ISSN: 2277-3754
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 1, Issue 2, February 2012

Multimodal System using Human Computer Interface

Milind .V.Lande, Makranand Samvatsar
milind.jdiet@gmail.com, makarand111@gmail.com

Abstract -- This paper introduces a new prototype system for controlling a PC by Face or Hand movements and also with voice commands. Our system is a multimodal interface concerned with controlling the computer. The selected modes of interaction are speech and gestures. We are seeing the revolutionary of computers and information technologies into daily practice. Healthy people use keyboard, mouse, trackball, or touchpad for controlling the PC. However these peripheries are usually not suitable for handicapped people. They may have problems using these standard peripheries, for example when they suffer from myopathy, or cannot move their hands after an injury. Our system has been developed to provide computer access for people with severe disabilities. This system tracks the computer user's Hand or Face movements with a video camera and translates them into the movements of the mouse pointer on the screen and the voice as button presses. Therefore we are coming with a proposal system that can be used with handicapped people to control the PC. Thus this system is needed for interpreting and fusing multiple sensing modalities in the context of human computer interface. This research can benefit from many disparate fields of study that increase our understanding of the different human communication modalities and their potential role in Human Computer Interface.

***Keywords*- IHM, Gesture, Speech, multimodal interface, handicapped people.**

I. INTRODUCTION

With the development of information technology in our society, we can expect that computer systems to a larger extent will be embedded into our environment. These environments will impose needs for new types of human-computer interaction, with interfaces that are natural and easy to use. In particular, the ability to interact with computerized equipment without need for special external equipment is attractive. Today, the keyboard, the mouse and the remote control are used as the main interfaces for transferring information and commands to computerized equipment. In some applications involving three-dimensional information, such as visualization, computer games and control of robots, other interfaces based on trackballs, joysticks and data gloves are being used. In our daily life, however, we humans use our vision and hearing as main sources of information about our environment. Therefore, one may ask to what extent it would be possible to develop computerized equipment able to communicate with humans in a similar way, by understanding visual and additive input. For many people with physical disabilities, computers form an essential tool for communication, environmental control, education and entertainment. However, access to the computer may be made more difficult by a person's disability. A number of users employ head-operated mice or joysticks in order to interact with a computer and to type with the aid of an on-screen keyboard. Head-operated mice can be expensive. In the UK, devices that require the users to wear no equipment on their heads, other than an infrared reflective dot, for example Orin Instrument's Head Mouse [2] and Prentke Romich's Headmaster Plus [5]. Other devices are cheaper, notably Granada Learning's Head Mouse [7], Penny and Gilles' device, Mouse Enhancer for paraplegics [1], Eye Gaze System [5] and No Hands Mouse [8]. However, these systems require the user to wear a relatively complex piece of equipment on their head, an infrared transmitter and a set of mercury tilt switches respectively. People with severe disabilities who retained the ability to rotate their heads have other assistive technology options. For example, there are various commercial mouse alternatives. Some systems use infrared emitters that are attached to the user's glasses, head band, or cap. Other systems place the transmitter over the monitor and use an infrared reflector that is attached to the user's forehead or glasses. The user's head movements control the mouse cursor on the screen. Having an additional window produces another problem. It occludes other window applications, making the windows desktop more cluttered and less organized. Gorodnichy [1] resolved this problem by introducing a new concept, called Perceptual Cursor, which serves both the purpose of marking a position (as normal cursor) and the purpose of providing a user with the feedback on how remote user motions are perceived by a sensor. As such, Perceptual Cursor does not replace the regular cursor, but rather is used in the interface in addition to it, taking its functionality only when requested by the user. For voice-based computer control system available in the market, we can find Control Your PC with Your Voice [2], Nuance Voice Control 2.0 [2] which are an easy software solution to enable you to control your computer, dictate emails and letters using a minimum of keystrokes or mouse clicks.

