduced by Biovision company exactly for the storage of movement. The BVH is one of the best formats for this kind of storage, with its sole disadvantage being the absence of the complete definition of the starting pose; yet this disadvantage is irrelevant to our problem. File in BVH format consists of two parts. The header contains the hierarchy of the model and the initial pose of the skeleton and description of the animatory part. The primary reason for using this file format is its simplicity and suitable for the our problem. It's very convenient that it is recognized by primary software on the market. The figure 1 shows a virtual human skeleton that is used for the formal description of the sign language reproduction.

Technology for gesture obtaining and storage

The history of 3d animation is much more than a decade old, thus the progress is essential. This work benefits from the Motion Capture technology for modeling a virtual human, which is designed for the sign language reproduction. Systems using Motion Capture appeared on beginning of 90ies of the past century. The problem of 3d animation was among the most important ones before the technology appeared. The underlying principle of the Motion Caption is quite simple: the real-world human plays the role for the virtual character. Typical Motion Capture implementation contains a number of signal emitters which are tightened to the real-world human. The information about the placement of the emitters is perceived by some detectors and thereon registered by a computer. Later on the information is processed by software, which reconstructs movements for the virtual character.

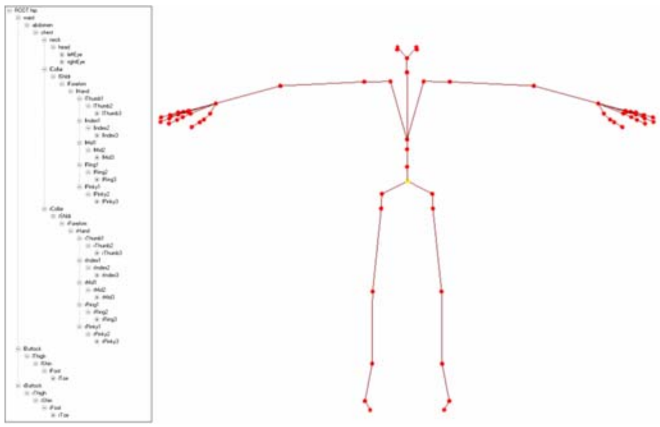


Figure 1.Virtual human skeleton

The primary disadvantage of existing Motion Capture systems is the relatively large cost of the equipments and the service for the data processing. Authors suggest much simpler implementation of the technology. Given as granted that for the sign language digitalization a relatively small amount of movements should be captured, the following is suggested:

In order to obtain a set of angles which reflect the changes of bones relatively to the initial state of the skeleton the following technological scheme is suggested:

1) Using 3 cameras a real-world human performing a custom gesture is recorded. The cameras are set on the same focus distance (L) from the object of the recording and thus fixate 3 projections: frontal, left and right (see figure 2). States of the skeleton characterize N frames of the recording (using 15 frames per second).
2) For the fixation of i-th state ($i = 1, \cdots, N$) of the skeleton we use the proper frames from the recordings (see figure 3). The skeleton is then set to correspond to the image (see figure 4).
3) Let's combine points of skeleton projection with the corresponding image of the real-world human (for the frontal, left and right frames), we obtain new coordinates: $(x_i^{front}, y_i^{front}, -1)$ – for the frontal, $(x_i^{left}, y_i^{180^\circ}, z_i^{left})$ – for the left projection and $(x_i^{right}, y_i^{180^\circ}, z_i^{right})$ – for the right projection. Where:

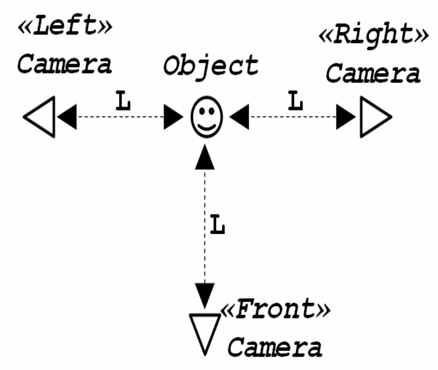
$$x_i^{front} = x_i^{180^\circ} + off_i^{front_x},\ y_i^{front} = y_i^{180^\circ} + off_i^{front_y},\ x_i^{left} = x_i^* + off_i^{left_x},\ x_i^{right} = x_i^* + off_i^{right_x},\ z_i^{right} = z_i^* + off_i^{right_z} \quad (1)$$

The shifts off_i obtained this way we will apply for all the frames which correspond to the gesture. For the more exact values off_i it is recommended for the first frame to make the real-world human stay in a T-position (or close to it).

