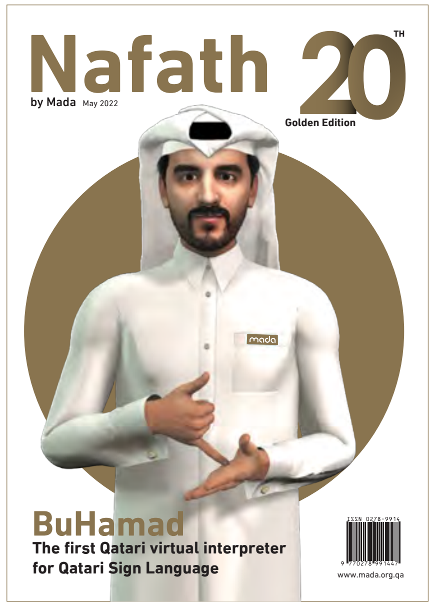
Nafath
by Mada May 2022

20TH Golden Edition

BuHamad

The first Qatari virtual interpreter for Qatari Sign Language

ISSN 0278-9914
9 770278 991447
www.mada.org.qa



Nafath
by Mada

Golden Edition
May 2022

ISSN (online): 2789-9152
ISSN (print): 2789-9144

Contributors
Hugo Jacome Andrade
Achraf Othman
Oussama El Ghoul
Al Dana Al Mohannadi
Shahbaz Ahmed
Shada Bennbaia
Sammy Sedrati

About Mada

Mada Center is a private institution for public benefit, which was founded in 2010 as an initiative that aims at promoting digital inclusion and building a technology-based community that meets the needs of persons with functional limitations (PFLs) – persons with disabilities (PWDs) and the elderly in Qatar. Mada today is the world's Center of Excellence in digital access in Arabic.

Through strategic partnerships, the center works to enable the education, culture and community sectors through ICT to achieve an inclusive community and educational system. The Center achieves its goals by building partners' capabilities and supporting the development and accreditation of digital platforms in accordance with international standards of digital access. Mada raises awareness, provides consulting services and increases the number of assistive technology solutions in Arabic through the Mada Innovation Program to enable equal opportunities for PWDs and the elderly in the digital community.

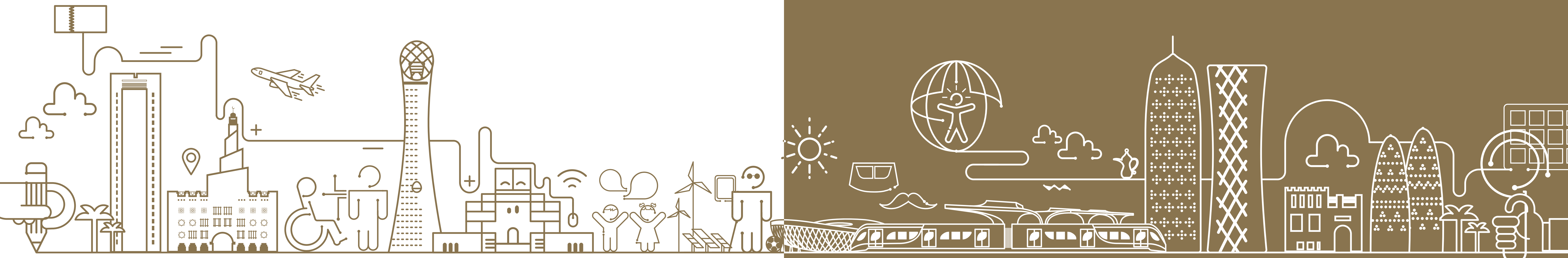
At the national level, Mada Center has achieved a digital accessibility rate of 90% amongst government websites, while Qatar ranks first globally on the Digital Accessibility Rights Evaluation Index (DARE).

Our Vision
Enhancing ICT accessibility in Qatar and beyond.

Our Mission
Unlock the potential of persons with functional limitations (PFLs) – persons with disabilities (PWDs) and the elderly - through enabling ICT accessible capabilities and platforms.

