Avatar-based Sign Language Training Interface for Primary School Education

Rabia Yorganci[1] Ahmet Alp Kındıroğlu[2] Hatice Kose[1]
[1]Istanbul Technical University [2]Boğazici University
34469 Istanbul, Turkey 34342 Istanbul, Turkey
{rabia.yorganci,hatice.kose}@itu.edu.tr, alp.kindiroglu@boun.edu.tr

ABSTRACT

The aim of this study is to present usable educational tools of students with special needs. Hearing impaired students cannot use text and speech based technological, educational material that is becoming a crucial tool for modern education. Like videos of signers, a signing avatar for TİD (Turkish Sign Language) would allow communicating information in the form of visual gestures, thus becoming usable when the use of text is unfeasible. However, compared to videos, animated avatars offer several advantages like easy reproducible of gesture sequences, control of point of view, adjusting the speed of the sign and being smaller in storage and bandwidth then videos. For this study, a signing avatar was developed to represent a portion of the social sciences course for Turkish primary education curriculum. To analyze the success and effectiveness of the interface, the performance of the text, and avatar-based interaction in a human-computer interaction scheme using elementary school students have been compared. The results demonstrate that avatar based tutoring was more effective in assessing the child's knowledge of certain sign language words. Since the aim was to make the signing avatar as comprehensible as possible, the results demonstrate that goal have been.

KEYWORDS

Human-computer interaction, Sign language synthesis, Children's education, Language education, Avatar

1 INTRODUCTION

In the cognitive development process of children, primary school education plays a crucial role. Studies indicate that a lack of access to primary school education for deaf children causes serious problems. Some of these problems are lack of language skills, presence cognitive impairments, problems in socially engaging with their peers and over-dependence on family members [1]. Unlike hearing children, who mostly learn their spoken language skills before starting primary school, deaf children often start primary school without knowing a spoken or sign based language. This can negatively affect age appropriate literacy learning, making an extra effort for education necessary.

One of the common approaches for the education of hearing-impaired children is teaching sign language based communication first and teaching written and spoken languages after, with the help of sign language. Sign languages are the communication medium of the hearing impaired. They are visual languages that rely on the use of the positioning and movements of hands, arms and upper body as well as facial gestures to convey meanings of concepts. Being designed to convey meaning rapidly in 3d space, sign languages do not have a well-defined written form that is practical. For that reason, the creation of educational support material such as books and worksheets designed to aid and asses learning are mostly done using videos which are hard to create, edit, store and transfer.

In primary school education, the influence of classical written materials and technological materials are important as they al-

low students to continue the learning and assessment processes without the direct attention of the instructor. In addition, by using technologies in the classroom, it has been shown that student learning and motivation for learning could increase. However, deaf children who cannot use and understand books or computer based material only participate in learning when an instructor or parent directly engages with them face to face. The exception to this is sign language videos, which are used by instructors to create materials and replay signs. However, the creation and storage and usage of videos all require great effort.

In this study, a signing avatar based tool that would serve as a method for generating video like materials for the primary school education of hearing impaired children is introduced. Compared to video-based materials, the avatar would have the advantage of being easier and less costly to create large amounts of materials, being easier to edit content, being easier to transfer to students allowing them to study at home with their parents and being easier to store. Content from the social sciences book of the Turkish primary school education was generated using the avatar and tests from the book were used to evaluate the system. The avatar was evaluated in terms of comprehensible by comparing it to textual input in school children. Children from grades three to five were asked to comprehend social studies questions in the form of the avatar based TİD signs textual questions. The students then chose the correct answer from among multiple answers in the form of images. The rest of the paper is organized as follows: In Section 2 the related works in the field are presented. In section 3, the signing avatar and the social science quiz methodology are explained. Finally, in sections 4 and 5, the experimental results are given and conclusions are presented.

