

BAR LIS: A Web Tool for Italian Sign Language Synthesis and Interaction

Luca Nardi, Matteo Rubini, Stefano Squartini,
Emanuele Principi and Francesco Piazza

Abstract This work presents the development and testing of BAR LIS (BAR in Italian Sign Language), a web application created in collaboration with the Ancona division of Ente Nazionale Sordi (ENS) and presented during X Masters Awards 2012 event in Senigallia, Italy. BAR LIS was structured as a small dictionary of words and related signs in LIS, Italian Sign Language (presented as 3D rendered animations) in order to ease communication between people attending the X Masters event and hearing impaired personnel of the ENS stall, which included a small coffee shop. A more extensive set of words was later created and tested with ENS members in order to study the impact of resolution of 3D models on comprehensibility and quality of the signs.

1 Introduction

Nowadays, the great development of WEB technologies and the increased capabilities of modern computers, smartphones and tablets, drive to the development of “open” language-to-signed-language systems, not linked to a particular

L. Nardi (✉) · M. Rubini · S. Squartini · E. Principi · F. Piazza
DII—Università Politecnica delle Marche, Ancona, Italy
e-mail: luca.nardi@live.it

M. Rubini
e-mail: rubini.matteo@hotmail.it

S. Squartini
e-mail: s.squartini@univpm.it

E. Principi
e-mail: e.principi@univpm.it

F. Piazza
e-mail: f.piazza@univpm.it

hardware/software configuration or “machine” but accessible to any device capable of connecting to the internet and running a modern browser. In this context, the A3Lab research group of Università Politecnica delle Marche started in 2012 a collaboration with the Ancona division of ENS (Ente Nazionale Sordi). Objective of the agreement was the development of WEB interfaces and systems, as well as more traditional tools, to improve the quality of life of Italian hearing impaired people. In fact, while many international studies have been conducted on this field ([1–3]), only a few regard specifically the LIS. Moreover, excluding one notable exception ([4–6]), they are all quite dated, don’t take into account the last technological developments and share the same “classical” approach, that is the development of a stand-alone, isolated system. The invitation of ENS people to X Masters Awards 2012 (a popular event taking place every year in Senigallia, with the aim to promote local associations operating in the field of social advancement) has given the opportunity to extend and field-test the tools developed during the collaboration between ENS and A3Lab in the context of the so-called A3LIS project. At the date, the A3LIS-147 Database had already been created (a video database of signs comprehending 147 words related to different situations of the common life [7]), and studies had been conducted in the field of automatic recognition of LIS signs and translation from LIS to Italian language [8]; what the A3LIS project still missed was an automatic synthesizer capable to translate Italian to LIS language. Since no prior work had been conducted in this particular field, a first phase of collection of the existing literature was carried out, followed by a study of the available technologies. Then, a new application was developed with these specifications:

1. The application should work “out of the box” on a standard notebook, without too demanding hw/sw prerequisites;
2. The application should work as a “video dictionary”, allowing the selection of simple or composite words and showing the related LIS signs, reproduced by a 3D actor (avatar);
3. It should feature different variations of the same word (i.e. hot/cold milk) re-using the common part of the animation whenever possible;
4. The user should be given the possibility to choose between a male and a female avatar.
5. WebGL technology [9] should be tested together with more “traditional” solutions in order to study pros and cons of each technique.

2 Development of BAR LIS Web Application

2.1 3D Animation

3D animation is based on the same principles of traditional animation. Illusion of movement is produced through very fast reproduction of single frames, which in the case of 3D animation are obtained from digital models. Project A3LIS animations

were created using the MakeHuman anthropomorphic model [10]. In total, 4 models were created (2 male and 2 female avatars), one with the highest possible resolution and skeleton complexity and the other with medium skeleton complexity and low resolution. All the models were then imported in Blender to make animations.

Blender is a very popular open source 3D rendering software. Among its features are the possibility of animating 3d models through different techniques, such as keyframes, animation curves or path-following animation. Animations for the A3LIS Project were created using keyframe animation. Each “keyframe” is obtained repositioning the models (moving the bones of their skeletons) using frames extracted by the original videos as a guide. After having created a sufficient number of keyframes, Blender is capable of obtaining through interpolation algorithms all the other frames needed to complete the animation.

