

A User-Friendly Sign Language Chat

Sebastian Fudickar¹, Karolina Nurzyńska²

Institute of Computer Science University of Potsdam, Institute of Informatics Silesian University of Technology

Key words: *Chat, Image recognition, Sign Language*

Abstract:

The introduced concept of a gesture based sign language chat enables the communication of hearing impaired people over networks, including concerns regarding low throughput and bandwidth characteristics. Existing applications handle that task not appropriate, since either high bandwidth networks are necessary or the extensibility of the vocabulary is complicated for untrained users. Thereby our approach facilitates existing image recognition techniques for the determination of gestures which are captured by a standard camera. The recognised gestures then are mapped to a gesture Id that is transferred to the participants, where it is rendered. Additionally, the uncomplicated extension of the vocabulary by the users is supported. Therefore a centralised server updates the locally stored dictionaries, if necessary.

1 Introduction

Usually a human interaction focuses on the sound world, where the communication is based on the speech and in which most information is conveyed via voice and other sounds. However, there are people who live in the world of silence. For them nothing can be heard, as they are hearing impaired. According to the research of Omer Zak [22] around one percent of people living in European Union in 1994 suffered from this disease, excluding people suffering hard hearing problems.

For all of them a voice communication is impossible or troublesome. Hence they have invented a sign language. The sign language consists of a grammar and a vocabulary. Usually the grammar is significantly different to the 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 expressions [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 internet technology concentrates mostly on mirroring the existing reality. Computer program designers try to make it as similar to the reality as possible. From one point of view it is easier to use it for people not familiar with computers, but on the other hand it narrows the horizons of creative thinking. Nowadays, the communication via Internet focuses on the mail and the telephone or video call while it could go much further. Thereby most of the current network based communication techniques are not adequate for hearing impaired people. The video supporting multimedia stream, which occurs as single adequate solution, requires high network bandwidths, which limits the range of possible scenarios.

As a result it is necessary to focus on the needs of hearing impaired and design systems which improve their communication without imposing any special capabilities from the user.

Therefore the communication systems should support the sign languages in a sophisticated manner.

2 Existing Applications

Current concepts for the task of a sign language based communication over digital networks are discussed in this section. Thereby two approaches are mainly focused.

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.

3 Concept

We introduce the concept of a chat for the sign language based communication, which overcomes the deficiencies of the existing approaches that we have mentioned in previous sections. Current hardware enables even standard computers to process complex calculations in soft real-time. The development of high performing algorithms in the field of the image processing additionally enables a new approach to this problem. Thereby we focus mainly on usability aspects and the reduction of resulting network traffic.

In general, our system enables the sign language data acquisition, which is composed of the hands shapes and its position mainly in the relation to the head position. In the next step of meaning recognition the sign is mapped to a gesture description according to the language specific gesture database. Afterwards, the meaning is encoded, what optimizes the size of data, and transferred on a pre-initialized data stream to the receivers, where it is rendered.

3.1 Gestures recognition

During the last decades the problem of hand gesture recognition has been widely researched. First of all, the systems must be able to work with both static and dynamic gestures. The static gestures are characterized mostly by the fingers shapes and positions as well as the hand position in the relation to the head. In the case of dynamic gestures those information is still

important, but 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. Different methods for the input information acquisition have been developed. Therefore the group of systems based on electrical gloves is the most represented [2][4]. Those systems have very high recognition ratio for big vocabulary consisting of thousands of signs, however they are very expensive and uncomfortable in utilization as the user must wear the sensors and has to be connected to the computer. The second groups of systems utilize markers like presented by Grobel *et al.* [5][6]. In this case the video camera is utilized for image capturing. Here the user is obliged to where the colour gloves to improve the hand detection, nevertheless the efficiency of those systems is significantly lower and the average image recognition dictionary consists of some hundreds of gestures. Finally, the less researched but also being very promising domain for further development are systems which use video camera for image capture only (without any additional markers) [17]. On the one hand the recognition rate for this approach in comparison to the previously mentioned solutions (less than hundred signs) is significantly lower. But, on the other hand, the reduced acquisition costs as well as the essential increase of usability makes this approach very attractive to the non laboratory dedication. Therefore, this approach is suggested as a part of the data acquisition module in the presented system. The natural way to obtain the information about the hands and head positions in the image is the utilization of skin colour detectors like suggested in [12].

