Multiple Language, User-Friendly Sign Language Chat

K. Nurzyńska[1], S. J. F. Fudickar[2]
[1] Institute of Informatics, Silesian University of Technology, Gliwice, Poland
[2] Institute of Computer Science, University of Potsdam, Potsdam, Germany

***Abstract*—The aim of this work is to introduce a concept of a gesture based sign language chat which enables the communication of hearing impaired people over networks. Existing applications handle that task not appropriately, since either high bandwidth networks are necessary or the extensibility of the vocabulary is complicated for untrained users. The chat needs a standard camera for data acquisition. Then the gesture is recognized using sign language recognition module and mapped to the gesture ID stored in the gesture database. Only the ID is transferred via the Internet to diminish the bandwidth overload. On the participants side the gestures are rendered. Moreover, the system supports uncomplicated extension of vocabulary by the user and easy word to word translation between different sign languages.**

***Index Terms*—Chat, Image Recognition, Sign Language**

I. INTRODUCTION

In recent years the development of the Internet infrastructure and its bandwidth has allowed the computer to become one of the most powerful medium of the communication. First attempts to send the information via networks were the ground beam for email, which is widely used nowadays. Later on the communication software which enabled instant messaging was developed and when the bandwidth was sufficient the voice communication was established and finally the video stream was transmitted in the online conference systems.

The designers of the computer systems pay special attention to resemble the activities well known for the user from the daily life onto the computer screen. As a result the main computer screen looks like a desk. This assumption assists in learning new programs functionality and in general orientation how to manage the computer system. However it narrows the possibility of creative thinking as the authors concentrates on transmitting user's habits from the real world into the computer world. Therefore the computers are well suited for the majority of people.

Nevertheless the computer systems can find many new applications, which have not been possible yet. For example the communication systems for people with disabilities can be designed. In our work we focus on the hearing impaired people, who live in the world of silence. According to Omer Zak [22] around one percent of people living in European Union in 1994 suffered from this disease, excluding people suffering hard hearing problem.

Hearing impaired persons cannot hear anything. Therefore the sound-centred communication is impossible or troublesome. By this fact those people invented sign languages for communication. Sign languages consist of a grammar and a vocabulary. Usually the grammar is significantly different to the one of spoken and written languages. Whereas the vocabulary is composed of many hand gestures and hand movements which convey the most important information, but which are supported by the whole body movement and facial expression [20]. Considering the differences in the way the hearing impaired observe the world, they encounter huge difficulties while learning and using the writing language, which is so common in daily communication.

The aim of our research is to create a possibility of easy communication for hearing impaired people. In our concept we assume that to assure the comfort and functionality of the system it must support the sign language transmission via the Internet and work well also when the high quality bandwidth is inaccessible.

In Section 2 the previous systems of similar functionality are described. The Section 3 concentrates on the main idea introduction. And finally the Section 4 gives a brief conclusion.

II. DOMAIN OVERVIEW

The sign language problem is widely researched. However most researches focus the problem of sign language recognition whereas the idea of possible communication creation was lesser studied. In our case the sign language recognition is the fundamental issue for the complete system functionality yet in this work we take under consideration the overview of the whole system performance.

Among the sign language recognition systems can be distinguished three main branches of development which differ in the data acquisition manner. The systems which allows towork with large vocabularies utilize some motion capture devices, like electrical gloves, to obtain the hand shape information and motion pattern. Due to such input data acquisition there is less input error and the recognition rate is high [2][4]. Although, wearing the electrical devices is inconvenient, not to mention the costs. Other systems utilize cameras with additional markers. The camera records the video and in this case the most difficult task consists of hand gesture recognition from the image data. The markers are used to improve the hand detection, as usually by the markers the coloured gloves are understood like was used in the work of Grobel [5][6]. Due to the hand shape description error the performance of such systems is worse. Finally, there are systems which also exploit camera but do not impose the user to wear anything [17]. In this case the recognition task becomes very difficult as the skin colour must be recognized, like described in the work [12]. For movement description and recognition the Hidden Markov Model (HMM) are