2 RELATED WORKS

Using technological tools in education research is the preferred method, especially for researchers studying with children. There are various commercial games for education and therapy of babies and infants both on mobile devices [2], [3] and desktop computers [4]. Several gaming-based projects have been designed for the therapy and education of children with special needs, too. As a part of this area, there are different studies developed for hearing-impaired children [5]–[8]. The ICICLE (Interactive Computer Identification and Correction of Language Errors) project aimed to create an educational system for the deaf children to provide them with individual lectures and guidelines by computer-aided commands [9]. There are various assistive applications for children with autism focusing on imitation, face and emotion recognition and speech therapy [10]–[12].

Unlike spoken languages which could be represented through text or speech, sign languages are complex linguistic structures that are composed of complex visual and spatial elements. The representation of multiple parts of the human body and their motion in 3D space around the signer cannot be represented efficiently in written form. For that reason, generation of novel sign language material is often a tedious task that requires the presence of people performing gestures in front of a camera.

While there exist markup languages such as Hamnosys or SIGML [13] to represent sign languages, they are often inefficient both for writing and for reading. Therefore, to digitally generate sign language material and for interaction purposes, the use of human-like digital output mediums becomes a necessity. Development of virtual reality human models or avatars for sign language is a relatively novel research area. With the progress of research in computer graphics, it is technologically possible to create human animations capable of performing sign languages. In computer graphics, many types of animated

human characters have been built for different purposes from automobile and aircraft crash simulators to motion analysis, gaming, and entertainment. In [14], Webber et.al. summarizes the desired properties of human models as being structured like a human, moving like a human, looking like a human and having the same size and movement restrictions as a human. In the market for human model development, there exist many modeling tools capable of creating high-quality models. These can be prioritized according to their category. Blender, Zbrush, 3ds Max, and Maya are more artistic means, while Rhino, Solidworks, Autodesk are more engineering oriented. The models created by these tools can be realized and animated using graphics libraries like OpenGL and DirectX or used with 3d engines such as Unity or Unreal 4. The level of quality in computer generated human models have long since passed the point that allows the building interfaces, which can present information from computers putters deaf via avatars in real time. However, such signing avatars are not readily available in everyday life. The main hurdle for the development of such software lies with the challenge of providing the applications the right instructions to generate linguistically comprehensible information.

According to Kipp et.al., existing methods to translate written languages to animation can be grouped into two broad approaches as concatenative and articulatory methods [15]. In the concatenative approach, movements of concepts in sign language are created individually. Therefore, to represent a sentence in Sign Language, some concatenation of these individual signs becomes necessary with possible smoothing to make transitions natural. There are several examples of avatars using concatenative methods [16]–[18]. As a more natural alternative to concatenative methods, the articulatory methods involve computationally generating the sequence of signs at runtime. There exist many avatars using articulatory synthesis such as GesSyCa [19], SignSynth [20]. Articulatory approaches were also used in the EU projects VISICAST and eSIGN to allow the conversion of written material to avatar animation for the hearing impaired. In the VISICAST project [21], the focus was on converting written material to sign languages. In that project, HamNoSys notation was used to represent sign languages. In the eSIGN project, which followed VISICAST the transition from concatenated to computational methods was achieved through the utilization of a markup language called SIGML. Motion/Sign Capture tools were used to capture upper body movements and facial expressions to represent them using a 3D avatar. In this project, text written in English were translated to the British, German and Dutch sign languages (BSL, GSL, DSL). In the Text and Sign Support Assistant project developed as a part of VISICAST, a post office application was developed using an avatar signing in BSL [22]. The ESIGN project was commissioned with the goal of translating written online materials in English to BSL using an avatar. As an output of the project, an application was developed with a weather forecast themed vocabulary. In the development phase of the application, first, the content of the original web page was translated to BSL via linguist specialists. The corpus generated by this translation was first manually made using the created avatar and added to a dictionary. The final output of the website was presented by displaying the corresponding content found in the dictionary of the avatar on the web page. Another EU project following in the footsteps of VISICAST and ESIGN was DictaSign [23]. This project made use of human's video recordings instead of avatar generated synthetic videos. It recorded corpora in Greek, British, German and French Sign Languages and focused on the conversion of these to one another and their corresponding written languages.