About Nafath

Nafath aims to be a key information resource for disseminating the facts about latest trends and innovation in the field of ICT Accessibility. It is published in English and Arabic languages on a quarterly basis and intends to be a window of information to the world, highlighting the pioneering work done in our field to meet the growing demands of ICT Accessibility and Assistive Technology products and services in Qatar and the Arab region.



Content Page

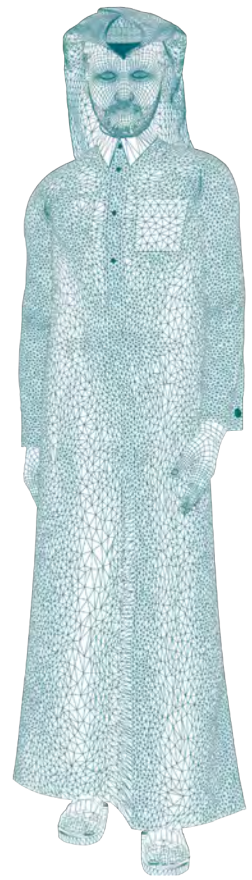
Page 8
BuHamad
The first Qatari virtual interpreter for Qatari Sign Language
Supported by Mada Innovation Program
Achraf Othman, Oussama El Ghoul

Page 15
Fostering Accessibility during the journey at Museums
QR-Code solution to make Museum artefact description accessible in Sign Language
Al Dana Al Mohannadi, Shahbaz Ahmed

Page 20
SpeakLiz by Talov
Toward a Sign Language Recognition mobile application
Supported by Mada Innovation Program
Hugo Jacome Andrade, Shahbaz Ahmed

Page 28
Understanding the challenges in creating Motion capture Sign Language Dataset
A Technical Perspective
Sammy Sedrati, Oussama El-Ghoul

Page 35
Toward an evaluation model for signing avatars
Shada Bennbaia



Page 41
Jumla Sign Language Annotation Tool
An overview
Achraf Othman, Oussama El Ghoul

Open call for papers

Nafath, an open access journal, solicits original research contributions addressing the accessibility, usability and key information resource for disseminating the facts about latest trends and innovation in the field of ICT Accessibility to enable persons with disabilities and the ederly. Nafath is focusing on theoretical, methodological, and empirical research, of both technological nature, that addresses equitable access and active participation of potentially all citizens in the Information Society.

Topics of specific interest

Important aspects and topics to be discussed evolve around (but are not limited to):

- Accessibility guidelines
- Accessible games
- Adaptable and adaptive interfaces
- Alternative and augmented Input /Output techniques
- Applications of assistive technologies in the mainstream
- Architectures, development methods and tools for ICT Accessibility
- Design for All and accessibility education and training
- Evaluation of Accessibility, Usability, and User Experience
- Innovative Assistive applications and environments and ICT Accessibility solutions
- Localization
- Novel designs for the very young, the elderly, and people with different types of disabilities
- Novel interaction techniques, platforms, metaphors, and devices
- Personalization techniques and personalized products and services
- Smart artifacts, smart cities and smart environments
- Web accessibility

In addition to the above, Nafath can host special issues, book reviews and letters to the editor, announcements (e.g. conferences, seminars, presentations, exhibitions, education and curricula, awards, new research programs), and commentaries (e.g. about new policies or legislation).

Why publish with us?

Nafath is registered and indexed by DOI. All issues have an ISSN number for online and print version.

To submit a paper please visit:
https://nafath.mada.org.qa/submit-your-paper/
or send it directly to the editors by email to:
innovation@mada.org.qa

Achraf Othman,
Oussama El Ghoul
Mada Center

BuHamad

The first Qatari virtual interpreter for Qatari Sign Language

As well as all spoken languages, Sign Languages SL has structured grammar and syntax. Although it was visual, multi-dimensional and mainly based on gestures, SL follows specific grammatical rules.

