Studies on Machine Processing Of Sign Language in the Context of Deaf Support. Application in Health Care – Interactive Service System for the Deaf

Dorota Szulc[1,2], Jakub Gałka[1,3], Mariusz Mąsior[1],Filip Malawski[1],Tomasz Jan Wilczyński[1,4], and Krzysztof Wróbel[1]

[1]AGH University of Science and Technology, Department of Electronics, Krakow, Poland
[2]JagiellonianUniveristy, Institute of Pedagogy, Krakow, Poland
[3]VoicePIN.com Sp. z o.o., Kraków, Poland
[4]Poradzis Sp. z o.o, Przyszowa, Poland

Abstract: *The article presents the architecture and the operation of a prototype dialogue system which enables the deaf to make an appointment with a medical specialist using sign language. It describes issues related to: concluding the logic of a dialogue, generating messages in Polish Sign Language using a virtual character (avatar), and acquiring and recognizing deaf signs. It also shows the resources and the building methodology of a PJM sign database created for the system. The system is described as a result of multiple works of the research group from the AGH University of Science and Technology and the VoicePIN.com companyregarding the machine processing of sign language. It is developed as a proof-of-concept of PJM sign recognition tools and of a Human-Machine Interface for the deaf.*

Keywords: *signs database*, *deaf support, dialogue system using sign language, sign language processing*

1. Introduction

The basic method of communication for approx. 70 million people all over the world is sign language [1]. The fact that this group's method of communication is different from those used by the rest of society does not leave access to information and services uninfluenced. The communication barrier which arises between those whose hearing is not impaired and the deaf hinders, and sometimes even makes it impossible for the deaf to e.g. settle matters in public administration offices.

In order to meet the expectations of the people who use sign language to communicate, researchers from the AGH University of Science and Technology in cooperation with a Krakow-based company called VoicePIN.com launched a 3-year project entitled "Virtual Sign Language Interpreter WiTKoM"in 2013 as part of the Applied Research Program of the National Centre for Research and Development. The main research objective of the project was to study possible applications of pattern recognition algorithms, machine learning, and natural language processing in the context of processing utterances in Polish Sign Language (PJM). The solutions found during the course of the project can be used for developing and improving elements of automatic systems for interpreting sign language into spoken language and vice versa.

One of the project's outcomes is the development of an interactive dialogue system which enables a deaf person to make an appointment with a selected medical specialist by using sign language. Earlier research conducted by the authors revealed that visiting a medical center is one of the situations in which a deaf person needs the most support in communicating. When a deaf person wants to make an appointment to see a doctor and ask about all the details connected with it, he/she faces the threat of not being understood or receiving an answer he/she may not understand. In such a situation, they often have to ask other people for help, even if their problem is of a private matter. The solution developed within the framework of the project is a prototype and can serve as a basis for developing advanced tools, taking advantage of the machine processing of natural languages which can support communication between the deaf and those without any hearing impairments.

In the second chapter, the architecture of the system will be presented, as well as the way it works, and the terms it uses. Theissues connected with concluding the logic of the dialogue will be discussed, as well as generating messages in Polish Sign Language by using a virtual character (avatar), and the acquisition and

recognition of deaf signs. In the third chapter, the authors will describe a sign database created for the system. They will draw attention to the sign acquisition methodology and the linguistic resources of the database. In the conclusion, the researchers will present the currently achieved effectiveness of the system and the perspectivesfor its development.

2. Structure and operation of the system

The system for the automatic registration of deaf patients consists of several independent and separately developed IT modules which are responsible for the subsequent stages of interpreting sign language. The system translates sign utterances into Polish, generates responses, questions, and messages in the Polish Sign Language, and performs it as an animation of a virtual character (avatar). The general structure of the system is presented in Fig.1.
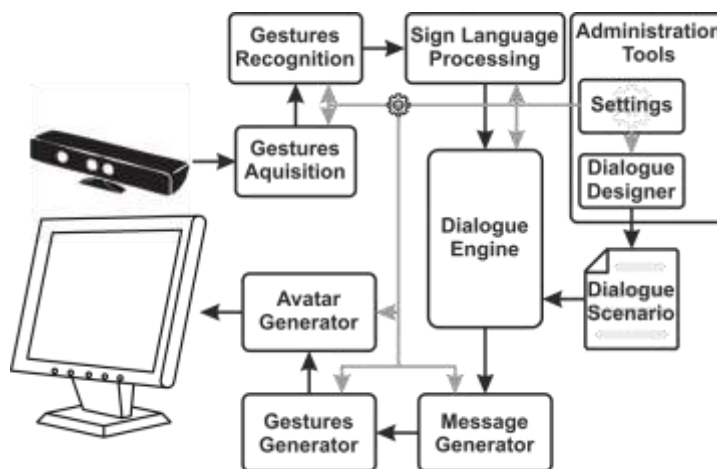
Fig. 1: The structure of the system for the automatic registration of a deaf patient