Delphian Desktop [32] predicts the destination of the cursor based on initial movement and rapidly 'flies' the cursor towards its target. Although these techniques were

ISSN: 2277-3754
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 1, Issue 2, February 2012

designed for single monitor conditions, they can be easily tailored for multi-monitor setups. Head and eye tracking techniques were proposed to position the cursor on the monitor of interest [3][4]. This reduces significant mouse trips but at the cost of access to expensive tracking equipment. Benko et al. propose to manually issue a command (i.e. a button click) to ship the mouse pointer to a desired screen [5]. Ninja cursor [6] proposes a technique to improve the performance of target acquisition, particularly on large screens. This technique uses multiple distributed cursors to reduce the average distance to targets. Each cursor moves synchronously following mouse movement. explore the possibilities for augmenting the standard computer mouse with multi-touch capabilities so that it can sense the position of the user‘s fingers and thereby complement traditional pointer-based desktop interactions with touch and gestures. They present five different multi-touch mouse implementations, each of which explores a different touch sensing strategy, which leads to differing form-factors and hence interaction possibilities. In addition to the detailed description of hardware and software implementations of their prototypes, they discuss the relative strengths, limitations and affordances of these different input devices as informed by the results of a preliminary user study.

II. MOTIVATION

Gesture recognition systems identify human gestures and the information they convey. Although relying on gesture as the primary source of command input to computers may sound like science fiction, the technology has rapidly progressed in some areas such as virtual reality, by relying on special hardware and wearable devices. This hardware is often not cost-effective and is infeasible for some applications; consequently, gesture recognition based on alternative methods of Data Acquisition is being considered. In this article we introduce A novel Method for gesture Recognition. In 2D space which we used for interpreting Hand and ARM Movements as gesture commands to a vision-based user interface.

Due to low processing time Real time application of this project are enormous .almost all consumer electronics equipment today user remote controls for user interface. So, this system can not only be used as an interface between MAN and Computer but also as an interface for other domestic/industrial appliances such as television, washing machine etc. on integrating with microcontroller, based on just one unified set of hand gesture, the system can be used to interface user’s hand into predefined commands to control one or more devices simultaneously. For those who are not unable to use other technology because of some natural or acquired disability.

III. HAND GESTURE (ARM POSITION) RECOGNITION

Here we have detected different Hand Gesture made by different ARM positions. For this we have used Camera to take Images and MATLAB as programming tool. Camera monitors some area and when somebody gets into the area and makes some hands gestures (with different arm positions) in front of the camera, we detect the type of the gesture. The method we have used uses a Motion Detection as its first step and then does some interesting routines with the detected object. When the hands gesture recognition is detected, we can have an application that may perform different actions depending on the type of gesture. For example, gestures recognition application may control some sort of device or another application sending different commands to it depending on the recognized gesture. Here, we have demonstrated a method to recognize different Arm positions. In this project we have detected 4 positions by analyzing horizontal histogram only. This idea can be extended easily to recognize 15 different gestures by also analyzing vertical histogram.

Motion detection and object extraction Before we can start with hands gesture recognition, first of all we need to extract human body, which demonstrates some gesture, and find a good moment, when the actual gesture recognition can be done. For these tasks we will use Motion detection. For object extraction task we are going to use the approach, which is based on background modeling. We need a frame in which we have no human, only background. But while extracting human, to recognize hand gesture, from a frame, we can have many other objects (not human) being detected from motion detection code which undesirable. So, we set few parameters- aspect ratio of object detected, size of object detected etc to hard bind the presence of human. This makes the method more Robust. We also used the concept of Adaptive Background. Since we may have some minor changes in the scene from time to time, like minor changes of light condition, some movements of small objects or even a small object has appeared and stayed on the scene. To take these changes into account we are going to change our background frame at a very slow rate.

Now, when we detected an object to process, we can analyze it trying to recognize a hands gesture. The idea of our hands gesture recognition algorithm we used is 100% based on histograms and statistics. This makes this algorithm quite easy in implementation and fast for real time implementation. The core idea of this method is based on analyzing two kinds of object’s histograms: Horizontal and Vertical histograms.

Histograms obtained are shown in figure [2]. As it can be seen from the histogram, the hands areas have relatively small values on the histogram, but the torso

ISSN: 2277-3754
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 1, Issue 2, February 2012

area is represented by a peak of high values. Taking into account some simple relative proportions of human's body, we may say that human arm thickness can never exceed 30% percent of human's body height (30% is the value we used). So, applying simple thresholding to the horizontal histogram, we can easily classify hands areas and torso area. To check if a hand is raised or not we are again using some statistical assumptions about body proportions. If the hand is not raised its width on horizontal histogram will not exceed 30% of torso's width, for example. Otherwise it is raised somehow. This has been achieved by analyzing only Horizontal histograms. By analyzing vertical histograms we can recognize exact hand's position when it is raised. The problem with above technique is sometimes it also detect shadows as a part of human body, which gives error in histogram analysis. Refer figure [3]. This we removed by setting some thresholds and removing that extra part.