2.2 WebGL and Javascript Picture Rotator

Since the aim of the project was to study the feasibility of WebGL, and not to develop a full-feature 3d model web application, a pre-existing WebGL engine was chosen, in order to minimize the development time. The chosen engine was the Levis WebGL Implementation, proposed by Marco Levis [11]. It was then enhanced by removing un-needed features and optimizing the 3d animation related code, which was too slow for the complex models needed to reproduce LIS signs and didnt take into account the time needed to load models. Then, a simple Javascript Picture Rotator was developed as an alternative to the WebGL Engine. In this case, the input is constituted by pictures created from the original 3d models. Both techniques were developed successfully and are worth considering in the development of a full-feature application. While WebGL is the most interesting one, because it allows to work directly with 3d models (which leads to the possibility to generate, edit and link animations in real time), but it requires much more computational power and the animations occupy more space. Therefore, in contexts with few words and little or no need to link them, such as a dictionary application, the Javascript Picture Rotator would be the better solution, while in contexts with many words which can change often or when linking words together to make whole phrases is important, such a translator application, the WebGL engine is preferable (Fig. 1).

2.3 Bar LIS Web App

BAR LIS was developed as a web application running on a simple USB Web Server. The Javascript Picture Rotator was preferred over the WebGL engine because its hardware requirements are much lower and, being the set of words related to the “coffee shop” context limited, the advantages of elaborating 3d models in real time were not needed. A simple animated web interface was created, using only CSS



Fig. 1 Male and female avatars for the BAR LIS web app

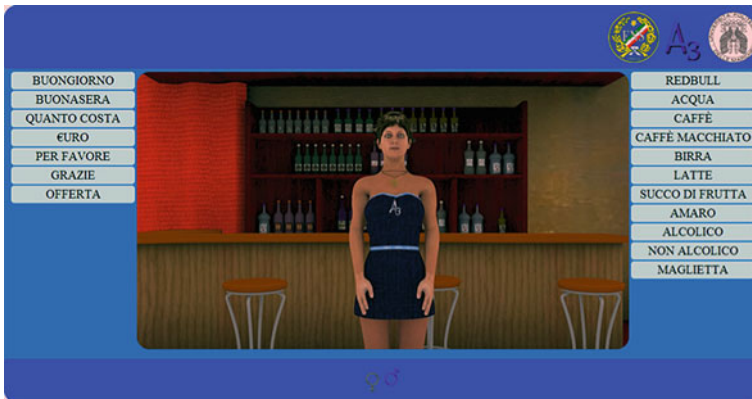


Fig. 2 BAR LIS interface

and HTML to keep the application as compliant to web standards as possible. The structure of the application interface is shown in Fig. 2.

The words and expressions featured in the application are:

- Water (still or sparkling);
- Alcoholic/ alcohol-free drink;

- Bitter (tonic liquor);
- Beer;
- Good morning;
- Good evening;
- Coffee;
- Macchiato (coffee with a drop of milk);
- Milk (cold or hot);
- How much does it cost?;
- Euro;
- T-shirt;
- Please;
- Thank you;
- Offer;
- Redbull;
- Fruit-juice.

Since not all the words needed were part of the A3LIS-147 Database, a new campaign of video acquisition was conducted in collaboration with ENS members.

3 Validation Tests

After having developed a test application such as BAR LIS, it was possible to use the same structure to develop tests in order to evaluate the quality and comprehensibility of animations and the impact factors such as model/picture resolution have on the overall quality of the system.

3.1 System Evaluation

The quantitative evaluation of an automatic LIS signs video dictionary is not a trivial problem, because many factors concur to the overall quality of a “good” system, such as:

- Comprehensibility of signs;
- Complexity of the single words reproduced;
- Quality of sign linking/mixing algorithms and techniques;
- Realism of avatars/digital actors;
- Number of words/expressions featured;
- Ease of use of the system;
- Hardware/software requirements and software optimization;
- Extensibility of the system.

Moreover, the higher or lower importance of each of these factors changes depending on the context in which the system operates. For example, speed and sign linking/mixing would be a major requisite for a real time translator, while for a dictionary application aspects such as photorealism and precision of signs would be much more important. Another aspect to take into account is the not-complete standardisation of the Italian Sign Languages [12]. Each LIS signer expresses him/herself in his/her own way and the same word is never signed in exactly the same way by two different people and the inevitable differences have a similar effect on the quality of communication of speaking English with a person of another country [13].