3 SOCIAL SCIENCE QUIZ

The aim of this project is creating avatar-based interactive software for the education of deaf children. To understand their attitude toward interaction with other people especially in their educational environment and to ensure that the truth of the animation, consultants whose native languages are TİD were consulted and each step of the project proceeded with their guidance and approval.

3.1 Research Questions and Hypothesis

Avatar based education tools and assessing methods were designed as a part of sign language training of deaf children at early ages. The interface was customized specifically for users who are hearing impaired and have reading difficulties. These experiments were designed to find the answers to two major research questions;

- The use of facial expressions in TİD: Differentiating non-verbal components of signs from facial expressions that are independent of the performed sign is an important part of corpus generation. Distinguishing the minimum required modalities of a sign from user specific additions allows the construction of an avatar that is concise. This process minimizes the confusion of children when they observe signs performed by an avatar.

- Effectiveness of TİD in primary school education: Deaf children learn spoken languages slower than sign languages. In the elementary education, teaching tools and curriculum are designed with the spoken language in mind. Therefore, when learning novel content, deaf children have to learn it in a language that they are not comfortable with. The use of a signing avatar enables instructors to teach their course to deaf children without the language barrier they would face when following the same curriculum using books.

3.2 Animation Generation

In this project, a free character for non-commercial usage, Mery[1], is used. She was created in Maya[2]. Maya was developed for 3D animation, modeling, simulation, and rendering software by Autodesk. The software offers an advanced modeling tool to create scenes and animations. Animations could be created either manually and automatically through a motion capture device. Maya supports Python and MEL script languages to develop animation systematically. Animations made for mobile and web-based applications were created using Maya. Besides Maya, using MotionBuilder[3], another software from Autodesk, is an option to create automatic animations.

3.3 Implementation on the Avatar

To generate the signs on the avatar, Mery, experiments were performed with both automatic and manual approaches. The primary software used for the generation of the avatar is Maya. In the studies prepared for this paper, manual implementation method has been used.

3.3.1 Manual generation of TİD sentences using Maya software

Maya lets the user create animations frame by frame. For upper body joints except face (such as eyes, mouth, nose) rotation (X, Y, and Z axis) values are used for creating animations. For face gestures, translation (mostly Y axis) values are used. .
In addition, by using a software called Trax Editor, a library of animation clips could be created for a character set and the clips could be used on different signs, whenever they were needed. This tool made creating sentence manually method easier on Maya.

[1] www.meryproject.com/
[2] www.autodesk.com/products/maya/
[3] www.autodesk.com/products/motionbuilder/

Besides the Trax Editor, different plugins were available. StudioLibrary[4] was used during the implementation phase.

3.3.2 Sign Library

To prevent the need to create the same hand shapes, arm postures and facial expressions for different signs, a library which includes poses and animations were created using Studio Library on Maya. The library for sign language is divided into three main sections; Animation (includes torso and arm animations, Fig. 1(a)), Face (poses, Fig. 1(b)) and Hand (poses, Fig. 1(c)). These poses and animations could be combined in any order. To create a sign or sign sentence using a number of previously created poses, the necessary body parts and time frame are selected. Interpolations between poses are calculated for each body part and animations for the selected joints are generated. Using this method, the process of creating new signs or sentences were greatly accelerated.

(a) Animation

(b) Face

(c) Hand

Figure 1. Examples from the Sign Library

3.4 Preliminary Test Scenarios

To evaluate usability and acceptability of Mery in a Sign Language Education environment, two test cases which targeted the torso and the face were designed. During the preparation phase of this tests, manual generation of signs using Maya software was used.

The first test included just the torso [24]. These tests were designed to teach signs to non-signers and evaluate the comprehensibility of those TİD gestures without facial expressions. Selected words were prepared on Maya, and a questionnaire was drawn up on the framework whose details are given in [7].