Consequently, the automatic generation of sign language should follow these rules. It is for this reason that Mada works on the development of a new framework that aims to support researchers and developers to create new innovative tools for the deaf. The objective is essentially to create tools that enhance the development of software using grammatically validated sentences. Signed languages are principally based on manual and facial gestures.

For this reason, the automatic generation of SL mainly deals with a 3D virtual signer. The use of avatars presents a mandatory task on SL generation. Since 2019, Mada has been working on designing and developing a personalized Qatari virtual sign language interpreter. The avatar is now published on the Mada website. It is used to translate the website content to Qatari SL. In this paper, we describe the project's objective, and we illustrate how it was designed.

Keywords
3D signing avatar, Sign language generation, Qatari Sign Language

Deaf communities in different countries employ different sign languages, which are not unified [1]. We found sign language (SL) in the United States, Australia, United Kingdom, France, Germany, Spain, Brazil, South Africa, China, Korea, Japan, and other countries. For Arab countries, we found that most of the existing research focused on the use of Arabic Sign Language or Unified Arabic Sign Language [2]. However, sign language differs from one Arab country to another, and Arab deaf communities use different sign languages, Such as Qatari Sign Language (QSL), Saudi Sign Language [3], Tunisia Sign Language [4]. To convey the information between deaf persons, the signer uses multiple channels to construct the Spatio-temporal environment [5], which makes the understanding and processing of sign language a challenge for researchers and opens as an active research field in different domains like linguistics [6], computational linguistics [7], and sociology [8]. From a computational linguistics angle, the computational sign language covers different disciplines:

(1) sign language recognition [9],
(2) construction and building of datasets in sign language,
(3) machine translation from spoken language to sign language [10],
(4) machine translation from sign language to sign language [11],
(5) sign language synthesis using conversation agents or avatars [12],
(6) sign language notation and annotation systems [13], and,
(7) sign language generation [7] and production [14].

Sign language production, translation, and recognition require large-scale datasets and corpora to build accurate systems [15], [16]. However, there are no publicly available large-scale datasets in any sign language with high-quality annotation available for computational sign language processing. This limit impacts the progress of research and decreases the accessibility to information and technologies for the deaf communities and implies a false sense of technological readiness for them.

To address this issue, Mada Center launched the "Jumla Sign Language" research project that aims to improve the accessibility of deaf people by using 3D virtual signers. "Jumla Sign Language" project is composed of 5 elements:

(1) the virtual sign language interpreter "Bu Hamad," a signing avatar available and live on Mada website that translates web content to Qatari sign language in real-time, taking into consideration all components of sign language [17] (Figure 1)
(2) the Mada Machine Translation to Sign Language that aims to provide a real-time translation of text in the Arabic language
(3) the annotation sign language system [18]
(4) an open-source repository containing scripts for body and facial recognition [18]
(5) "Jumla Dataset" a library containing annotated videos.

In this paper, we focus on the first component (The virtual signer "BuHamad").

The Avatar

This work aims to make a realistic Qatari Avatar that can automatically generate Qatari Sign Language. It is known that SL processing is an arduous task as it needs to be based on specific linguistics rules. However, many avatar-based translation systems use a word-by-word translation due to the lack of linguistic studies on Arabic SL [Ref]. Consequently, the lack of SDK and programming tools that help developers to build their applications.

According to many focus groups with Qatari deaf, the avatar should be realistic and produces natural movements and realistic facial expression. The designed avatar has a Qatari style wearing Qatari clothes Gutra and Thube. The creation of "BuHamad" requires solving many technical challenges. The avatar should support live and real-time animation. It should also be run on a web environment with low-speed internet connexion and low graphical resources. The skin is created by applying approximately 116.000 polygons and 71.000 vertices (Fig 2).