4) For the j-th ($j = 1, \cdots, N$) frame on the pair of images (frontal and left or frontal and right) we consider points of bones connections which have actually moved ($\left(x_{i,new}^{front}, y_{i,new}^{front}\right)$ and $\left(x_{i,new}^{left\ or\ right}, y_{i,new}^{left\ or\ right}\right)$) and then computer their 3d coordinates (relatively to the center of the skeleton):

$\left(x_i^{new}, y_i^{new}, z_i^{new}\right) = \left(x_i^{old}, y_i^{old}, z_i^{old}\right)$,

$$\text{where } x_i^{old} = x_{i,new}^{front} - off_{i,x}^{front},\ \ y_i^{old} = y_{i,new}^{front} - off_{i,y}^{front},\ \ z_i^{old} = x_{i,new}^{left\ or\ right} - off_{i,x}^{left\ or\ right} \quad (2)$$



Firgure 2. Placement of camera during recording

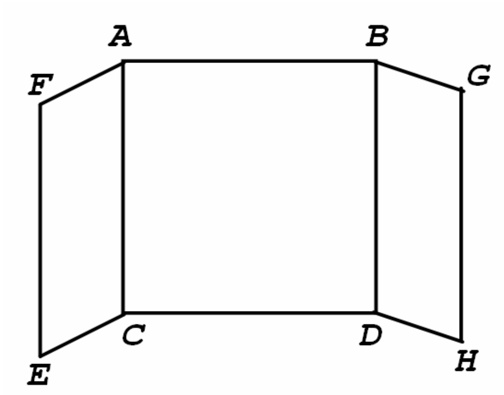


Figure 3. Planes for frames projection

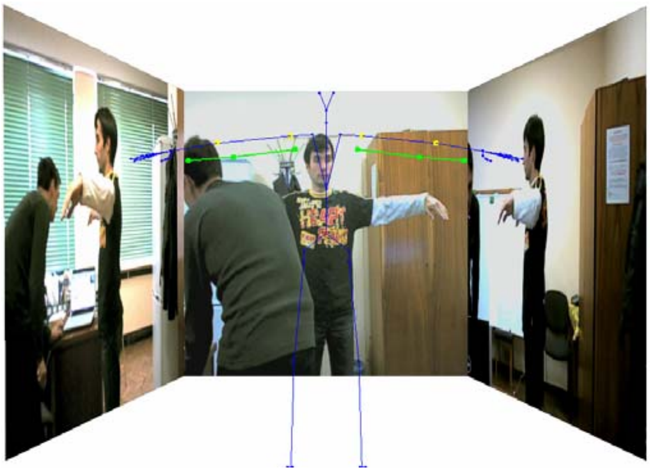


Figure 4. Working scene example for the angles detection

Then, given the 3d coordinates of the new location of the bones connections we compute the Euler angles, which is the equivalent information of the points coordinates in combination with the previous intersection location.

Consider coordinate system XYZ and two vectors in it: $r_1 = (x_1, y_1, z_1)$ and $r_2 = (x_2, y_2, z_2)$ (see figure 5). What are the Euler angles $\varphi_x, \varphi_y, \varphi_z$ for the rotation of X, Y, Z which will combine r_1 into r_2. For this:

1) Let's build a rotation matrix T, which transforms vector r_1 into vector r_2:

$r_1 = Tr_2$, where

$$T = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix} = \begin{pmatrix} \cos(\alpha_{X^{r_1}X^{r_2}}) & \cos(\alpha_{X^{r_1}Y^{r_2}}) & \cos(\alpha_{X^{r_1}Z^{r_2}}) \\ \cos(\alpha_{Y^{r_1}X^{r_2}}) & \cos(\alpha_{Y^{r_1}Y^{r_2}}) & \cos(\alpha_{Y^{r_1}Z^{r_2}}) \\ \cos(\alpha_{Z^{r_1}X^{r_2}}) & \cos(\alpha_{Z^{r_1}Y^{r_2}}) & \cos(\alpha_{Z^{r_1}Z^{r_2}}) \end{pmatrix}.$$

Given that the vectors are normalized, we obtain:

$$t_{11} = x_1x_2 + y_1y_2 + z_1z_2. \tag{3}$$

$$t_{21} = -y_1x_2 + \frac{x_1y_2}{\sqrt{x_1^2+y_1^2}} + \frac{y_1z_1z_2}{\sqrt{x_1^2+y_1^2}}. \tag{4}$$

$$t_{31} = z_1x_2 + z_2\sqrt{x_1^2+y_1^2}. \tag{5}$$

$$t_{12} = -x_1y_2 + \frac{y_1x_2}{\sqrt{x_2^2+y_2^2}} + \frac{z_1y_2z_2}{\sqrt{x_2^2+y_2^2}}. \tag{6}$$

$$t_{22} = y_1y_2 + \frac{x_1x_2 + y_1z_1y_2z_2}{\sqrt{x_1^2+y_1^2}\sqrt{x_2^2+y_2^2}}. \tag{7}$$

$$t_{32} = -z_1y_2 + \frac{y_2z_2\sqrt{x_1^2+y_1^2}}{\sqrt{x_2^2+y_2^2}}. \tag{8}$$

$$t_{13} = x_1z_2 + z_1\sqrt{x_2^2+y_2^2}. \tag{9}$$

$$t_{23} = -y_1z_2 + \frac{y_1z_1\sqrt{x_2^2+y_2^2}}{\sqrt{x_1^2+y_1^2}}. \tag{10}$$

$$t_{33} = z_1z_2 + \sqrt{x_1^2+y_1^2}\sqrt{x_2^2+y_2^2}. \tag{11}$$

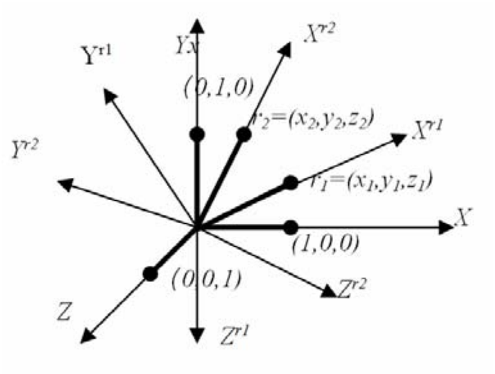


Figure 5. Coordinate systems: $XYZ, X^{r_1}Y^{r_1}Z^{r_1}, X^{r_2}Y^{r_2}Z^{r_2}$

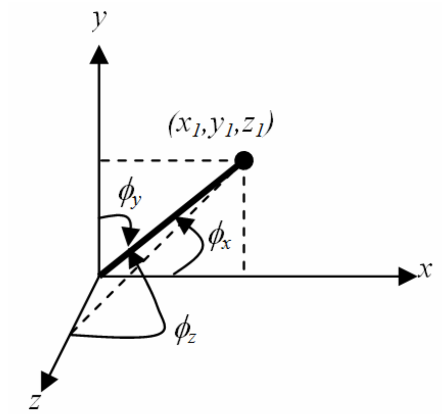


Figure 6. Directing angles for the rotation

2) We then compute Euler angles $\varphi_x, \varphi_y, \varphi_z$. It should be underlined, that the multiplication of matrices is not commutative, and thus 6 cases of multiplication precedence should be considered. Therefore, we compute angles for the following 6 orders of multiplication usage: ZYX, YXZ, YZX, XYZ, XZY, ZXY.