Real Time Implementation: Object extraction in this method is very robust. Since this method is based only on information from histograms, it is quite efficient in performance and does not require a lot of computational resources. We are getting this method working in real time with 3frames/sec. It is important to mention here that the time to process one frame is around 0.3 sec out of which 0.28sec goes just for Morphological operations (Opening and closing technique). This method may not work with sudden and large change of light intensity. We also need human to be wearing clothes of color i.e. of different color from that of colors present in the background.

IV. HAND GESTURE (SHAPES) RECOGNITION

Here gestures formed with different hand shapes are recognized and after implementing this in real time we used this for sign language implementation. A camera is used to take images and MATLAB is used as programming toolbox. In this project we first detect a hand in a frame of video and generate a radial histogram for that detected hand. We used two techniques for final recognition. One is PCA based and other is features from histogram based.

Hand/Band Extraction: In order to recognize hand gestures from Video, it is first necessary to detect information about the hand from raw data provided by webcam used. For this first, we are determining the color ranges of hand & marker and moreover, to make it independent of different user's hands, we are making this range dynamic with initial color calibration. This is achieved by making a rectangular box to find out the color ranges. In this project we analyzed 2 different color domains RGB and HSV. What we observed is, with HSV domain we can have small intensity variation while application is running but RGB give bad results even with small intensity variations. But in MATLAB to convert RGB image to HSV image is computationally very large, it took around 0.2 sec while whole code takes 0.3 sec. So, there was a tradeoff need to be made. We decided to go with HSV color domain. With some image processing techniques Hand and Band are extracted out. But these hands obtained are not ready to use directly. We need to first calibrate for YAW present in hand and for distance of hand from camera.

Scaling Calibration: To make the system robust, it is necessary to make the features of hands same even for hands of different users (hands may be large or small in size) and for different distances from camera. In initial calibration maximum distance between Hand's centroid and hand's boundary points is calculated. Then with a predefined reference value we rescale (expand or, contract depending upon scaling factor calculated from above) the whole hand. It is important to note that in this project we did scaling of hand, but this scaling is not dynamic i.e. user has to keep his hand at the same distance as he kept while initial calibration.

V. REAL TIME SIGN LANGUAGE IMPLEMENTATION

We used the second approach i.e. feature based hierarchical approach for Recognition in real time. The speed from our approach is 3 frames per second. We have recognized 19 different gestures in real time. In next step, we implemented Sign Language. Although numbers of gestures were not that large so we couldn't have implemented all the symbols. In our sign language we have used 3 gestures as transition gesture. Sign language working is explained in detail in next file. In the end we have implemented all alphabets, all numbers and few symbols. Research and Challenges in Statistical Machine Translation of Sign Languages.

While the first papers on sign language translations only date back to roughly a decade and typically employed rule-based systems, several research groups have recently focused on data-driven approaches. In a SMT system has been developed for German and German sign language in the domain weather reports. Their work describes the addition of pre- and post-processing steps to improve the translation for this language pairing. The authors of have explored example-based MT approaches for the language pair English and sign language of the Netherlands with further developments being made in the area of Irish sign language. In , a system is presented for the language pair Chinese and Taiwanese sign language.[11] The optimizing methodologies are shown to outperform a simple SMT model. In the work of , some basic research is done on Spanish and Spanish sign language with a focus on a speech-to-gesture architecture

ISSN: 2277-3754
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 1, Issue 2, February 2012

