A Cascaded Speech to Arabic Sign Language Machine Translator using Adaptation

Shereen A. Mohamed
Department of Mathematics and Computer Science, Faculty of Science, Alexandria University, Alexandria, Egypt

Mohamed A. Abdou
Informatics Research Institute, City for Scientific Research and Technology Applications, Alexandria, Egypt

Y. F. Hassan
Department of Mathematics and Computer Science, Faculty of Science, Alexandria University, Alexandria, Egypt

ABSTRACT
Cascaded machine translation systems are essential for Deaf people. Speech recognizers and sign language translators when combined together constitute helpful automatic machine translators. This paper introduces an automatic translator from Arabic spoken language into Arabic sign language. This system aims to integrate Deaf students into classrooms of hearing ones. The proposed system consists of three cascaded modules: a modified Arabic speech recognizer that works using adaptation, an Arabic machine translator, and a developed signing avatar animator. The system is evaluated on real case studies and shows good performance.

General Terms
Machine translation, Arabic Sign Language(ArSL), signing avatars

Keywords
Arabic Sign Language (ArSL), Deaf students, Example-based translation system, signing avatars

1. INTRODUCTION
Deafness is the most impairment that causes communication barriers between people. Sign language interpreters play a considerable role in reducing these barriers, but sometimes they are not available or their services become expensive.

Signing videos have appeared as an assistant tool that works side by side with interpreters, and sometimes be an alternative. However, they can't be considered an ideal alternative because it costs a lot to produce high quality videos. Also, it becomes cumbersome when the contents are changeable. Moreover and for consistency purposes, when making adjustments, the new amendments have to be recorded using the same signing person in the same clothing and with the same background. Add to that, the large disk spaces required to keep these videos, and the cost and time consumed for downloading signing videos especially when using dial up connections [1].

The constant evolution in the fields of computer hardware and multimedia has encouraged researchers to pay more efforts on development of signing avatars to address problems of signing videos. The developed avatars have offered the following features: The production cost of avatar-based contents is less. Signs sequences are easily modifiable. Consistency is not a problem; different signing avatars can be animated by the same signing data. The required disk space is much smaller. Negligible bandwidth demands are required to transmit avatar-based contents over a network. The rendering speed and view angle of signing avatars are adjustable. And finally, signing avatars keep the anonymization of the producer [1, 2].

This paper describes an avatar-based translation system from Arabic into Arabic Sign language (ArSL). The system tries to solve some of the problems that face Deaf students when they enroll in training courses, and facilitate the possibility of integrating Deaf students into classrooms of hearing ones and make them share the same training materials. As will be presented in next section, there are similar systems for ArSL, but this system is differentiated in developing and integrating: 1) An adapted Arabic speech recognizer. 2) An Arabic into ArSL translation algorithm. 3) A parallel corpus.

2. RELATED WORK
After reviewing the current literature, efforts made to facilitate communication between Deaf and hearing Arabic people can be classified into: 1) Efforts to accurately recognize ArSL. 2) Efforts to translate Arabic into ArSL.

Regarding translation into ArSL, the developed systems are scarce, and most relied on the use of images and signing videos in expressing the translation. Halawani et al [3] in 2013 developed a speech to ArSL translation system to translate the Arabic alphabet and numbers. The system used Microsoft windows 7 speech recognition engine to receive the uttered speech, and jpg format images to express the ArSL translation. Almohimeed et al [4] in 2011 developed a system that translated Arabic text into ArSL. The system translated complete sentences, and represented the translation by means of signing videos. Mohandes [5] in 2006 developed an Arabic into ArSL translation system. The system is based on a database of signing videos. Once the user enters an Arabic word that is exist in the system database, its corresponding signing video is played. Otherwise, the word is finger spelled using a sequence of images representing each letter of the word.