The second set of tests aimed to increase the comprehensibility of the avatar by adding facial expressions. In sign language, face gestures include facial expression of the six primary emotions. After observing the sign language translation of the primary school written educational material, it was concluded that there are also other facial expressions frequently used in TİD, such as bored, and question. These expressions were implemented on Mery with using Maya to combine torso gestures. [25]

3.5 Interaction Scenario

Based on the results of the isolated sign based preliminary tests on face and torso, sentence tests were created. These sentences were generated with the aid of the Sign Library. The trials began with demographic information questions; name, age, gender, and level of TİD knowledge. To evaluate the success of Mery, three test versions were created which only had text, text and human's video and text and Mery's video. With this test, the aim was to observe the effectiveness of the signing avatar as an educational tool (only-text vs. text with videos) and to measure the comprehensibility of the animated character (human vs. Mery). An example test question is given in Fig. 2.

4 EXPERIMENTS AND RESULTS

4.1 Experimental Setup

A Social Science Course test was designed for primary school children. First, ques-

[4] www.studiolibrary.com

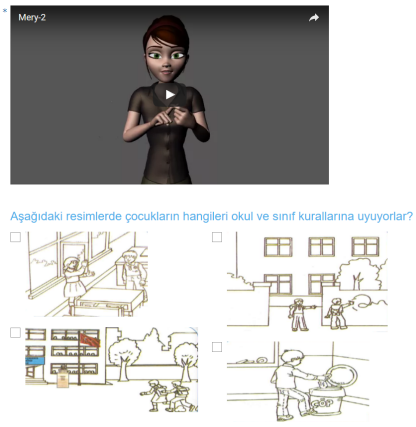

Figure 2. Example test question on Social Science Quiz

tions were selected from the class-book offered by the primary school curriculum of Turkey. These sentences were performed by a sign language instructor in front of a video camera. In addition, skeleton coordinates of all words in the book were collected with Kinect 2 sensor as isolated signs.

Selected sentences were created on Mery using Maya. After the creation phase, web-based questionnaires were prepared. Three version of the questionnaire were designed, which differed by question format; 1. Only text, 2. Text and human's video, and 3. Text and Mery's video. In the experiments, the first and the third formats were tested.

The experiments with primary school children were conducted in a room at the school for the hearing impaired. One experimenter had an assistive role of explaining the rules to each child one-by-one. Children completed the test one child at one time. Therefore they were not exposed to the questions before the test. Images that were the answers of the questions were copied from directly from the social studies class-book. 25 students age of 9 – 12 (average age: 10.12±1.37) from Dost Eller Primary School for Hearing Impaired Children completed the tests. Of the 25 attendees, 13 were girls and 12 were boys. 15 of attendees specified that they were familiar to TİD.

4.2 Results

In this study, questions with avatar based sign videos with texts were compared with text-only questions and preliminary results were obtained. Questions on this experiment were multiple choice questions meaning any question could have more than one right answer. Number of correct and wrong answers per question are given on Table 1.

Table 1. Number of correct and wrong answers of the sentence based test

#	Correct	Wrong
1	2	2
2	2	2
3	2	1
4	1	1
5	1	1
6	2	1
7	2	2
8	1	2
9	1	2
10	1	2
Total	15	16

Every participant completed the test session without any interruption or help from their instructor. The questions in the experiments were taken from the first-grade social studies book. However, since hearing impaired children obtain the necessary communication skills for the book at the third to fifth grades, attendees of the experiment were selected from those classes. Overall accuracy ratios of the experiments are presented on Table 2. As can be observed from the table, the ratio of the correct answers was higher for the experiments with the avatar based sign videos and lower for the text only version (Fig. 3).