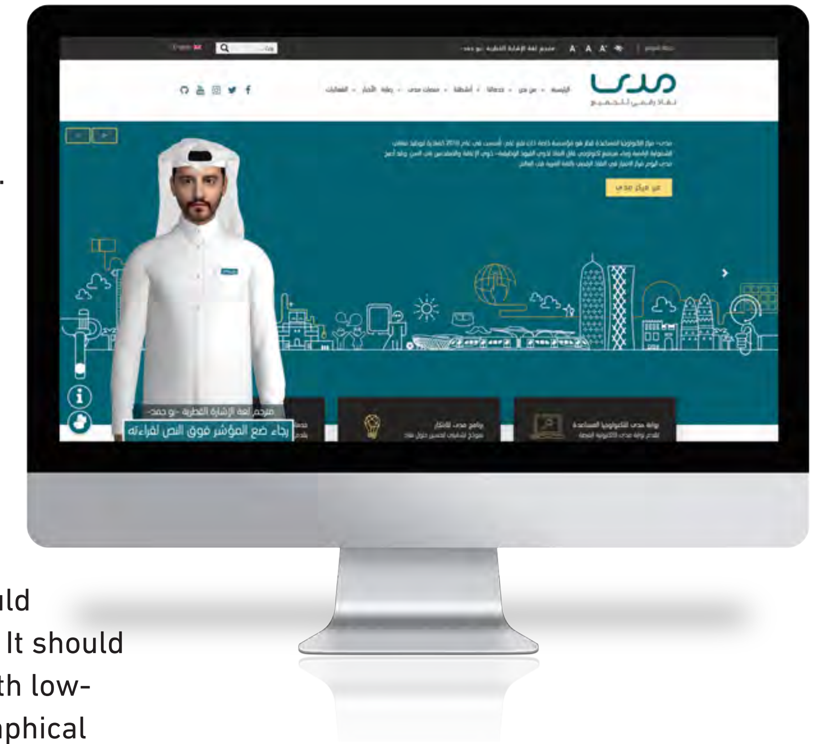


Figure 1.
BuHamad avatar on Mada website

Figure 2.
The design of the skin, beard, and clothes of the avatar

In order to animate the skin, we used a skeleton of 101 bones and applied 38 blend-shape to animate the face (Fig 3). The animation is just a description of changes along a timeline. For our 3D object, we adopted two ways of transforming its triangle mesh to create an animation:

- **Animation through skeletons** attached to the character: Rigid transforms are applied per bone. The skeleton was skinned to the skin of the character. By skinning, we mean the process of attaching the mesh of a 3D model to the virtual skeleton or rig so that when we move a character's rig, the skin moves with the rig, creating a quality character animation. This animation is used to make body and hand gestures.
- **Animation through the morphing of the mesh**: that is, moving each vertex of the mesh separately and storing its new location or by describing its change through particular functions. This technique is resources consuming. For this reason, we used it only to generate face animation.

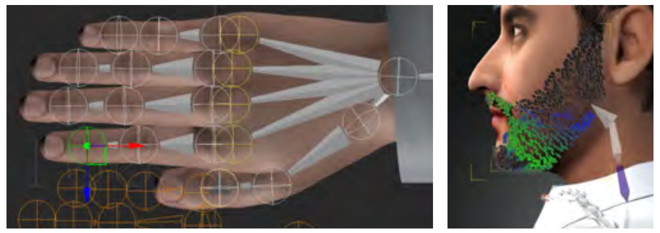


Figure 3.
Skinning the model

We developed a set of tools and services to ensure the character's animation. This framework allows developers to easily integrate the character into their websites and translate texts to sign language.

Framework architecture

Mada SLWebplayer is designed to help developers to add signing avatars into their websites without the need to develop them from scratch. The Service is designed to guarantee high availability (using cloud servers) and extendibility of each component. We adopted 5 tier architecture organized as follows:

- Website Server
  The web server contains the main website which includes the avatar.
- Mada SLK Server
  The Server that provides the sign language toolkit proposed by Mada. The toolkit allows developers to integrate the avatar player into their websites, as described in the next section.
- Mada RealTime SL Streamer
  Once moving the cursor over a paragraph, a dedicated script on SLK automatically generates a request to start receiving sign language animation as a compressed binary stream.
- Signs Database Server
  A server containing the SL animations database (Words + Annotated Sentences)
- Mada SLMT Server
  The Deep Learning Machine translation used to translate Arabic text to ArSL.