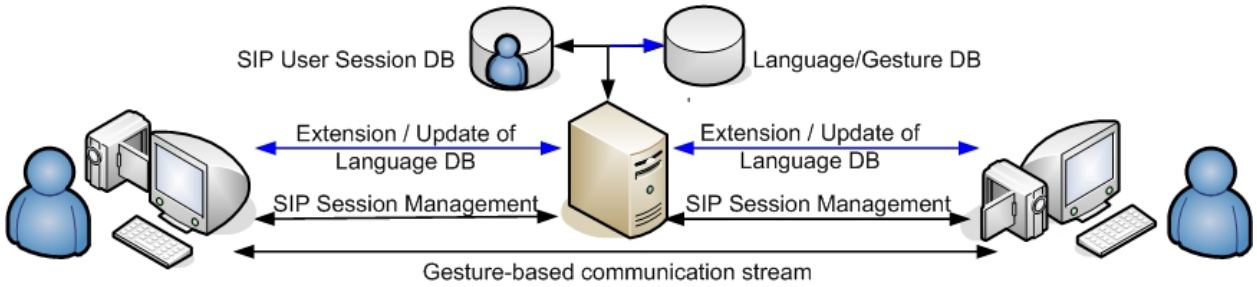

Figure 1: Network flow including the communication and the extension of language libraries

exploited [9][18][21] however the template matching [7], neural networks [8], and statistical analysis [10] were also researched.

There have been also discussed some systems taking under consideration the problem of communication of hearing impaired people. There have been researched two different approaches.

The first approach concentrates on the recording, transferring and the presentation of video-streams. Thereby the pre- and post-processing of this approach is high-performing, since an image based sign-recognition is not necessary. A fundamental characteristic of the video-streams is the high amount of data. The availability of internet connections supporting high speed and throughput to ensure an adequate transfer of the video based recorded signs is a resulting necessity. By the fact, that such high speed throughput internet access must be available to all participants of a communication, this approach is not optimal.

Another approach that focuses on the reduction of the data amount in the networks was introduced by Ohene et al. [14]. The Mak-Messenger allows the manual selection of images, which predefined symbols are then transferred to the other participants of a communication afterwards. On the receiver site an image that corresponds to the selected sign, is presented. Thereby the requirements for the utilized networks can be reduced tremendously. This approach lacks in several aspects like extensibility of supported languages, as well as usability. Regarding usability aspects of a communication platform for hearing impaired the manual selection of single gestures is not adequate, since the quantity of gestures is comparable with words included in spoken languages. Since each word/gesture is represented by the image as well as a representing code, the extension of the vocabulary results as a problem. Therefore an adequate image and the description must be available on all instances of the application, which could be just achieved through a centralized extension. As a result the grade of complexity for extending a vocabulary is oversized.

III. SIGN LANGUAGE CHAT CONCEPT

The aim of this work is to introduce the concept of sign language chat based communication which overcomes the deficiencies of the other solutions mentioned in previous sections. In our work we concentrate mainly on the comfort of use and the reduction of network overload. We assume that current hardware development enables standard computers to process complex calculation in soft real-time therefore the high performing algorithms in the field of image processing and virtual object rendering enable a new approach to the this problem.

In this section firstly is given a brief overview of the system work-flow and then we focus on each part of the system performance. The main issues in the work are divided into following issues:

- input data recognition,
- output data rendering,
- data description,
- transfer optimalisation,
- additional features.

A. General Concept Description

The main concept is to enable the sign language communication between two or more end-users connected via the Internet. The system should allow for two work scenarios. First is responsible for new user serving which is composed of actions connected with adding the user IP and nick to *SIP User Session database* and transferring the chosen sign language database from *Gestures database* to the client side (we assume that the system may work with many sign languages databases). The second scenario enables the chat connection between signers. In this case in the first step the connection between two computers is established using the IP data from *SIP User Session database* what is followed by the chat communication. This functionality is depicted on the fig. 1.