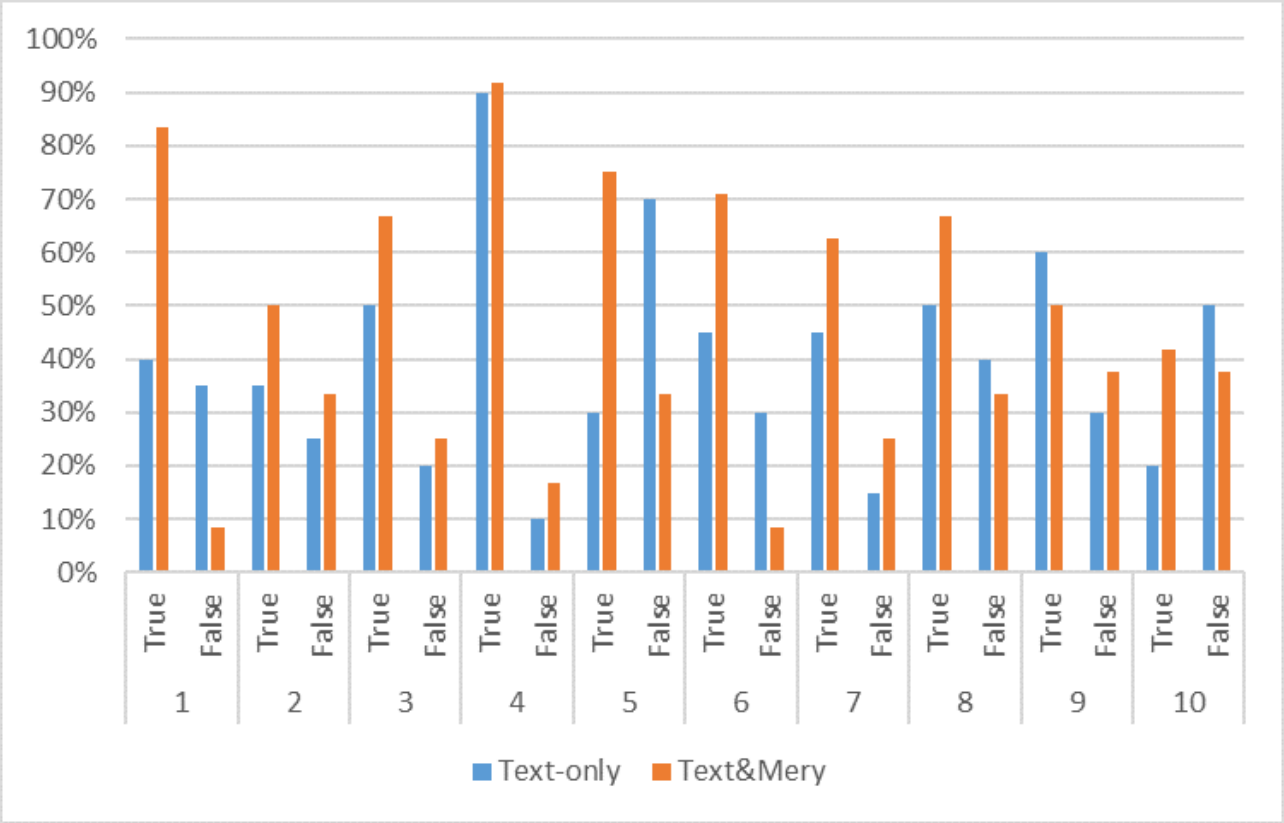

Figure 3. Question-based true and false answers

Table 2. Overall accuracy ratios

	Correct	Wrong
Text-only	45.33%	32.50%
Text&Mery	66.11%	27.08%

4.3 Discussion

The social science test with avatar was prepared on three phases; word-based signs video test includes arms and torso [24], face gesture tests [25] and sentence-based test to evaluate the effect of supported material to textbooks for primary school students. First, two phases were initial steps of the final test.

Hearing-impaired children have adaptation problems on a regular curriculum which is usually based on the text. Learning literacy is a challenging obstacle for these children. Although most students are learning it in the first year of the primary school, the average age of learning is higher for these special children. Because of that, hearing-impaired children lag behind their peers in education. According to teachers, parents of deaf children refuse sign language training in schools. This opinion postpones sign language courses in primary schools. Despite the age of participants is approximately 10, almost half of them (Table 3) could not speak in TİD. Also, most of these children have not ability on spoken Turkish. Even if some of the students claimed that they know TİD, lack of their abilities is observed during the test.