a) Multiplication order ZYX, ($c() = \cos(), s() = \sin()$):

$$\begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix} = \begin{pmatrix} c(\varphi_y)c(\varphi_z) & c(\varphi_z)s(\varphi_x)s(\varphi_y) - c(\varphi_x)s(\varphi_z) & c(\varphi_x)c(\varphi_z)s(\varphi_y) + s(\varphi_x)s(\varphi_z) \\ c(\varphi_y)s(\varphi_z) & c(\varphi_x)c(\varphi_z) + s(\varphi_x)s(\varphi_y)s(\varphi_z) & -c(\varphi_z)s(\varphi_x) + c(\varphi_x)s(\varphi_y)s(\varphi_z) \\ -s(\varphi_z) & c(\varphi_y)s(\varphi_x) & c(\varphi_x)c(\varphi_y) \end{pmatrix} \tag{12}$$

Then

$$\varphi_z = arctg\left(\frac{t_{21}}{t_{11}}\right) + n\pi . \tag{13}$$

$$\varphi_y = arctg\left(\frac{-t_{31}}{\sqrt{t_{11}^2 + t_{21}^2}}\right) + n\pi \ . \tag{14}$$

$$\varphi_x = arctg\left(\frac{t_{32}}{t_{33}}\right) + n\pi \ . \tag{15}$$

b) The angles for the other orders are computed in a similar way.

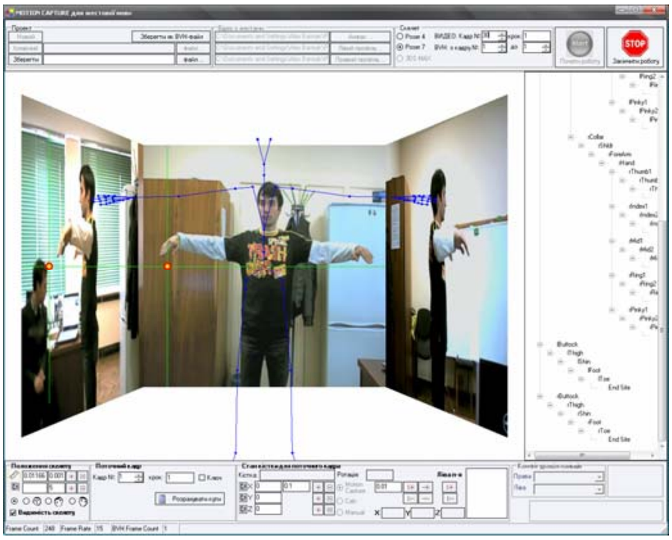


Figure 7. The application which implements motion capture technology

Model for face mimics synthesis for the pronunciation process and a technology for visemes computation.

Face mimics play a role of an additional information channel in the sign language. In order to support this additional channel and make gesture reproduction look more real, special research work has been conducted and then face visemes have been introduced. From the numerous researches [4,5] we conclude that visual alphabet contains no more than 6-15 elements. The actual number depends on the required level of natural look.

Using results of the researches on face mimics and visuals mentioned above a set of visemes has been creates for 6 vocals and 32 consonants (see table 1), as required for the Ukrainian language phonetics. Phonemes have been classified into 15 groups, excluding additional group for the idle state (no pronunciation). We define viseme as a position of lips which is naturally observed during pronunciation of a particular phoneme.

For each group of phonemes (i.e. for each viseme) frames of face mimics were computer. These frames are further used for pronunciation animation (see figure 8).

TABLE 1.

viseme	*phoneme*	*viseme*	*phoneme*	*viseme*	*phoneme*	*viseme*	*phoneme*
1	а	5	і, и	9	п, б, м	13	р
2	е	6	й	10	в, ф	14	л', р'
3	о	7	ш, ж, ч, дж	11	т, д, н, л	15	т', д', н'
4	у	8	к, г, х, ґ	12	с, з, ц, дз	16	**idle state**

3d gesture animation and face mimics reproduction model and its implementation

For the process of gesture and face mimics animation synthesis the following formal description is suggested. The formal description using proper set of parameters and algorithms. The 3d virtual human model which implements the animation and face mimics has such attributes:

Viseme	"у"	"а"	"тднл"	"л'р'"	"сзцдз"	"пбм"	с/сп
MBO							

Figure 8. Sample visemes for Ukrainian sign language

$V = \{V_i : V_i = \{x, y, z\}\}$ – set of triangles' vertices for the triangulation of the 3d human model;

$N = \{N_i : N_i = \{x, y, z\}\}$ – set of normals for the vertices; $T = \{T_i : T_i = \{u, v\}\}$ – set of texture coordinates for the vertices; $V^{ind} = \{V_i^{ind} : V_i^{ind} = \{k_1, k_2, k_3\}\}$ – set of indices which specify the order of triangles construction; $I = \{I_i : I_i = \{img\}\}$ – set of pictures for the textures

For the skeleton-based animation modeling it is necessary and sufficient to compute the new values of vertices (V). To reach this goal we suggest to using so-called skinning method. Skinning can be defined as an algorithm of computing new values of vertices as a weighted sum of points which belong to different bones of the skeleton. The model of skeleton-based animation can be formalized as follows:

$MH = \{MH_i : HM_i = \{k, \{l_1, \cdots l_m\}, d_i, Glb_i, Order_i\}\}$ – description of the simplified human skeleton (bones hierarchy) for the implementation of the skeleton-based animation, where MH_i – i-th bone of the skeleton ($i = 0, \cdots, N-1$, N – number of bones in the skeleton); k – index of the parental bone; $\{l_1 \cdots l_m\}$ – set of children's indices, $d_i = [x_i, y_i, z_i]^T$ – coordinates of the ending point of the bone in the coordinate system which center is in the beginning of the bone; Glb – vector which is used for determination of the bone's coordinates in the global coordinate system; $Order_i$ – order of rotations.

$\mathrm{Skin} = \{Skin_i : Skin_i = \{(IndexVertex_1, Weight_1), \cdots\}\}$ – set of vertices which influence the current point.

For each vertex v skinning is computed as follows:

$$v_j^{'} = \sum_{i=0}^{N} \{(v_j * IBM_{H_i} * JM_{H_i}) * JW_{H_i}\}, \quad (16)$$

where: n - number of bones related to the vertex v;

IBM_{H_i} - inverted bind-pose matrix for the bone H_i; JM_{H_i} - moment matrix for the bone H_i; JW_{H_i} - weight coefficient for the influence of the bone H_i on the vertex v.

For the modeling of animation process we use so-called morphing method. It is a way of smooth change of properties from one state to another. The states in between are called key states. Intermediate states are computed based upon key states using weight coefficients. Thus it is a key part for the animation. The morphing process for face mimics is described as follows:

Face mimics is built for a virtual 3d human model is built using a segmented (or targeted) morphing of the head in a common state and the head showing the most distinct face mimics.

The formula for relative morphing of M basic morphs is:

$$V^{'} = V + \sum_{m=1}^{M} w_m \cdot TM_m \ , \qquad (17)$$

where w_m - weight coefficients, V – mesh of the basic model without morphing, TM_m - the mesh of the distinct face mimics. The result is a linear combination of meshes.

Method of segmented morphing allows to form several facial expressions based of a relatively small amount of morphs. It enables animation as well.

Additional advantage of segmented morphing is a feature to animate jaws independently of eyes, lips and emotional expression.

For the 3d reproduction of gestures and face mimics an application has been created which implements the skeleton-based (for the gesture) and morphing animation. The application benefits from OpenGL 3d API and shows 3d human (see figure 9) using the attributes mentioned above. The application uses morphing and skinning to show animation (see figure 10).