3. SIGN LANGUAGES
Sign languages are mother languages and primary means used by Deaf to communicate with each other inside their communities and with the hearing communities. They are complete languages; each has its own grammar and phonology, independent and unrelated to the national spoken languages. Contrary to what many believe, they differ from country to country and sometimes from region to another within the same country. As with hearing communities, communities with close cultures can influence their respective sign languages. For example, the British SL has influenced greatly Australian and New Zealand SL and is now considered as one known as BANZSL (BSL, Auslan, and NZSL) [6]. Although there are numerous attempts to express sign languages in written form, there is no official written form for these languages [7]. Sign languages are visual

languages; they depend on the use of the 3D space surrounding the signer along with different signer's body parts to express a sign, the basic grammatical unit of a sign language [6].Sign languages linguists have shown that signs consist of two components [7, 8]: 1)*The manual components* which are the main components that include hand configuration, place of articulation, hand movements, and hand orientation. 2) *The nonmanual components* which are the additional components that include facial expression and body posture

Sign languages are complex, the same sign can be used to express the meanings of different words, and some words have no equivalents [9]. Also, they are accurate languages; any slight difference in the signing process can lead to entirely different meaning. E.g., in ArSL, both words "Easy" and "Not important" are expressed using the same sign, but with a small variation in lip movement.

4. CHALLENGES FACING TRANSLATION INTO ANIMATED ARSL

Signing avatars are still far from being real signers [10]. Although the development of many signing avatar projects and the innovation of different technologies to control the animation, challenges for giving these signing avatars the natural human movements, face expressions, and lip movements still need more research. Also, there are no standard evaluation methods to assess the signing avatar performance [11]. Researchers depend on the extent to which the signs are understood by the Deaf users in evaluating the signing process. Despite the importance of Deaf feedback, the existence of scientific measurements is necessary to evaluate the accuracy of signing, compare between different avatars, and measure the progress done in the way of producing a lifelike signing avatar.

Developing systems that translate Arabic into ArSL faces many problems. These problems can be classified into:

- Inexistence of Arabic speech recognizers and/or validated Arabic phoneset
- Lack of Arabic Sign Language documentation or corpora
- Inadequate number of research publications

5. A PROPOSED CASCADED ARSL TRANSLATOR

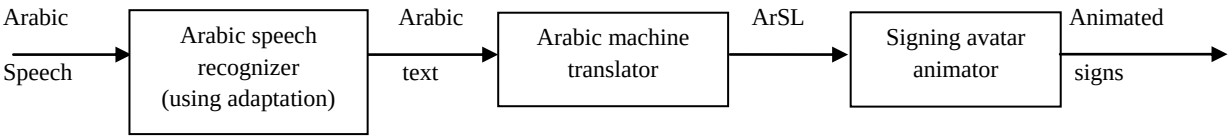

Fig 1: The proposed system components

5.1 Arabic Data Set

The proposed system tries to provide a training course in classrooms including both Deaf and hearing students. The Microsoft Word 2007 program has been selected as the translation domain. Reviewing the program, 54 basic sentences, describing how to execute different tasks, and explaining different components of the program's main window, have been constructed. These sentences have been translated into sign language, and a video for each signed sentence has been recorded, by a team of three ArSL experts. A file containing each sentence with its corresponding ArSL translation has been created. This corpus was increased to 205 sentences by incorporating different variants for Arabic sentences (maintaining the ArSL translation). The signed sentences were written using gloss notation, where each sign was represented by its meaning in Arabic.

5.2 Structure of the Cascaded System

The proposed Arabic to ArSL translation system consists of three basic modules: an adapted Arabic speech recognizer, an Arabic machine translator, and a signing avatar animator module. The module diagram of the system is shown in figure 1. The main function of the first module, the Arabic speech recognizer, is to receive the Arabic speech and decode it to its textual form. The objective of the second module, the Arabic machine translator, is to receive the Arabic text produced by the previous module and translate it to its equivalent in ArSL by searching a database of examples to find an equivalent or a similar enough to the Arabic text. Finally, the third module, the signing avatar animator module, expresses the translation graphically by playing the sequence of signs by means of a signing avatar.