The experimenter children did not know some fundamental concepts like natural disasters and abstract concepts such. While they were taking the test, some of them asked the meaning of “earthquake”, “hibernation”, etc. both in spoken Turkish and TİD. The curriculum assumes these concepts are included in the first grade which causes a greater problem for deaf kids compared to their hearing counterparts.

Fourth question (“Please find Anıtkabir, the

Table 3. TİD knowledge level of the participants

	Text-only	Text&Human	Text&Mery	Overall
Do not know	2	0	1	3
I know a little	3	1	3	7
I know	5	2	8	15

resting place of Atatürk") has highest accuracy rate on both scenarios. Besides the overall accuracy ratio, the gap between accuracy ratios of the first and the fifth questions ("Which tools do you think the kids in the picture require to paint a painting? Please check the appropriate boxes." and "Which of the children shown below has not been inoculated?") with the other questions show that the children did not know of the concepts present in these questions.

5 CONCLUSIONS

Avatar-based Sign Language Training Interface is a part of an ongoing nationally funded project which aims to create a TİD based educational tool and dictionary specialized on the primary school curriculum. As preliminary experiments, usability tests for the arms and facial expressions were conducted. Prior to getting the responses from the participants, sentences selected from the Social Science Exercise Book of the first-year student of primary school were created. The book was used for usability studies of the avatar with the primary level education kids. The test was designed with the aim of comparing the educative effectiveness of the avatar compared to textual educational tools.

The results indicate that avatar based tutoring was more effective in assessing the child's knowledge of certain sign language words. One major source of the problem in both avatar and textual based assessment was that the children were not familiar with the curriculum. For example, they did not know what an earthquake was. These experiments demonstrated that signing avatar was a viable alternative to textual educational materials for hearing-impaired children.

An overall look at the interaction experiment scheme demonstrates that the use of virtual non-textual tools is useful in the learning process of children. Avatar technology, which has been shown to be a close and successful imitation of videos may be used to fill that gap, as they may provide a more readily available and cheaper support to the education of primary school children with special needs.

6 FUTURE WORKS

One of the limitations of the work in this thesis was that the corpora were limited to a small vocabulary size and interactions and learning experiments with children did not exceed a limited time. While that was sufficient to prove the hypothesis of this project, the completion of the overall project corpora will allow researchers to examine the effect of using technological aids for the period of an entire semester. Such a work would require a significant amount of collaboration and monitor with randomized tests to keep track of student improvement at different intervals.

One step that will be completed towards achieving the overall project corpora is the implementation of fully automatic generation of avatar based content using Kinect camera and leap motion sensors. By reducing Maya based involvement to a minimum, avatar videos will be produced much faster by linguists without a need for designer involvement.

Acknowledgement

This study is a part of an ongoing project supported by the Scientific and Technological Research Council of Turkey under the contract TUBITAK FATIH 114E263.

REFERENCES

[1] C. Vaccari and M. Marschark, *Communication between parents and deaf children: implications for social-emotional development*, 1997. DOI: 10 . 1111 / j . 1469 - 7610.1997.tb01597.x.

[2] *Lingling language learning game.*

[3] *Mybabydrum android application.* [Online]. Available: http://www.dooet.com / android / My - baby - drum - 3376814129.

[4] *Little mozart 2 - lirec project.* [Online]. Available: http://mozart.imagina.pt/.

[5] N. Adamo-Villani, "A virtual learning environment fordeaf children: design and evaluation.," *IJASET -International Journal of Applied Science, Engineering,and Technology*, vol. 16, pp. 18–23, 2006.

[6] P. Uluer, N. Akalin, R. Yorganci, H. Kose, and G. Ince, "A new robotic platform as sign language tutor," in *IEEE/RSJ International Conference on Robots and Intelligent Systems (IROS) International Workshop on Assistance and Service Robotics in a Human Environment*, 2013.

[7] A. Ozkul, H. Kose, R. Yorganci, and G. Ince, "Robostar: an interaction game with humanoid robots for learning sign language," in *IEEE International Conference on Robotics and Biomimetics (ROBIO)*, IEEE, 2014, pp. 522–527.