Network communication uses a protocol WebSocket. It provides a full-duplex communication channel over a single TCP connection. This protocol was standardized by the IETF as RFC 6455 in 2011, and the WebSocket API in Web IDL is being standardized by the W3C. The SL animation data is sampled and sent at regular time intervals, for which the length depends upon the configuration of the Mada Realtime SL Streamer. Typical sampling rates lie between 20 and 60 Hertz. The update rate of the real-time network stream can be modified separately.

In this paper, we presented a new framework for integrating a 3D virtual signer into websites or desktop applications to make it accessible to the deaf by generating sign language. The proposed avatar is realistic, and it produces natural hand gestures and convincing facial expressions. The avatar is now online published on the website of Mada. The feedback of the deaf is looked promising. The work presents the first step of a big project that aims to improve the implementation of SL into websites.

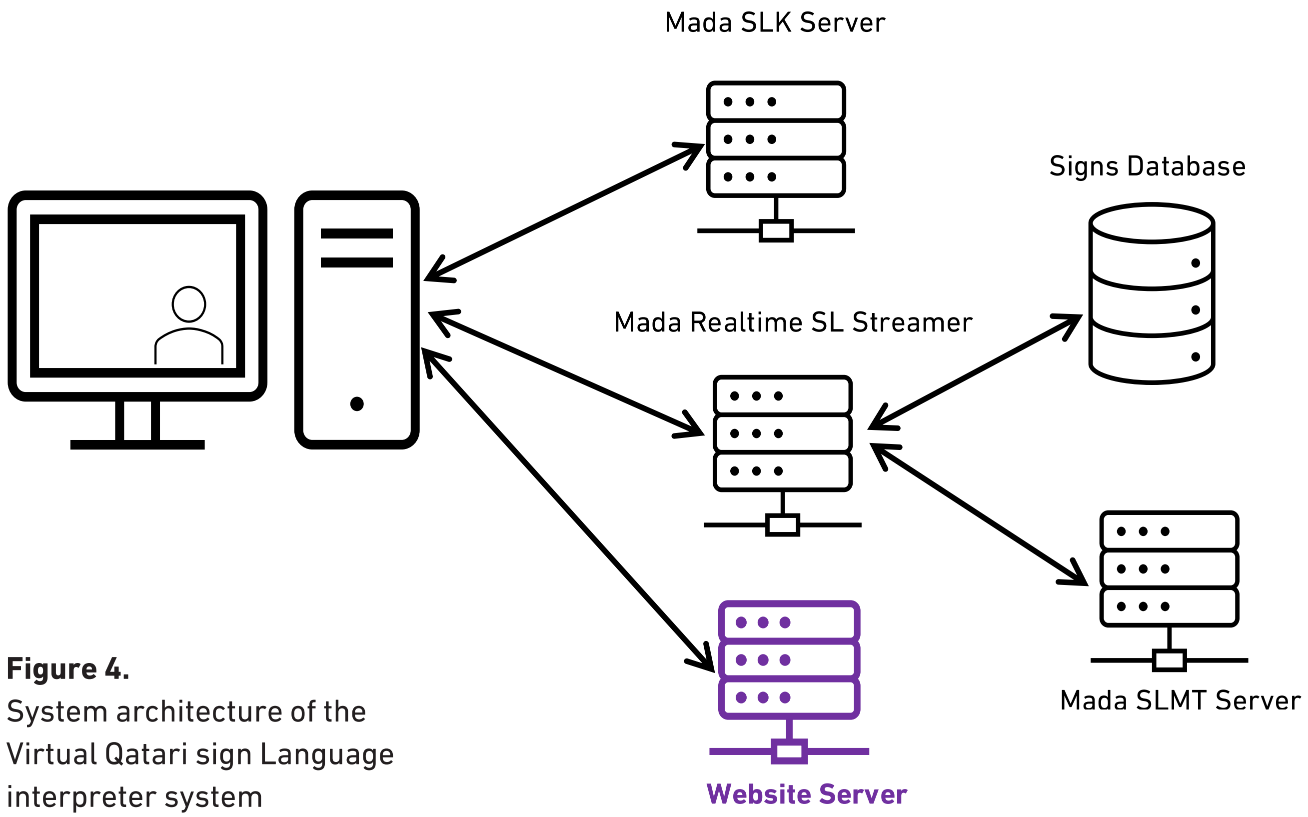


Figure 4.
System architecture of the Virtual Qatari sign Language interpreter system

We at Mada Center highly appreciate and are grateful for the team of the Mada Innovation Program and contributors from the Qatari deaf community, The Qatari Center of Social Cultural for the Deaf and Audio Education Complex. Mocap cleaning tasks and the Jumla sign language project have seen the light thanks to their support and contribution.

References

[1] W. Sandler and D. Lillo-Martin, Sign language and linguistic universals. Cambridge University Press, 2006.

[2] M. A. Abdel-Fattah, “Arabic sign language: a perspective,” J. Deaf Stud. Deaf Educ., vol. 10, no. 2, pp. 212–221, 2005.

[3] Y. O. M. Elhadj, Z. Zemirli, and K. Ayyadi, “Development of a bilingual parallel corpus of Arabic and Saudi Sign Language: Part I,” in Intelligent Informatics, Springer, 2013, pp. 285–295.

[4] M. Jemni, S. Semreen, A. Othman, Z. Tmar, and N. Aouiti, “Toward the creation of an Arab Gloss for arabic Sign Language annotation,” in Fourth International Conference on Information and Communication Technology and Accessibility (ICTA), 2013, pp. 1–5.