5.2.1 An Introduced Arabic Speech Recognizer Using Adaptation

This module receives the uttered Arabic sentence entered via a microphone and converts it to its written form. The module makes use of the CMUSphinx speech recognition toolkit [12], a widely used Hidden Markov Model-based speech recognition toolkit developed at University of Carnegie Mellon, to achieve its function. The process of decoding speech requires three main files: an acoustic model that contains acoustic properties for each phone, a language model that restricts word search, and a dictionary that contains mapping from words to their phones. Figure 2 shows the module components. The main problem faced the development of this module is that the toolkit doesn't support Arabic language. So to solve this problem, the existing US English acoustic model has been adapted to recognize Arabic. The adaptation process could be summarized as follows:

- Each of the 205 Arabic sentences in the system corpus was recorded to a separate file, at a sampling rate of 16 KHz in mono with single channel following the adaptation instructions of the library.
- These recording were processed to extract the acoustic features required to adapt the existing US English model

- An Arabic language model was created for determining which word could follow previously recognized words
- An Arabic dictionary, describing the pronunciation of all the Arabic words in the system corpus, was created using the English phoneset.

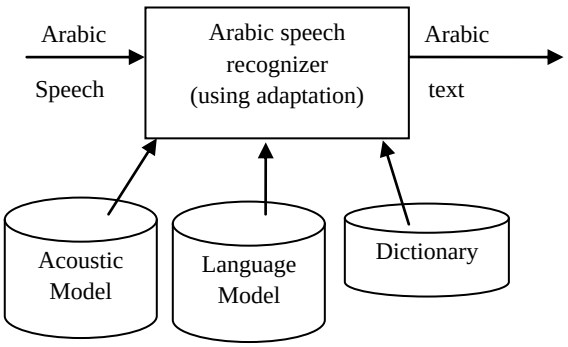

Fig 2: The Arabic speech recognizer module's components

5.2.2 Creation of an Arabic Corpus

Here, we introduce an Arabic machine translator module that receives the Arabic text recognized by the previous module and translates it to its corresponding in the ArSL. Despite the presence of several translation prototypes, most of them are not suitable for ArSL because of two problems: 1) The Lack of ArSL Corpora. 2) The Lack of grammar and structure studies on ArSL. In this work, the example-based prototype was assumed to overcome these problems.

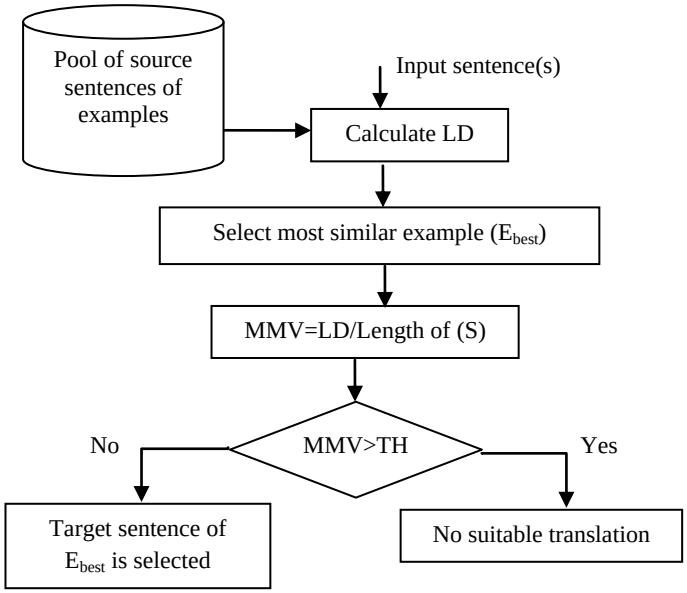

Fig 3: An illustration of the translation process

As the EBMT model depends on the existence of a bilingual corpus, a file containing the 205 Arabic sentences (the source language of the system) and their corresponding translation in ArSL (the target language of the system) has been created.

The ArSL sentences have been written using the glosses, where each sign has been represented by its similar meaning in the Arabic language. It is important to highlight that it was not usual to represent a sign by a gloss because most of the terms used within the Microsoft word program like "Title bar", 'Status bar", and others do not have specific signs in ArSL, so they have been described semantically using sequence of signs for each.