[8] H. Köse, P. Uluer, N. Akalin, R. Yorganci, A. Özkul, and G. Ince, "The effect of embodiment in sign language tutoring with assistive humanoid robots," *International Journal of Social Robotics*, vol. 7, no. 4, pp. 537–548, 2015.

[9] C. Greenbacker and K. McCoy, "The icicle project: anoveriew.," In: First Annual Computer Science ResearchDay, Department of Computer and InformationSciences, University of Delaware, Tech. Rep., 2008.

[10] *Autismspeaks mobile application.*

[11] K. Dautenhahn, C. L. Nehaniv, M. L. Walters, B. Robins, H. Kose-Bagci, N. A. Mirza, and M. Blow, "Kaspar - a minimally expressive humanoid robot for human-robot interaction research," *Special Issue on "Humanoid Robots", Applied Bionics and Biomechanics*, vol. 6, no. 3, pp. 369–397, 2009.

[12] S. Powell, *Helping Children with Autism to Learn.* London: David Fulton Publishers, 2000.

[13] R. Elliott, J. Glauert, R. Kennaway, and K. Parsons, "Visicast deliverable d5-2: sigml definition," ViSiCAST project, Tech. Rep., 2001.

[14] B. L. Webber, C. B. Phillips, and N. I. Badler, "Simulating humans: computer graphics, animation, and control," *Center for Human Modeling and Simulation*, p. 68, 1993.

[15] M. Kipp, A. Heloir, and Q. Nguyen, "Sign language avatars: animation and comprehensibility," in *Intelligent Virtual Agents*, 2011, pp. 113–126.

[16] J. A. Bangham, S. J. Cox, R. Elliott, J. R. W. Glauert, I. Marshall, S. Rankov, and M. Wells, "Virtual signing: capture, animation, storage and transmission-an overview of the visicast project," in *IEEE Seminar on Speech and Language Processing*

for Disabled and Elderly People, IET, 2000, pp. 1–6.

[17] M. Verlinden, C. Tijsseling, and H. Frowein, "A signing avatar on the www," in *Gesture and Sign Language in Human-Computer Interaction*, Springer, 2001, pp. 169–172.

[18] H. Sagawa, M. Ohki, T. Sakiyama, E. Oohira, H. Ikeda, and H. Fujisawa, "Pattern recognition and synthesis for a sign language translation system," *Journal of Visual Languages and Computing*, vol. 7, no. 1, pp. 109–127, 1996. DOI: 10.1006/jvlc.1996.0007.

[19] T. Lebourque and S. Gibet, "High level specification and control of communication gestures: the gessyca system," in *Computer Animation, 1999. Proceedings*, IEEE, 1999, pp. 24–35.

[20] A. B. Grieve-smith, "Signsynth : a sign language synthesis application using web3d and perl," *Synthesis*, pp. 134–145, 2002, ISSN: 16113349. DOI: 10.1007/3-540-47873-6_14.

[21] *Visicast at uea*. [Online]. Available: http://www.visicast.cmp.uea.ac.uk (visited on 04/30/2016).

[22] S. Cox, M. Lincoln, and J. Tryggvason, "Tessa, a system to aid communication with deaf people," *Proceedings of the fifth international ACM conference on Assistive technologies. ACM*, 2002.

[23] E. Efthimiou, S. E. Fontinea, T. Hanke, J. Glauert, R. Bowden, A. Braffort, C. Collet, P. Maragos, and F. Goudenove, "Dicta-sign–sign language recognition, generation and modelling: a research effort with applications in deaf communication," in *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, 2010, pp. 80–83.

[24] R. Yorganci, N. Akalin, and H. Kose, "Avatar tabanlı etkileşimli işaret dili oyunları," in *Uluslararası Engelsiz Bilişim 2015 Kongresi*, Manisa, 2015.

[25] E. Aklan and N. Tarlakazan, "Avatar based interaction studies gestures and facial expressions in sign language with virtual avatars," Istanbul Technical University, Tech. Rep., 2016, Graduation Project.