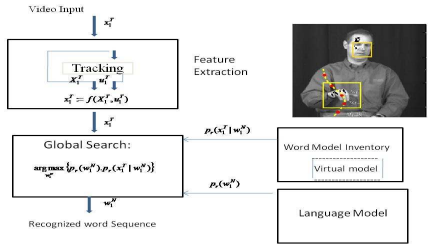

Fig 1: Bayes' decision rule used in ASR and ASLR systems

VI. SPEECH RECOGNITION

Speech recognition requires the computer to accept spoken words as input and interpret what has been spoken. To make the job of understanding speech easier for the computer, a method of speech input called command and control is used. Speech Recognition is technology that allows a computer to identify the words that a person speaks into a microphone. We used Microsoft Agent version 2.0 that provides a library for more natural ways for people to communicate with their computers. And also we used The Lernout & Hauspie True Voice Text-to-Speech (TTS) Engine that provides speech output capabilities for Microsoft Agent so we can hear what the characters are saying through your sound speakers. The commands available to the user are the following: "left click" (or just "click"), "double left click" (or just "double click"), "right click", "double right click". The user can also keep a button pressed so as to highlight a group of objects. The command "down", "up" change the selection area of the mouse per example in the menu (File, Edit, Insert…) and we can select also the menu with the voice command. This set of available commands allows executing meaningful tasks on the computer since all the main mouse click operations are available.

VII. SYSTEM ARCHITECTURE

Humans perceive the environment in which they live through their senses—vision, hearing, touch, smell, and taste. They act on and in it using their actuators such as body, hands, face, and voice. Human-to-human interaction is based on sensory perception of actuator actions of one human by another, often in the context of an environment. In the case of human computer interface, computers perceive actions of humans. To have the human– computer interaction be as natural as possible, it is desirable that computers be able to interpret all natural human actions. Hence, computers should interpret human hand, body, and facial gestures, human speech, eye gaze, etc. Some computer-sensory modalities are analogous to human ones. Computer vision and Automatic Speech Reorganization mimic the equivalent human sensing modalities. However, computers also possess sensory modalities that humans lack. They can accurately estimate the position of the human hand through magnetic sensors and measure subtle changes of the electric activity in the human brain, for instance. Thus, there is a vast repertoire of human-action modalities that can potentially be perceived by a computer. The modalities are discussed under the two categories of human-action modalities and compute sensing modalities. A particular human-action modality (e.g., speaking) may be interpreted using more than one computer-sensing modality (e.g., audio and video).The action modalities most exploited for gesture interpretation system are based on hand movements. This is largely due to the dexterity of the human hand which allows accurate selection and positioning of mechanical devices with the help of visual feedback.

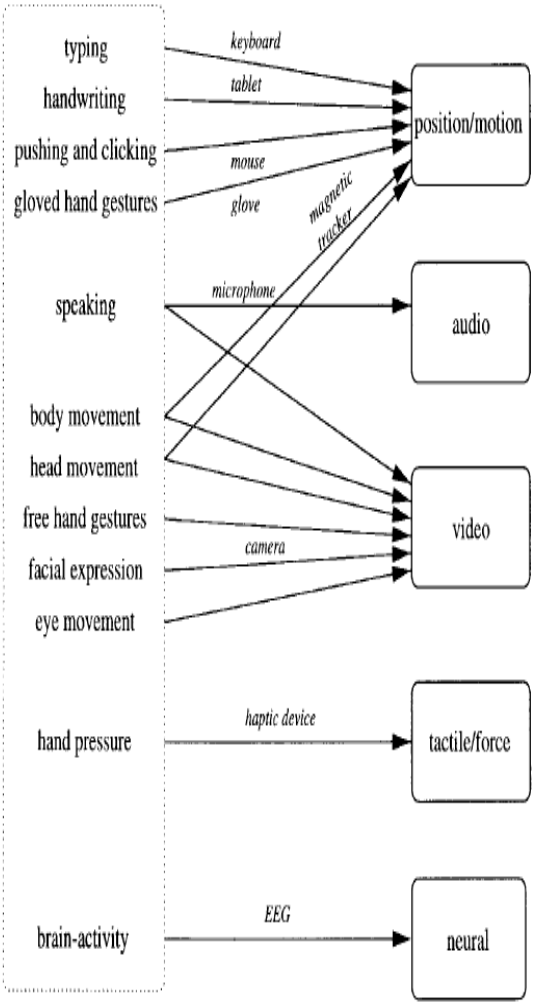

Fig. 2: Mapping of different human-action modalities to computer-sensing modalities for Human Computer Interface

Multiple human actions, such as facial expressions and hand or eye movement, can be sensed through the same "devices" and used to infer different information. Appropriate force and acceleration can also be applied easily using the human hand. Thus, the hand movement is exploited in the design of numerous interface devices—keyboard, mouse, stylus, pen, wand, joystick, trackball,