Each pair of (source sentence, target sentence) is called an example. When a sentence is to be translated, the module matches it against source sentences of all the examples in the corpus to find an equivalent or a similar enough example using the Levenshtein distance (LD) [13]. LD is one of the most popular edit distances. It was invented to measure the variation between two strings, but in this module, the LD is applied on complete words. Figure 3 shows the flowchart of the translation process.

Once the most similar example E_{best} with the minimum LD value is selected, a mean modification value (MMV) is calculated. This value is vital in determining how close the E_{best} to the input sentence S. the MMV is then compared to a threshold value (TH). TH is a percentage value that refers to the maximum number of modifications in the sentence S with regard to its length. If MMV is less than or equal to TH, the target sentence of the example E_{best} is considered the correct translation. Otherwise the module outputs a "No suitable translation found" message.

To relax the constraints imposed by the examples, and make the same example suitable for translating more than a sentence, the examples have been generalized as follows:

1) Each word in the system vocabulary was assigned to a class.
2) Examples were rewritten using classes' names

So to translate the sentence like

"To add a picture, click insert then picture button"

First of all, each word is replaced by the class name it belongs to. So the sentence would be

To <Actions> a <Objects>, click <Tabs> then <Buttons>"

Then, the translation process continues as described before. Once a match is found, the target sentence is processed to decode the classes' names with the original sentence words.

5.2.3 The Signing Avatar Module

The main function of this module is to express the translated ArSL sentences graphically by means of a signing avatar. The module makes use of the JASigning (Java Avatar Signing) software, a synthetic virtual human signing system developed at the University of East Anglia [14].

As the JASigning plays motion data generated from SIGML files, a database of SIGML files has been created. The creation process could be summarized as follow:

- For each sign, specifications of both the manual and nonmanual components were notated and saved to a separate SIGML file with the same name of the sign's gloss.
- Sequences of signs that define the program's technical terms were specified the same way as individual signs
- After creation, the database of signs has been reviewed by the team of ArSL experts, for correction and validation.

- Finally, full ArSL sentences have been constructed from separate SIGML files, to be ready for use by the module.

6. RESULTS AND DISCUSSION

The system starts receiving the uttered Arabic sentence via a connected microphone. Once an end of speech is automatically detected, the Arabic speech recognizer module starts its processes to recognize the uttered sentence. The recognized text is displayed in a textbox. If the uttered sentence was not perfectly recognized, the textbox can be used to enter it as a text. By clicking a translate button, the Arabic machine translator is invoked. The module receives the sentence from the textbox, and starts its mission to translate it. The speed at which the signing avatar plays the signs can be controlled through a speed slide control.

6.1 The Proposed System Evaluation

6.1.1 The Arabic Speech Recognizer

To evaluate the performance of this module, a new test dataset that consists of 50 audio files of 50 different Arabic sentences using all words in the system corpus was recorded. Regarding the Arabic dictionary, it was created using the English phones where each Arabic phone was mapped to its corresponding or most similar in English with the help of a phonetics dictionary. The Word Error Rate (WER) metric was used to evaluate the recognition accuracy. The evaluation gives a WER of 34%. By comparing the hypothesis sentences to those of the reference, we find that the reasons of this high error rate are: 1) Some words were never recognized, these words contain phones that have no equivalents in English. 2) The whole system is based on adaptation of Arabic phones to match English phones.

6.1.2 The Arabic Machine Translator

The objective of evaluating this module is to measure the translation accuracy, and evaluate the performance of the developed translation algorithm in addressing the variations that can occur in the sentences.

A dataset of 150 Arabic sentences that consists of: 1) 50 sentences from the system corpus. 2) The same 50 sentences with only one word changed in each. 3) The same 50 sentences with two words changed in each, was created.

The main task is to compare sequences of signs of the translation output with that of the reference translation, and find to what extent they are similar. The more they match, the better the candidate translation is. The translation quality is evaluated at different threshold values, and using two performance metrics: Sign Error Rate (SER) and BiLingual evaluation understudy (BLEU). Table 1 shows results.