Three different tests have been conducted to evaluate these different parameters:

1. Comprehensibility/quality of animations;
2. Impact of video resolution and model complexity on the overall quality of the system;
3. Quality of the Blender animation mixing algorithm in producing whole sentences.

All the tests were administered to the ENS members of Ancona. The tests were developed starting from the BAR LIS application and distributed both in the form of a USB stand-alone application and published online. The age and gender of each candidate is registered, and candidates are asked if they contributed to the creation of the A3LIS-147 Database.

3.2 Test 1: Comprehensibility and Quality of Animations

Objective of this test was to evaluate the comprehensibility of animations. The test is composed by 20 questions. In each question, the animation corresponding to a single word is shown; the candidate is asked to recognise the word represented and to express the confidence of his choice with a number from 1 (no confidence) to 5 (absolute confidence).

The test has been administered to a sample of 23 elements extracted from the members of the Ancona division of ENS. 13 of them are male (mean age 47.1), 10 are female (mean age 42.4). Of these, 3 people (2 men and 1 woman) had contributed to the creation of A3LIS-147 Database.

Two features were achieved in the development of this test:

1. The test is a free response test;
2. Each animation corresponds to a single word or concept, there is no context;
3. The test engine verifies that the candidate has played the animation before giving the answer.

Features 1 and 2 ensure that the candidate is given no clues from which deducing the meaning of the sign.

56.74% of signs were recognised correctly, and the mean confidence is 61.96%. In both cases, women had better results than men (58.50% recognition and 62.80% confidence vs 55.38% and 61.31%). Clustering results by age groups, it is possible to

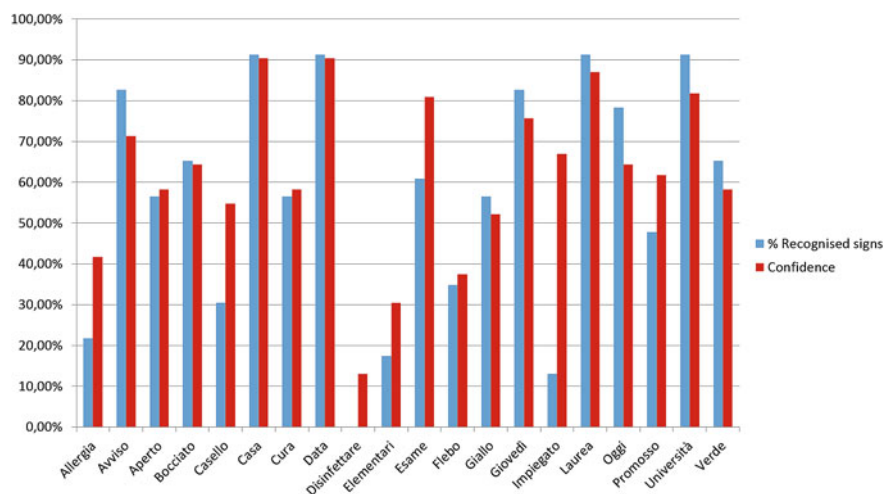


Fig. 3 Test 1 results: % of recognised signs and confidence of given answers

see that percentage of recognition and confidence decrease sensibly with increasing age while unanswered questions increase.

All the results were discussed with the help of an expert of LIS language from ENS. The overall results appear satisfactory, with much words recognised in more than 70% of interviewed subjects, even if in LIS the meaning of a single sign depends on the sentence containing it much more than in the spoken language. Moreover, some signs differ from one another for the facial expression (i.e. “allergy” and “itch”) [14], while all the animation created featured a neutral facial expression. In fact, in more than 80% of wrong answers the word recognised has the same sign of the correct word (Fig. 3).

3.3 Test 2: Impact of Video Resolution and Model Complexity on the Overall Quality of the System

Objective of this test was to evaluate the impact of video resolution, quality of picture rendering and complexity of the 3D model (i.e. number of polygons) on the overall quality of the system. The test is composed by 20 questions. In each question, an original video taken from the A3LIS-147 Database is shown together with two animations created from that video using high quality and low quality settings. The candidate is asked to tell which animation is the most accurate reproduction of the original video; moreover, he/she is asked to evaluate the overall quality of each animation with a mark from 1 (not accurate at all) to 5 (extremely accurate).