Since the connection between two end-users is established the chat starts. The whole process of chatting is divided into some individual parts repeated for each individual sign, see fig. 2. There could be distinguished part responsible for data acquisition which allows for sign recognition and matching with the data stored in the copy of *Gesture database* on the client side. That allows for retrieving gesture ID which is transmitted on the other client end where this ID is utilized to render the gesture according to rendering description defined by the ID in the *Gesture database*.

B. Gesture Recognition

While designing the gesture recognition module most important is to make this interface user-friendly and assure its high quality recognition. There are three main attitudes which solve such a problem, as has been already mentioned. The systems which were verified for large vocabulary recognition [2][4] use motion-capture sensors, which is very uncomfortable for the user. Therefore we have decided to exploit the solution where the data is captured with a camera, for instance a web-camera, and the user just shows the gestures without any additional markers [5][6], like the coloured gloves. This imposes the implementation of the skin colour detectors to find the position of hands and head on the images [12]. However

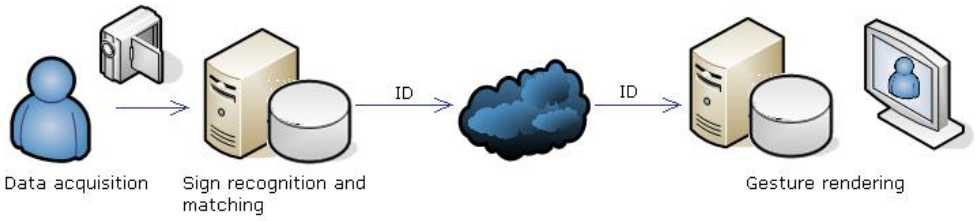

Figure 2: Information flow during the chatting phase.

this hand shape description schema never will be as exact as the data from sensors, yet according to the research of Bowden et al. [1] the characterization of the gesture just by the hand movement with the information of the starting and the ending position is adequate. The information about the movement is stored with the exploitation of the stochastic system HMM, which enable to describe the gesture movement independently from the different user speeds.

C. Gesture Description

The *Gesture database* stores data necessary in two stages of system work: gesture recognition and gesture rendering. For each stage was created a gesture description vector, both refer to the same gesture ID to assure easy data access. Additionally, the IDs of similar gestures among different sign languages databases also should be the same to enable easy translation.

The gesture description vector on the recognition needs stores the information about the hand shape and the movement of the gesture. It is worth notice that in sign language can be distinguished static and dynamic gestures. In the first group the hand shape plays the most important role as there is no other information. However for dynamic gestures usually it is enough to store the starting and ending hand shape and the movement trajectory. During the system work the recognition module returns the gesture description vector which is compared with the gestured database to retrieve the gesture ID which is then transmitted.

The gesture description vector for the sign rendering, on the contrary, should contain all the necessary information for 3D gesture animation. There have been proposed many notation schemas which differ in the level of precision and easiness of gesture description. For example, the HamNoSys – Hamburg Notation System [23] introduces very precise graphical description, although it results in relatively complicated definitions that are difficult to create and read by the humans. On the other hand, Bowden et al. [1] notices that for proper recognition the most crucial information is stored in the hands position relatively to each other and to the body, movements of the hands and the proper hands shapes. This draws to suggestion that such basic information might be sufficient as well for gesture animation. Nevertheless, there is applied the Szczepankowski's gestographic notation [19], following the solutions described by Francik and Fabian [3] as it is a ready solution. This notation characterizes the easiness of usability while creating the gesture description as well as adequate gesture description that allow for proper animation. Although, it is very detailed notation, sometimes it lacks exactness and needs human intuition to properly retrieve the gesture. But as was shown in Thetos system [24] it gives good results. Each gesture specification, called a gestogram, consist of one or more sections, which may appear in any number and order:

- hand configuration,
- hand orientation,
- hand location,
- relations between hands,
- direction of the hands motion,
- additional parameters of movement.