After collecting the input data the step of recognition takes part. First systems were based on the technique of templates matching [7], neural networks [8], statistical analysis [10], but all of these methods shown to be prone to the dynamic properties of the gestures. Generally, none of them could properly recognize among similar gestures shown in different speed. The idea of exploiting the stochastic system, hidden Markov model (HMM) [9][18][21], came as a solution to this problem and nowadays is the method used broadly.

3.2 Gestures description

There are necessary two different gesture description vectors for each gesture. The gesture description vector used in the recognition module holds the hand shape description on each part of the gesture movement, usually in the starting and the ending point of the gesture and the movement trajectory. That information is compared with the vector generated by the recognition module and then the system decides which sign was recognized. The recognition is equivalent with finding the identification number connected with the sign in the database.

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.

3.3 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 3.2.

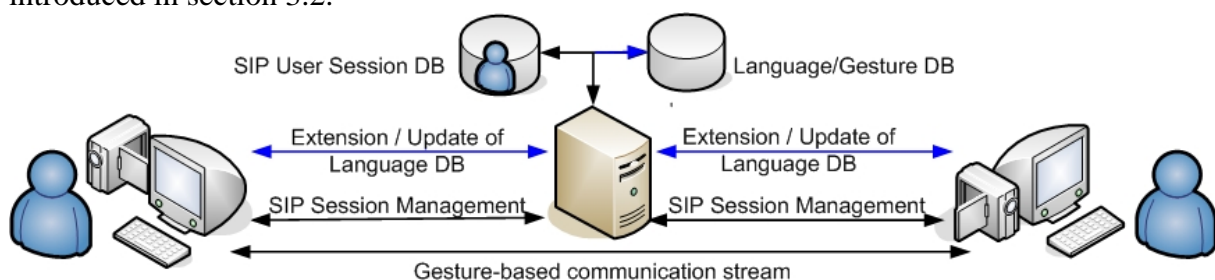


Figure 1: Network flow including the communication and the extension of language libraries

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 3.5. 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.

3.4 Gesture rendering

The gesture rendering module renders 3D gestures according to the received gesture descriptions on the receiver side. Thereby it is a part of the network transfer reduction as only the sign identification number is transferred.

Our 3D rendering approach is based on the ideas introduced by Francik *et al.* [3]. The rendered body parts are limited to the ones that are relevant for the expression of gestures. These body parts 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.

3.5 Usability aspects

Essential elements concerning the usability of the sign language chat include traditionally known things of standard communication applications, like reduced delays, the necessity of feedback but also focus on much more specific characteristics. These specific characteristics cover the extension of the gesture libraries, the distinction of chatting and non chatting phases as well as the response in case of missed gestures through a packet lost.

In case of recognised gestures that are unknown to the system, the gesture library will be extended by them. In that case the user has to train the recognizer with the new gesture like shown in [11][13]. Additionally the rendering description has to be generated through the utilization of an avatar. Afterwards the word must be mapped to the appropriate written word.

The distinction of chatting and non chatting phases is essential, since non communicational gestures could be interpreted as communicational gestures otherwise. Therefore we suggest the usage of a specific gesture for starting and stopping the chat phase, in combination with a user-feedback.

If an unrecoverable packet loss occurs during communication, the receiver must be informed about the missing gesture through a feedback, since missing gestures may change the sense of a sentence totally and may lead to misunderstandings. Therefore we include an incremented id in each transmitted package and identify the chronology depending on them. When packages are lost, we recognize that on receiver side by a non continuous id flow and react with a user-feedback.

4 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, <http://sun.iinf.polsl.gliwice.pl/sign/ias021.pdf>
- [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.; Srinivas 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 Markov 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://sun.iinf.polsl.gliwice.pl/sign/>, 2007

Author(s):

Sebastian J. F. Fudickar, B. Sc. SE
 University of Potsdam, Germany
 Faculty of *Networktechnologies and Multimedia Teleservices*
 Institute of Computer Science
 Sebastian@Fudickar.eu

Karolina Nurzynska, MSc EE
 The Silesian University of Technology,
 Faculty of *Automatic Control, Electronics and Computer Science*
 Institute of Informatics,
 16 Akademicka Street
 44-100 Gliwice
 Poland
 Karolina.Nurzynska@polsl.pl