The test has been administered to a sample of 14 elements extracted from the members of the Ancona division of ENS. 12 of them are male (mean age 45.3),

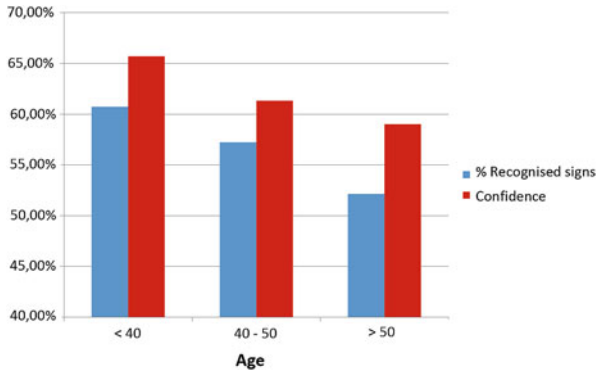


Fig. 4 Results by age

2 are female (mean age 47.5). Of these, 2 men had contributed to the creation of A3LIS-147 Database.

Globally, the high resolution animation is preferred in 50.64 % of cases, achieving an average quality mark of 2,5 out of 5, while the low resolution animation achieves an average mark of 2,3 (Fig. 4).

Even if the high quality animation is preferred over the low quality one, the difference in the evaluation of the two versions is not significant (5 %). This is particularly relevant if we consider that the high resolution model counts 7502 vertices against the 655 vertices of the low resolution model. The difference in number of polygons is not relevant in this particular scenario, in which the final application works with pre-rendered pictures, but it is decisive if the WebGL technique is adopted, because a lower number of vertices means .obj files 5 times smaller (600kB vs 3MB) and therefore an invaluable gain in terms of speed of data transfer and elaboration time. The overall judgement of both the high and low resolution models is not completely satisfactory. If test 1 shows that the comprehensibility of signs is good, this results suggest that photorealism and accurate reproduction of movements are factors to be improved in future development and that some features, such as facial expressions, are very important in the overall quality of signs.

3.4 Test 3: Quality of the Blender Animation Mixing Algorithm in Producing Whole Sentences

Objective of this test was to evaluate the possibility to use Blender animation mixing algorithms to obtain complex animations of whole sentences starting from animations of the single words. Contrary to the first two tests, Test 3 was conducted in a more informal way. The test is composed by 3 animations of these complete sentences:

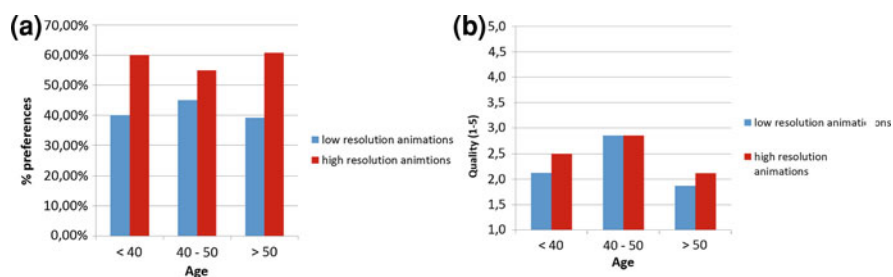


Fig. 5 Test 2 results: high resolution versus low resolution animations



Fig. 6 Blender animation mixing tool

- Where do I have to deliver the modules for the final examinations?
- The Ancona toll booth today is closed.
- How much is the train ticket to Rome?

The sentences were composed using only words already available in the A3LIS-147 Database and the related animations were connected with the Blender animation mixing tools in the correct order. The resulting complete animations were then administered to a selected group of expert LIS signers from the Ancona division to ENS, asking to evaluate the overall quality of the sentence (Fig. 5).

All the people consulted evaluated the overall quality of the sentences as very good to excellent.

The challenge of connecting two or more animations together was already dealt with while making animations of complex expressions for the BAR LIS application (i.e. expressions such as “cold milk” or “how much does it cost”). In that case, each animation was created starting and ending in the same neutral position, so that the animations could be reproduced one after the other without any particular expedient. The resulting complex animation, while perfectly comprehensible in most cases, is however perceived as strange or unnatural. The problem of the “neutral position” could be lessened without resorting to complex animation interpolation techniques by using an “intermediate position”, with arms bent and hands at chest height, so that the distance traversed by the hands while going from the end of a sign to the intermediate position and from the intermediate position to the starting position of the next sign is shorter.