All those parameters are stored in the gesture database and are utilized for gesture rendering.

D. Network Transfer

Our concept focuses on the size optimized network-transfer of the recognized gestures in between the users that participate in a specific session. Additionally, the extension of the stored gestures is as well supported over a network connection with a central server. Thereby the necessary network transfer in our application can be separated (like depicted in the fig. 1) into the following elements:

- session management,
- gesture based communication stream,
- extensions and updates of / from the centralized database.

The integration of a session management becomes necessary for our concept, since a communication stream must be established between several participants. Since we assume that in most cases the users IP addresses are not known commonly the usage of a mapping between the usernames and their current IP addresses is necessary either. The integration of the Session Initiation Protocol (SIP) that was introduced by Rosenberg et al. [16] is adequate to solve these necessities. Therefore SIP is de facto the standard in the domain of session based communication and enables the establishment of any session based communication streams between unlimited amounts of participants. The basic functionality of the SIP protocol can be enhanced by several extensions. Especially the SIP Instant Messaging and Presents Leveraging Extensions (SIMPLE) [15]are relevant for our concept, since they enable a presentity functionality including status reports of subscribed contacts. By this, users can presume the availability of their contacts. Additionally, it supports a peer to peer based content transfer.

For initializing a gesture based communication stream between several users a SIP session is established. Afterwards the recognized gestures are transferred through a TCP/IP based communication stream. During the transfer the gestures are represented by the gesture id, like introduced in section 3.3.

The adaptation of this representation leads to a significant reduction of the message overhead, by which the usage of our application even in network environments with restricted bandwidth is possible.

Since each transmitted packet contains a complete gesture description a packet-loss in the communication stream is significantly more critical than for applications that focus on audio or video streaming. Therefore the usage of the TCP/IP connection and a buffer is essential in our scenario to avoid information loss. In case of occurring packet-loss the missing packets are retransmitted by the TCP/IP functionality. The buffer stores the received packages for 10 ms on a receiver site to enable the retransmitting of lost packages. As a result the presentation of the gestures in the proper chronology is achieved. If a package still is not received in time it is recognized by the system through an incremented package id and is reported to the user additionally.

The extension of a language specific library can be done during communication like described in the section 3.6. Therefore the new gesture and rendering description is transmitted to the central server that extends the language specific database accordingly. During the initialization progress of a client application all local language libraries are checked whether are present by requesting the central server. In case of differences, the specific language library is updated by downloading the current version from the central server. The extension and update process is independent from the communication process. We recognize the necessity that if the language specific library has been extended by a gesture this must be available to the other participants directly, since the information is essential during the rendering process but this case was excluded from our current concept for concise reasons.

E. Gesture Rendering

Due to the reduction of network data transmission arises the necessity of gesture rendering module creation. In this module the received gesture ID retrieves from database the gesture description vector for the rendering needs. The data stored in this vector describes the actions of an avatar which is rendered using 3D rendering approach based on the ideas introduced by Francik et al. [3]. To minimize the stored data there are only described the behaviour of the crucial body elements for the expression of the gestures. Those elements are divided into several points following the position of anatomy joints. The position of each point is included into the gesteographical description. In case of the dynamic gestures the gesteographical description may consist of several positions per point. To assure the natural look of gestures during the animation process the whole gesture animation is divided into parts where static description of the objects (hands, head and body) is known and then the passages between those frames are interpolated. Moreover, there must be taken care that each body parts do not penetrate each other and the proper joints bend during movement.

The animation module is designed with OpenGL animation unit and the open Microsoft COM interface was used for implementation, therefore an uncomplicated integration is assured.