Figure 9. 3D Virtual Human

Figure 10. Frames of animation for the gesture "you"

For testing the correctness of gestures appropriate software has been created (see figure 11):

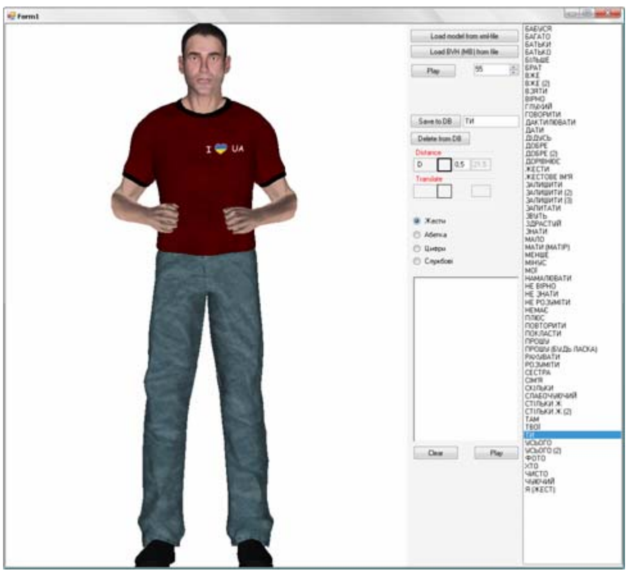


Figure 11. Software for gesture reproduction testing

Implementation of Ukrainian sign language

A specialized software has been created to implement the Ukrainian sign language. The software uses the same methodology as in Ukrainian schools for deaf children. Those underlying teaching materials are issued by the Ministry of Education of Ukraine [6] and are targeted at a beginner's level.

The functionality included in software has 3 informational categories (topics, words and sentences) and viewport with virtual human who shows gestures. The primary information category is "topics". It contains main methodological information for each lesson: points to make, skills to teach (or to learn), necessary information for comprehension and explanations. It also contains a list of gestures relative to the topic and sample sentences with the gestures (see figure 13).

The categories "Words" (see figure 14) and "Sentences" are secondary. They contain all the gestures and samples of sentences respectively.

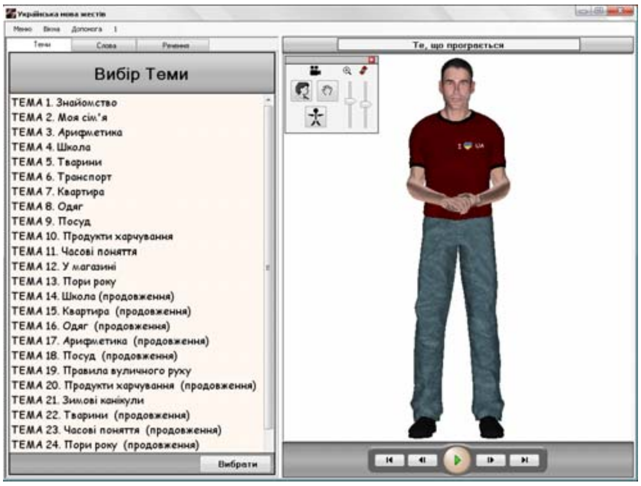


Figure 12. Software application "Ukrainian Sign Language"

Gesture reproduction viewport has a special meaning. First of all, because of the feature to demonstrate custom gesture's dynamics, that is, to show a gesture frame-by-frame. Because the gestures were digitized using recordings of native speakers and assuming that the software will be widely spread, it gives a rise to a standard of the sign language. The feature of showing a gesture frame-by-frame is a true competitor to a live teacher who needs to demonstrate gestures very slow and with lots of explanations during the teaching process. Using a single program for teaching sign language means that all the deaf children on all the territory of Ukraine will know the same gestures and thus communicate more effectively. This is how a standard of the sign language can emerge.