4 Conclusion

We developed BAR LIS, a simple “dictionary-like” web application and field-tested it with good results during the X Masters Awards 2012 event. During the development of such application, we were able to study the different techniques available to achieve reproduction of 3D animation inside an internet browser, WebGL and reproduction of pre-rendered frames. While WebGL is the most interesting technique for future development, because it allows to mix and edit the animations in real time, the reproduction of pre-rendered frames is the most advisable technique when data exchange and computational power are limited resources, such as in mobile applications. We also used BAR LIS as a base for the development of tests in order to evaluate the quality of animations. The results displayed a good overall comprehensibility of the signs, with many words recognised in more than 70 % of cases. The main difficulties were related to different words corresponding to the same sign, words not known at all by the people interviewed and facial expressions, which weren’t taken into account in the making of the animations. Furthermore, use of a more compact, less complicated human model appears feasible, since the high resolution model marks were not significantly better than the low resolution model ones. This suggests that the making of a full WebGL application, at first discarded for the complexity of models and the amount of computation power required to deploy such an application in a real time context, is indeed possible using much simpler models without losing quality. Concerning the mixing of simple animations to obtain whole sentences, complex interpolation algorithms such as those used inside Blender achieve the best results, but it seems possible that comparable results could be obtained in the future by determining the best intermediate position of the single animations.

Acknowledgments The authors express their gratitude to all people at the Ancona Section of Ente Nazionale dei Sordi (ENS) for their support in the development of this work.

References

1. E.A.R. Group: EMBR character animation engine. <http://embots.dfki.de/EMBR/> (2011)
2. Hanke, T.: HamNoSys—Representing Sing Language Data in Language Resources and Language Processing Contexts, University of Hamburg—Institute of German Sign Language and Communication of the Deaf
3. Moemedi, K.A.: Rendering an Avatar from signwriting notation for sign language animation. MSc Thesis, University of the Western Cape—Department of Computer Science (2010)
4. Bertoldi, N., Tiotto, G., Prinetto, P., Piccolo, E., Nunnari, F., Lombardo, V., Mazzei, A., Damiano, R., Lesmo, L., Del Principe, A.: On the creation and the annotation of a large-scale Italian-LIS parallel corpus, In: 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (2010)
5. Lombardo, V., Battagliano, C., Damiano, R., Nunnari, F.: An avatar based interface for the Italian sign language. In: 2011 International Conference on Complex, Intelligent, and Software Intensive Systems (2011)

6. Prinetto, P., Shoaib, U. Tiotto, G.: The Italian sign language sign Bank: using wordnet for sign language corpus creation. In: International Conference on Communications and Information Technology (ICCIT), Aqaba (2011)
7. Fagiani, M., Principi, E., Squartini, S., Piazza, F.: A New Italian Sign Language Database, *Advances in Brain Inspired Cognitive Systems*, vol. 7366. LNCS Springer, Heidelberg (2012)
8. Fagiani, M., Principi, E., Squartini, S., Piazza, F.: A new system for automatic recognition of Italian sign language. In: *Proceedings of WIRN 2012*, Springer LNAI (2012)
9. OpenGL—WebGL standard 2.0. <http://www.khronos.org/webgl/>
10. MakeHuman Project. <http://www.makehuman.org/>
11. Levis, A.P.: <http://pralevis.com>
12. Bertone, C.: *La struttura del Sintagma Determinante nella Lingua dei Segni Italiana (LIS)*, Venezia: MSc Thesis—UniversitàCa' Foscari Venezia—Dipartimento di Scienze del Linguaggio (2006)
13. Sandler, W., Lillo-Martin, D.: *Sign Language and Linguistic Universals*. Cambridge University Press, Cambridge (2006)
14. Rossini, P., Reilly, J., Fabbretti, D., Volterra, V.: Aspetti non manuali nelle narrazioni LIS. In: C. Bagnara, Chiappini, P., M.P. Conte e M. Ott (a cura di), *Viaggio nella città invisibile. Atti del 2 Convegno nazionale sulla Lingua Italiana dei Segni*. Genova, 25–27 settembre 1998. Pisa: Edizioni del Cerro