F. Usability Aspects

Except the essential functionality of the sign language chat, which allows the chat based communication with sign language over the internet, we want the system to fulfil some other usability aspects. First of all we concentrated on the traditional problems in standard communication applications connected with reduction of delays and the necessity of feedback, however some specific characteristics are also taken under consideration. The system should enable the possibility of sign language extension, assure proper work in case of packet loss and distinguish between chatting and non-chatting phases. Additionally, by proper building of the *Gesture database* structure it can allow basic translator among different sign languages.

During the conversation may happen that some gestures might be unknown for the system, therefore the module for extending the library is suggested. Such module task is to create the gesture description vector by training the recognition module with the new gesture like shown in [11][13]. This part allows the proper recognition, however for whole system work the gesture rendering description vector needs to be created. This can be done or by typing the codes of gesture movement for people familiar with the notation utilized in the system or by exploiting the avatar and setting appropriate movement which will be mapped to the description vector.

The recovery of lost packets is possible due to the fact that each packet contains the part of the conversation is numbered with the successive number, therefore when one number lacks the application asks for the retransmission of the lost packet. When the retransmission is impossible the end users are informed about the loss, otherwise the message sense could be corrupted.

For chatting system it is very important to distinguish between the chatting and non-chatting phases, since non communicational gestures could be interpreted as communicational gestures otherwise. As a solution we suggest to add a gesture which starts and ends the chatting phase.

The last additional functionality is the translation between sign languages. The translation concept is very easy however it does not solve problems connected with different grammar structure, but allows for easy user communication between languages. We assume that for each sign language the database structure will be very similar and the ID describing the recognition and rendering vector is constant for the gesture of the similar mining between all sign languages databases. Therefore user signing a message in one language could be rendered using other sign language library.

IV. SUMMARY

We have introduced a concept to support the sign language based communication of hearing impaired people through digital networks. To reduce the necessary network load and to increase the usability we suggest the usage of an image processing module. Thereby the recognised gesture can be mapped to a gesture id that represents the expressed word. This gesture id is transferred to the participants of a communication, where the appropriate gesture is rendered. This approach results in a significant reduction of network traffic in comparison to existing applications. Additionally, we support the

uncomplicated extension of the existing vocabulary through the user, which represents another essential restriction of existing approaches. We utilize a centralised server-component which stores and extends current vocabularies. In case of recognised extensions the dictionaries of the clients are updated. Through these approaches a prospective way for network based communication of hearing impaired people is achieved.

REFERENCES

[1] Bowden R., Windridge D., Kadir T., Zisserman A., Brady M.: A Linguistic Feature Vector for the Visual Interpretation of Sign Language, In Tomas Pajdla, Jiri Matas (Eds), Proc. 8th European Conference on Computer Vision, ECCV04. LNCS3022, Springer-Verlag (2004), Volume 1, pp391- 401.

[2] Chan-Su L.; Zeungnam B.; Gyu-Tae P.; Won J.; Jong-Sung K.; Sung-Kwon Kim.: Real-time recognition system of Korean sign language based on elementary components, Proc. of 6th IEEE International Conference on Fuzzy Systems, 1997, vol. 3

[3] Francik J.; Fabian P.: Animating Sign Language in Real-Time, 20th IASTED International Multi-Conference Applied Informatics, Innsbruck, Austria, pp. 276-281

[4] Gaolin Fang; Wen Gao; Debin Zhao: Large vocabulary sign language recognition based on fuzzy decision trees, IEEE Transactions on Systems, Man and Cybernetics, 2004vol. 34

[5] Grobel K.;. Assan M.: Isolated sign language recognition using hidden Markov models, IEEE International Conference on Systems, Man, and Cybernetics, Computational Cybernetics and Simulation, 1997, vol. 1

[6] Grobel K.; Heinz H.:.Video-based handshape recognition using a handshape structure model in real time, Proc. of the 13th International Conference on Pattern Recognition, 1996 vol. 3