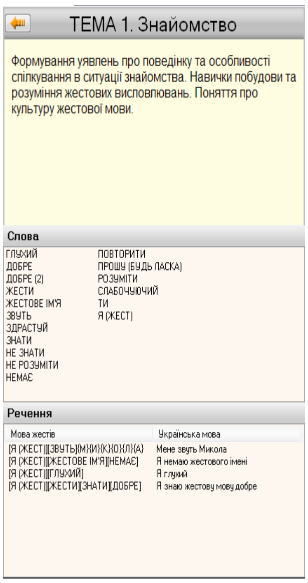


Figure 13. "Topic" category

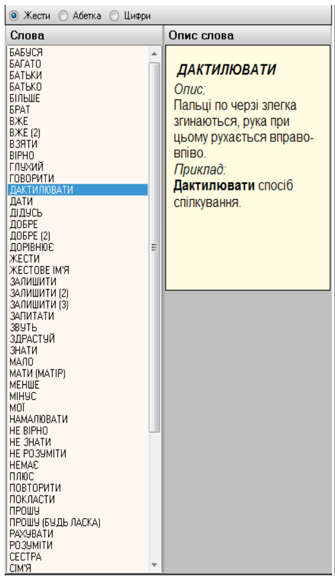


Figure 14. "Words" category "Meeting new people"

Conclusions

A set of gestures has been digitized using motion capture technology and material with recorded gestures. Gesture reproduction on a virtual human showed the technology to be able to repent naturally looking gestures, all very close to the video they have been digitized from.

Appropriate software has been created, which contains several lessons of the Ukrainian sign language. The lessons have been taken from the materials for schools for the deaf children [6].

The created software is a candidate to represent a standard of the Ukrainian sign language. It can solve the problem of differences between gestures with the same meaning, which are inevitably present as a result of teaching from different speakers.

Further development is aimed on improvement of the suggested technology:

- taking into account natural limits for each connection of bones in order to have a virtual model capable of intelligent controlling;
- fill the database with the major set of gestures of the Ukrainian sign language, that is, to create a standard of the language;
- create a mean for semantic connection of sentences in the spoken language to sentences in the sign language.

Acknowledgment

We are very grateful to Michael Rodda a Chartered Psychologist, Professor Emeritus at the University of Alberta, and Foreing Member of the Academy of Pedagogical Sciences of Ukraine for the support of our research. It is very important for us.

Bibliography

[1] Kryvonos Yu.G., Krak Yu.V., Barmak O.V., Ternov A.S. "Information technology of nonverbal communication with deaf people", Artificial Intelligence, no 3, 2008, p. 325-331. (on Ukrainian)

[2] Krak Yu.V., Barmak O.V., Gandha O.S., Ternov A.S., Shatkovskii N.N. "Computer system for communication with deaf people". Advanced Studies in Software and Knowledge Engineering, International Book Series "INFORMATION SCIENCE & COMPUTING", Number 4, Supplement to the International Journal "INFORMATION TECHNOLOGIES & KNOWLEDGE", Volume 2, 2008, p.161-165.

[3] Lander J. "Working with Motion Capture File Formats. Game Developer". Miller Freeman Inc. USA., 1998

[4] Bilodid J.K. "Modern Ukrainian language". Kyiv, Naukova dumka, 1969. (on Ukrainian)

[5] Darris Dobbs and Bill Fleming. "Animating Facial Features and Expressions". Charles River Media, 1999.

[6] Hryschenko E.S., Stiopkin V.V. "Ukrainian sign language for primary school". Kyiv, Bohdana, 2004. (on Ukrainian)

Authors' Information

Iurii Krak – the senior scientist, Institute of Cybernetics of National Academy of Science of the Ukraine, address: 40 Glushkov ave., Kiev, Ukraine, 03680; e-mail: krak@unicyb.kiev.ua

Iurii Kryvonos – the Deputy Director, Institute of Cybernetics of National Academy of Science of the Ukraine, address: 40 Glushkov ave., Kiev, Ukraine, 03680

Olexander Barmak – the senior scientist, Institute of Cybernetics of National Academy of Science of the Ukraine, address: 40 Glushkov ave., Kiev, Ukraine, 03680;e-mail: barmak@svitonline.com

Anton Ternov – the junior scientist, Institute of Cybernetics of National Academy of Science of the Ukraine, address: 40 Glushkov ave., Kiev, Ukraine, 03680

Bohdan Trotsenko – the junior scientist, Institute of Cybernetics of National Academy of Science of the Ukraine, address: 40 Glushkov ave., Kiev, Ukraine, 03680