The main functional elements from the user's point of view are the physical components for the acquisition of signs and the graphical interface of the computer software (presented in the user window). The stand for serving deaf individuals is equipped with a processing unit (computer), a video sensor – in the form of an RGB camera and/or depth sensor (e.g. Kinect), and a standing or seating place (depending to local conditions) for the person using the system. The acquisition sensor used determines if the stand requires a uniform static background or if any is allowed, including changeable backgrounds. The communication application – shownon the screen in front of the user - displays speech utterances in sign language performed by an avatar. The entire system setup allows for ergonomic and intuitive handling by the deaf person - gesture performance is conducted toward the screen.

The dialogue is aimed at scheduling an appointment with a selected medical specialist in a requested or available timeframe. The dialogue starts with the avatar greeting the deaf patient: *Good morning, I am an automatic interpreter of sign language and I will help you make an appointment. Which specialist would you like to see (e.g. internal medicine specialist, cardiologist, dermatologist, etc.)?* Subsequent messages from the avatar – the automatic receptionist – include: *Would you like to make an appointment with a specific specialist (name and surname)? Please spell the name and surname of the selected specialist. Please specify the time (date and time) of your appointment. The nearest available dates are: .... Please tell me your name and surname so I that can book your appointment. Please provide the telephone number for contact purposes (SMS contact). Thank you, your appointment has been made. The date of your appointment with ..... is .... . If you have any questions or would like to cancel the appointment for any reason, please text us at ..... or send us an email.* The whole dialogue logic must be properly designed in order to ensure the intuitiveness and ease of use of the communication. The system has a built-in set of special applicationsused for cooperation with system modules. The application for dialogue design allows to definethe dialogue stages as well as the messages delivered by the avatar and to determine how the system responds to the commands provided by the user.

The dialogue designers and systemadministrators have at their disposal a set of tools for the diagnosis and control of the system. The person responsible for the operation of the system has extensive management and modification capabilities, e.g. by selecting the dialogue scenario and time settings available for the user's statements. It is also possible to control all parameters of the system modules, including the algorithms responsible for pattern recognition, gesture classification, and recognition confidence levels.

In a basic dialogue system, utterances addressed to deaf people can be pre-recorded as video clips by a real person. Thus, there is no need for the creation of a module responsible for displaying avatar gestures,as it is possible to simply play an appropriate recording.However, a slightly more advanced system, with the dialogue logic dependent on the user's responses (e.g. The*date of scheduled visit to the doctor ... is ....* supplemented with information extracted from the conversation) would require too many possible combinations of statements to be recorded. The variable part of the message could be recorded separately, but the combined video would not be smooth, making it difficult to understand its content. This problem can only be solved by the automatic online generation of sign language animations (using the avatar). This module performs the entire statement, so it can smoothly connect all sign gestures. Another advantage of the animated character is the ability to highlight its features for easy understanding, e.g. by increasing the mimics and hands size. The animation is generated based on the provided positions of the virtual bones over time. All of the individual signs which might appear in the dialogue must be determined by the positions of the bones. Such format is not very convenient, which is why the module allows designers to define signs in the HamNoSys notation [2].It allows for easier understanding of the representation of the sign, which can still be processed by the computer.

The possible answers in sign language to each of the questions asked by the avatar have been identified in consultation with deaf people. The set of training recordings of signs was gathered based on the adopted dialogue scenario of "making a doctor's appointment". The process of building a database of sign gestures is presented in the second chapter.

For a system to recognize the signs carried by the deaf, it was necessary to implement an extraction module for the features of signs and train a classifier to recognize them. Each sign has its own features which have to be extracted. Feature differentiation is necessary, otherwise sign utteranceswould be unreadable for the computer. Generally, thereare three divisions of features: displacement-based, velocity and movement-based, and hand shapes [3]. Unfortunately, in many cases these assumptions are not sufficient. In the case of extensive scenarios some of gestures have more than one variants. Therefore it was necessary to create further high level algorithms to distinguish signs between each other. But the main problem based on overall system integration from the begging of data acquisition up to sign recognition.That is why all of the acquired signs were strictly selected to satisfy the use case with emphasis on integrity, cohesion, and being universal for deaf people.

In terms of implementation, the system can operate on any stationary PC-class unit with the following minimal parameters: Intel i7 class processing unit, USB 3.0 port compatible with Kinect 2.0, Windows 8, and Kinect 2.0. The system is also ready to be operational on mobile devices (smartphone/tablet) equipped with suitable sensors (camcorder and/or structured light sensor) in the future.