Table 1. Summary of different SER and BLEU gained at different threshold values

Threshold Value	SER %	BLEU
0.125	68.8	0.24
0.2	34.6	0.69
0.25	29.8	0.72
0.3	22	0.81
0.35	3.4	0.98

The best translation results obtained are 0.98 BLEU and SER of 3.4%, which have been found at threshold value TH of 0.35. This means that the Arabic machine translator module will perform well when changing at most one word in a sentence of three words. The main reason for the gained performance is working in a restricted domain.

6.1.3 The Signing Avatar Animator

The evaluation of this module depended on subjective information from Deaf users questionnaires. The process has been carried out with the contribution of 10 Deaf users (7 deaf, 3 hard of hearing), who use the ArSL as their mother language, and with a fair to professional levels of command of Microsoft word 2007. The evaluation was carried out with one Deaf user individually. For each user, a brief description of the system objectives is introduced, the module interface is illustrated, an overall explanation of the questionnaire is provided, and a short time (10-15 minutes) was given to practice the module. Figure 4 shows the module interface used in the evaluation process. Users' questions and comments were collected with help of an ArSL interpreter involved during the evaluation process.

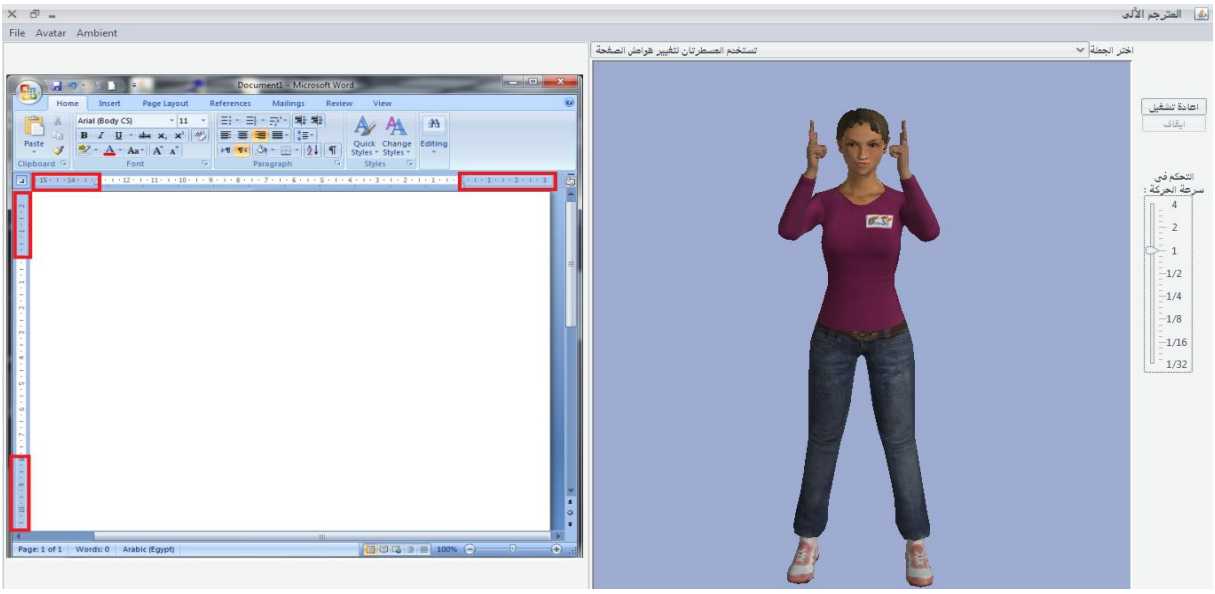

Fig 4: The module interface used in the evaluation process

The negative comments on the signing avatar performance can be summarized in the following points:

- Facial expressions need more efforts to look natural
- The eye movements must be synchronized with and watching the hands movements.
- As the ArSL depends on movements performed by the upper part of the body, seeing a complete standing character is not needed. Instead, the signing avatar pane can be better exploited by showing only the upper part of the signing avatar with a higher zoom, for more clarity of the signs.