[7] Hernandez-Rebollar J.L.; Kyriakopoulos N.; Lindeman R.W.: A new instrumented approach for translating American Sign Language into sound and text, Proc. of 6th IEEE International Conference on Automatic Face and Gesture Recognition, 2004

[8] Ming-Hsuan Yang; Ahuja N.; Tabb M.: Extraction of 2D motion trajectories and its application to hand gesture recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, vol. 24

[9] Kobayashi T.; Haruyama S.: Partly-hidden Markov model and its application to gesture recognition, Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing, 1997, vol. 4

[10] Malassiotis S.; Aifanti N.; Strintzis M.G.: A gesture recognition system using 3D data, Proc. of 1st International Symposium on 3D Data Processing Visualization and Transmission, 2002

[11] Nurzyńska K., Duszeńko A.: An Overview of Polish Sign Language Learning Tool with Sign Recognizer Feedback, II International Conference of Interactive Mobile and Computer Aided Learning, IMCL'2007, 18-20 April 2007, Amman, Jordan, ISBN 978-3-89958-276-5

[12] Nurzyńska K.: Are the skin colour detectors stable in changing lighting conditions? (Comparison of skin colour detectors for signed language recognition). VIII International Workshop for Candidates for a Doctor's Degree, OWD'2006, Wisła, Poland, October 2006

[13] Nurzyńska K., Duszeńko A.: Interactive System for Polish Signed Language Learning. International Journal: Emerging Technologies in Learning, iJET, vol. 1, No 3 (2006)

[14] Ohene-Djan J., Zimmer R., Bassett-Cross J., Mould A., Cosh B.: Mak-Messenger and Finger Chat, Communications Technologies to Assist in the Teaching of Signed Languages to the Deaf and Hearing, ICALT 2004

[15] Rosenberg J.: Session Initiation Protocol (SIP) Extensions for Presence, RFC 3856 Internet Draft, IETF Network Working Group, August 2004

[16] Rosenberg J., Schulzrinne H., Camarillo G., Johnston A., Peterson J., Sparks R., Handley M., Schooler E.: SIP: Session Initiation Protocol, RFC 3261 Internet-Draft , IETF Network Working Group, June 2002

[17] Starner T.; Weaver J.; Pentland A.: A wearable computer based American sign language recognizer, 1st International Symposium on Wearable Computers, 1997

[18] Starner T.; Pentland A.: Real-time American Sign Language recognition from video using hidden Markov models, Proc. of International Symposium on Computer Vision, 1995

[19] Szczepankowski B.: Wyrównywanie szans osób niesłyszących, Warszawa, WSiP, 1999

[20] Szczepankowski B.: Lektorat Języka Migowego – Kurs Wstępny, Polski Związek Głuchych, Centralny Związek Spółdzielni Inwalidów, Warszawa 1986

[21] Vogler Ch.; Metaxas D.: Parallel Hidden Marcov Models for American Sign Language Recognition, Proc. of the International Conference on Computer Vision, 1999

[22] http://www.zak.co.il/deaf-info/old/demographics.html, 2007

[23] Sign Language Notation System http://www.sign-lang.uni-hamburg.de/projects/HamNoSys.html, 2007

[24] Thetos system homepage http://thetos.polsl.pl, 2007

AUTHORS

K. Nurzyńska is with the Institute of Informatics, Software Division, The Silesian University of Technology, ul. Akademicka 16, 44-100 Gliwice, Poland (email: Karolina.Nurzynska@polsl.pl)

S. J.F. Fudickar is with the Institute of Computer Science, Network and Multimedia Teleservices Division, University of Potsdam, Germany (e-mail: Sebastian@Fudickar.eu).

Paper presented at ICL2007 conference, Villach, Austria, September 2007.

Manuscript received 25 October 2007. Published as sumitted by the authors.