3. Sign databaseand acquisition methodology

The vast majority of sign language database content is generally dividedby gesture character (letter, digit, animals etc.) [4], without distribution by usecases.For the purposes of the described system,a sign database was required which: fulfils the patient's registration dialogue, contains a lot of ways the deaf could respond (finger spelling, simple signs, complex signs), takes account a lot of the variants of signs. As the main objective was to use the system in real conditions (medical facilities), it was necessary to have a system which will work user-independently. The database must contain signs performed by different people. To better train the system,the signs had to be repeated at least several times. The important thing was providing high performance gestures for sign recognition in real conditions for PJM.

The authors had started recording the signs for the database in April 2015. The recordings were carried out up to June of the same year. In total, they were attended by 17 deaf and 3 hearing lectors of PJM. Subjects were aged 17 to 65and they were from the southern part of the Poland. According to the scenario of the "making a doctor's appointment" dialogue, the following recordings were made: 28 signs of medical specialties, 38 signs of the fingerspelling alphabet, 38 signs of cardinal numbers in the range from 0 to 2000, 9 signs of ordinal numbers

in the range from 1 to 9, weekdays, months, 26 signs to specify the time,and 9 other signs which complement the dialogue.

The database was recorded in a dedicated studio (located at the VoicePin.com company office in Krakow, Poland). The recorded person was being illuminated from four directions by four soft-light lamps, which produced the proper amount of light for the cameras operating at 30 frames per second as well as provided an evenly lit scene without any shadows. There was a green background behind the person, which facilitated further processing of the recorded data by allowing to easily extract the person in the image. The person was seated in a distance of approximately 1 meter from the cameras. These were mounted on tripods, approximately 1 meter from the floor, facing directly the main area of signing - the torso and the head. Two types of cameras were used. The first type was the Kinect 2 sensor, which provided high quality color images as well as depth images, which store information regarding the distance to the objects in front of the sensor. The second type was a set of two PS3Eye stereo cameras, which provided medium quality color images from two perspectives. This allowed to obtain depth information from stereo reconstruction - a method similar to vision processing in the human brain.

The recordings were made with dedicated software, which we designed and implemented specifically to optimize the process of the acquisition of the data. The software allowed to record simultaneously from all cameras and provided automated synchronization mechanisms for all data streams. The operator was only required to press the 'start/stop' buttons for each recording, which resulted in a very efficient manner of work. The recordings were automatically stored in a dedicated structure of folders, taking into account the current session, person, gesture, and repetition. Due to such automation, the obtained database required no further preprocessing and could be used directly for the development of recognition algorithms.

The recording process also required solving a number of problems associated with the linguistic aspect of Polish Sign Language as well as the methodology of working with deaf individuals.

It must be mentioned that Polish Sign Language is a language which has still not been studied sufficiently and its vocabulary is not yet well-defined. The biggest challenge were the signs used for doctors' specialties. Polish Sign Language lacks most of the signs used to refer to doctors' specialties. The names of specialties are expressed by signs referring to symptoms, body parts and other associations with the given specialty. The same specialty can be expressed differently by different deaf individuals depending on e.g. the health problems they want to present to the doctor. The names of specialties usually consist of a series of (several) individual signs which can be expressed by the deaf in a different order and in different variants (see Tab.I). The multivariant character of the signs is one of the features of Polish Sign Language. A sign performed by different individuals can also differ because of their individual signing manner and the way they perform a given sign.

TABLE I: Different variants of signs for [ORTHOPAEDIST] and [LARYNGOLOGIST]

orthopaedist_x = [DOCTOR_x] + [GROUP_x] + [LEG_x] + [TO BREAK]
orthopaedist_y = [DOCTOR_y] + [GROUP_x] + [SHOULDER-PAIN]
orthopaedist_z= [DOKTOR_x] + [SKELETON_x]

laryngologist _x = [DOCTOR_x] + [GROUP_y] + [PROBLEM_x] + [EAR_x] + [THROAT_x]
laryngologist _y = [PHISICIAN_x] + [EAR_y] + [AUDIOMETRY GRAPH]
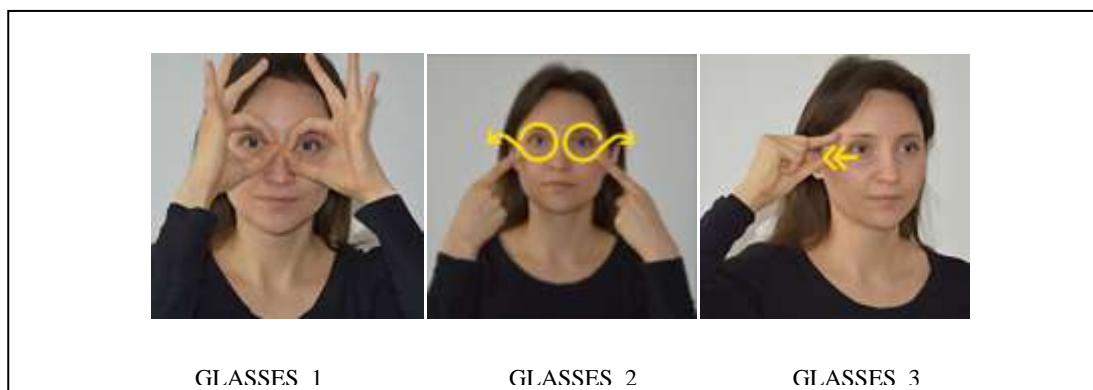laryngologist _z = [THROAT_x]