ISSN: 2277-3754
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 1, Issue 2, February 2012

etc. The keyboard provides a direct way of providing text input to the computer, but the speed is obviously limited and can only be improved to a certain rate. Similarly, hand movements cause a cursor to move on the computer screen (or a 3-D display). The next level of action modalities involves the use of hand gestures, ranging from simple pointing through manipulative gestures to more complex symbolic gestures such as those based on American Sign Language. With a glove-based device, the ease of hand gestures may be limited, but with non-contact video cameras, free-hand gestures would be easier to use for Gesture Interpretation System. The role of free-hand gestures in Gesture Interpretation System could be further improved (requiring lesser training, etc.) by studying the role of gestures in human communication. A multimodal framework is particularly well suited for embodiment of hand gestures into human computer interface.

In addition to hand movements, a dominant action modality in human communication is the production of sound, particularly spoken words. The production of speech is usually accompanied by other visible actions, such as lip movement, which can be exploited in Gesture Interpretation System as well. Where the human is looking can provide a clue to the intended meaning of a particular action or even serve as a way of controlling a display. Thus, eye movements can be considered a potential action modality for Gesture Interpretation System. The facial expression and body motion, if interpreted appropriately, can help in human computer interface. Even a subtle "action" like a controlled thought has been investigated as a potential candidate for human computer interface.

IX. RESEARCH DESIGN

The interaction of humans with their environment (including other humans) is naturally multimodal. We speak about, point at, and look at objects all at the same time. We also listen to the tone of a person's voice and look at a person's face and arm movements to find clues about his feelings. To get a better idea about what is going on around us, we look, listen, touch, and smell. When it comes to human computer interface, however, we usually use only one interface device at a time—typing, clicking the mouse button, speaking, or pointing with a magnetic wand. The "ease" with which this unimodal interaction allows us to convey our intent to the computer is far from satisfactory. An example of a situation when these limitations become evident is when we press the wrong key or when we have to navigate through a series of menus just to change an object's color. Gesture interpretation systems today are unnatural and cumbersome.

However, concurrent use of two or more interaction modalities may loosen the strict restrictions needed for accurate and robust interaction with the individual modes. For instance, spoken words can affirm gestural commands, and gestures can disambiguate noisy speech. Gestures that complement speech, on the other hand, carry a complete communicational message only if they are interpreted together with speech and, possibly, gaze. The use of such multimodal messages can help reduce the complexity and increase the naturalness of the interface for human computer. For example, in computer-vision based gesture recognition, in addition to the input from the images, the gesture recognition could be influenced by the speech, gaze direction, and content of the virtual display. To exploit this multimodality, for example, instead of designing a complicated gestural command for the object selection, which may consist of a deictic gesture followed by a symbolic one (to symbolize that the object that was pointed at by the hand is supposed to be selected), a simple concurrent deictic gesture and verbal command "this" can be used. Another pragmatic reason for using multiple modalities in gesture interpretation system, particularly with redundant input, is to enable physically or cognitively handicapped people access to computers (or computer-controlled devices).. With multimodality built into the gesture interpretation system, the need for building special-purpose interfaces for individual disability will be greatly eased.

X. LIMITATION OF STUDY

In this project we have built system which is reveals real system model, as real system model is cost effective. This project does not consider factor such as speed of moving wheels.

XI. CONCLUSION AND FUTURE WORK

The multimodal system is aimed for the disabled people, which need other kinds of interfaces than ordinary people. In the developed system the interaction between a user and a computer is performed by voice and head movements. To process these data streams the modules of speech recognition and head tracking were developed. This system was applied for hands-free operations with Graphical User Interface in such tasks as Internet communications and lunching applications. We showed theoretically and practically that this technology could be used to operate computers hands-free. Our prototype exhibits accuracy and speed, which are sufficient for many real time applications and which allow handicapped users to enjoy many computer activities. The experiments have shown that in spite of some decreasing of operation speed the multimodal system allows working with computer without using standard mouse and keyboard Thus the developed