7. CONCLUSION

This paper presents an Arabic into ArSL machine translation system. It consists of three modules: An adapted Arabic speech recognizer, an Arabic machine translator, and a signing avatar animator. The system was evaluated using the BLEU and SER performance metrics to evaluate the translation process. The best results obtained are 0.98 BLEU and SER of 3.4%, which have been found at threshold value of 0.35.The graphical representation of the translated ArSL sentences was accepted to a great extent by Deaf. The use of a signing avatar to express the sequences of signs was received successfully by 92% of Deaf students. The negative feedback about the avatar was its naturalness. More efforts are still needed to make signing avatars act and look more real.

8. ACKNOWLEDGMENTS

We introduce all our thanks to the eSIGN consortium and the University of East Anglia for permitting the use of the eSIGN Editor, and the JASigning virtual signing system respectively. Special thanks to John Glauert (at University of East Anglia, School of Computing Sciences), and Thomas Hanke (at University of Hamburg, Institute of German Sign Language and Communication of the Deaf) for their prompt support during the installation of the JASigning. We introduce all our thanks to Asdaa Association for sophisticating of Deaf and Hard of Hearing for their support in ArSL, and participation in assessing the system.

9. REFERENCES

[1] Kennaway, R.,J.Glauert, and I.Zwitserlood(2007). Providing Signed Content on the Internet by Synthesized Animation. In ACM Transactions on Computer-Human Interaction (TOCHI) journal. Volume 14, Issue 3, Article No. 15

[2] Jaballah, K., &Jemni, M. (2013). A Review on 3D Signing Avatars: Benefits, Uses and Challenges. International Journal of Multimedia Data Engineering and Management (IJMDEM), 4(1), 21-45.

[3] Halawani, S. M., Daman, D., Kari, S., & Ahmad, A. R. (2013). An Avatar Based Translation System from Arabic Speech to Arabic Sign Language for Deaf People. International Journal of Computer Science & Network Security,13, 43-52.

[4] Almohimeed, A., Wald, M., & Damper, R. I. (2011, July). Arabic text to Arabic sign language translation system for the deaf and hearing-impaired community. In Proceedings of the Second Workshop on Speech and Language Processing for Assistive Technologies (pp. 101-109). Association for Computational Linguistics.

[5] Mohandes, M. (2006). Automatic translation of Arabic text to Arabic sign language. AIML Journal, 6(4), 15-19.

[6] Caridakis, G., Asteriadis, S., &Karpouzis, K. (2014). Non-manual cues in automatic sign language recognition. Personal and ubiquitous computing, 18(1), 37-46.

[7] Dreuw, P., Stein, D., Deselaers, T., Rybach, D., Zahedi, M., Bungeroth, J., & Ney, H. (2008). Spoken language processing techniques for sign language recognition and translation. Technology and Disability, 20(2), 121-133.

[8] Stokoe, W. C., Casterline, D. C., &Croneberg, C. G. (1976). A dictionary of American Sign Language on linguistic principles.Linstok Press.

[9] Cooper, H. and R. Bowden (2009). Learning Signs from Subtitles: A Weakly Supervised Approach to Sign Language Recognition (2568-2574) Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 22-25 June 2009, Miami, Florida, USA

[10] Efthimiou E, Fotinea SE, Sapountzaki G (2007) Feature-based natural language processing for GSL synthesis. Sign Lang Linguist 10(1):3–23

[11] Kipp, M. Heloir, A. , Nguyen, Q. 2011. Sign Language Avatars: Animation and Comprehensibility.IVA 2011, LNAI 6895, pp. 113-126.

[12] CMUSphinx website: http://cmusphinx.sourceforge.net/. Last access date 20/11/2015

[13] Levenshtein V (1966) Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady

[14] Elliott, R., Bueno, J., Kennaway, R., &Glauert, J. (2010, May). Towards the integrationn of synthetic slannimation with avatars into corpus annotation tools. In 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, Valletta, Malta (p. 29).