Fig. 2: Different ways of performing the sign for [GLASSES]

In order to study the way the deaf sign the names of individual specialties, those included in the study were shown pictures of specialists and later asked to sign the name. The pictures were chosen in the manner which would not suggest to those included in the research to use signs referring to individual specialists. Thanks to the pictures, it was possible to trigger the sign in those deaf whose Polish language competence is low (a pilot study carried out before the recording revealed a poor understanding of verbal designata used to refer to specialties by the deaf).

Every variant of the collected signs was tagged appropriately and described according to the following criteria: a) the arrangement of fingers (the corresponding sign of the Polish or international manual alphabet, a cardinal number sign, or a modification of previously mentioned ones), b) the position of hand in respect to the body (at the level of the head, chest, waist, on the right, on the left, in the middle), c) the position of the hand in space (hand positioned vertically – fingertips facing upwards, downwards, horizontally, or diagonally – determining the position of the back, the bottom, and the edge of the hand), d) the direction of movement (e.g. downwards, upwards, rightward, forward), e) the type of movement (e.g. circular, straight, wavy, energetic, slower).

The table below presents the basic statistics and parameters describing the database.

TABLE II: Statistics and parameters of the database

Number of signers	20	Parameters of recordings:
Number of (semantically) different signs	167	
Number of signs, including different variants	390	- 2 PlayStation 3 Eye cameras: image color VGA (640x480), 30 fps
The average number of variants denoting medical specialists	6,3	
Number of samples	24 372	- Microsoft Kinect 2: image color VGA (640x480), 30 fps image depth (512x424), 30 fps
Average length of one sample	3 s	
Number of gesture classes performed by more than 10 users	158	
Number of unique signs with more than 10 repetitions per user	114	
Total duration	20 h	

The authors aim to make the database publicly available. Great emphasis was given to the video processing stage of the collected recordings. It is planned to publish the set of features describing every recorded motion as well. The set of features is internally diverse. It consists of features of the body parts tracking and the depth image from the Kinect sensor. In addition, they were designed handshape features and the powerful motion field approach based on the optical flow algorithm. Full set of varied features gives the opportunity to use the database in the context of the automatic sign language recognition systems. Collected features are an integral part of the database and they are useful for machine learning research community. The database is also interesting in terms of linguistic analysis.

4. Conclusions

The presented system, at the current stage of development, recognizes up to one hundred sign symbols with an average accuracy of 91% (80-91% depending on the type of sign). The work on the extension of the symbols' dictionary and on an increase in system efficiency continues.

The entire described system was developed as a proof-of-concept of an efficient Humane-Machine Interface for deaf. The primary goal of the work was to research and define the possibilities of sign gesture recognition usage in a fully automatic dialogue system. Such systems are well-known in the area of speech communication, but are still new and unfamiliar for users of sign language.

Among the developed solutions are methods of gestureacquisition, advancedimage processing and recognition algorithms, as well as automatic translation and natural language processing procedures. A database of recorded gestureswas created and will be used to further work on image processing and issues related to machine learning.

The described system will be expanded in the future. The development of a sign language dictionary will allow building more complicated and usable dialogues. Further development of the processing algorithms and sign gesture acquisition will help in embedding sign language translation systems into mobile devices. Finally, complementing the translation system with an ASR (Automatic Speech Recognition) module may result in a useful tool for communication between the deaf and the hearing.

Undertakings such as described workundoubtedlybring progress in developing successful technological solutions that will support and assist deaf people in functioninga hearing society.

5. Acknowledgements

6. References

[1] World Federation of the Deaf, "Sign Language", available at: http://wfdeaf.org/human-rights/crpd/sign-language (accessed 30 March 2016).

[2] T. Hanke, "HamNoSys-representing sign language data in language resources and language processing contexts," in *Proc. LREC 2004*, Workshop proceedings : Representation and processing of sign languages, 2004, pp. 1-6.

[3] L. Rung-Huei, and O Ming, "A Real-time Consinous Gesture Recognition System for Sign Language", in *Proc. IEEE ICAFGR, Japan,* 1998,pp.558-567.

[4] M. Kawulok, J. Kawulok, and J. Nalepa, "Database for hand gesture recognition", available at: http://sun.aei.polsl.pl/~mkawulok/gestures/ (accessed 30 March 2016).