ISSN: 2277-3754
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 1, Issue 2, February 2012

assistive multimodal system can be successfully used for hands-free PC control for users with disabilities of their hands or arms .In Arm position Recognition, we have analyzed only horizontal histogram to differentiate 4 hand positions. By analyzing vertical histogram also, we can have 15 different gestures (Arm positions).After improving Scaling, PCA with different distance from camera can be analyzed and if accuracy is high, then can be implemented in real time with video implementation. Training of ANN: Analyzing with different number of layers and different number of neurons per layer, a good trained model can be obtained. This is expected to be more Robust than current hierarchical classification. Extracting more features with Multiple View Point concept, as there are more data points to analyze per histogram or, we can say we have multiple signatures for each gesture and extracting different features from each signature. Hierarchical approach can be implemented together with PCA or, with ANN for better results. This can be in a way that first we find number of fingers in the gesture from which we can have 6 classes and with 6 (1 for each class) different Eigen Spaces for PCA we can expect better results. Apart from Sign language implementation it can be used in many other applications like operating games, creating virtual Keyboard.

REFERENCES

[1] Yan Meng and Yuyang Zhang and Yaochu Jin "Autonomous Self –Reconfiguration of Modular Robots by evolving a Heirarchical Model" IEEE transaction on Computational Intelligence Magzine, pp 43-54 Feb 2011.

[2] S.A. Chhabria and R.V. Dharaskar, "Multimodal interface for disabled persons" in International Journal of Computer Science and Communication, 2011.

[3] Rajeev Sharma, Mohammed Yeasin, Member, Ieee, Nils Rahnstoever, Ingmar Rauschert, Guoray Cai, Member, Ieee, Isaac Brewer, Alan M. Maceachren, And Kuntal Sengupta, ?Speech–Gesture Driven Multimodal Interfaces for Crisis Management Proc Of The Ieee, Vol. 91, No. 9, September 2003.

[4] Boucher, R. Canal, T.-Q. Chu, A. Drogoul, B. Gaudou, V.T. Le, V.Moraru, N. Van Nguyen, Q.A.N. Vu, P. Taillandier, F. Sempe, and S. Stinckwich. :"A Real-Time Hand Gesture System based on Evolutionary Search". In Safety, Security Rescue Robotics (SSRR), 2011 IEEE International Workshop on, pages 16, 2011.

[5] M. Segers, James Connan, "Real-Time Gesture Recognition using Eigenvectors" Vaughn Private Bag X17 Bellville, 7535, volume III,2009.

[6] Rami Abielmona ,Emilm.Petriu, Moufid Harb and Slawo Yesolkowki,"Mission Diven Robotics for Territorial Security" Model"IEEE transaction on Computational Intelligence Magzine ,pp 55-67 Feb 2011.

[7] Melody Moh, Benjamin Culpepper,Lang Daga, "Computer Vision and Pattern Recognition "IEEE , 2005. CVPRW '05.Conference on, page 158, June 2005.

[8] Boukje Habets, Sotaro Kita, Zeshu Shao, Asli Özyurek, and Peter Hagoort "The Role of Synchrony and Ambiguity in Speech–Gesture Integration during Comprehension" 2011.

[9] R. Sharma, V. I. Pavlovic, and T. S. Huang, ?Toward multimodal human-computer interface, Proc. IEEE, vol. 86, pp. 853–869, May 1998.

[10] R. Stiefelhagen, C. F¨ugen, P. Gieselmann, H. Holzapfel, K. Nickel and A. Waibel ?Natural Human-Robot Interaction using Speech, Head Pose and Gestures Proceedings of the Third IEEE International Conference on Humanoid Robots - Humanoids 2003.

[11] Benoit Legrand, C.S. Chang, S.H. Ong, Soek- Ying Neo, Nallasivam Palanisamy, ?Chromosome classification using dynamic time warping , ScienceDirect Pattern Recognition Letters 29Dec 2008.

[12] Y. Tamura, M. Sugi, J. Ota, and T. Arai, ?Estimation of user's intention inherent in the movements of hand and eyes for the deskwork support system, in IEEE/RSJ IROS, (USA), pp. 3709–3714, Nov. 2007.

[13] CHIU, Y.-H., C.-H. WU, H.-Y. SU and C.-J. CHENG: Joint Optimization of Word Align-ment and Epenthesis Generation for Chinese to Taiwanese Sign Synthesis. IEEE Trans. PAMI, 29(1):28–39, 2007.

[14] STEIN, D., J. BUNGEROTH and H. NEY: Morpho-Syntax Based Statistical Methods for Sign Language Translation. In 11th EAMT, pp. 169–177, Oslo, Norway,