[5] T. Shanableh, K. Assaleh, and M. Al-Rousan, “Spatio-temporal feature-extraction techniques for isolated gesture recognition in Arabic sign language,” IEEE Trans. Syst. Man Cybern. Part B Cybern., vol. 37, no. 3, pp. 641–650, 2007.

[6] B. Woll, “The History of Sign Language Linguistics,” Oxf. Handb. Hist. Linguist., 2013.

[7] D. Bragg et al., “Sign language recognition, generation, and translation: An interdisciplinary perspective,” in The 21st international ACM SIGACCESS conference on computers and accessibility, 2019, pp. 16–31.

[8] W. C. Stokoe, “Sociology in Sign Language Studies,” Sign Lang. Stud., vol. 36, no. 1, pp. 227–231, 1982.

[9] R. Rastgoo, K. Kiani, and S. Escalera, “Sign language recognition: A deep survey,” Expert Syst. Appl., vol. 164, p. 113794, 2021.

[10] J. Bungeroth and H. Ney, “Statistical sign language translation,” in Workshop on representation and processing of sign languages, LREC, 2004, vol. 4, pp. 105–108.

[11] A. Othman and M. Jemni, “Statistical sign language machine translation: from English written text to American sign language gloss,” ArXiv Prepr. ArXiv11120168, 2011.

[12] H. Maarif, R. Akmeliawati, and T. S. Gunawan, “Survey on language processing algorithm for sign language synthesizer,” Int. J. Robot. Mechatron., vol. 4, no. 2, pp. 39–48, 2018.

[13] A. Schembri and O. Crasborn, “Issues in creating annotation standards for sign language description,” in sign-lang@ LREC 2010, 2010, pp. 212–216.

[14] R. Rastgoo, K. Kiani, S. Escalera, and M. Sabokrou, “Sign Language Production: A Review,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 3451–3461.

[15] A. Duarte et al., “How2Sign: a large-scale multimodal dataset for continuous American sign language,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 2735–2744.

[16] H. R. V. Joze and O. Koller, “Ms-asl: A large-scale data set and benchmark for understanding American sign language,” ArXiv Prepr. ArXiv181201053, 2018.

[17] “Mada – Digital Access for All.” https://mada.org.qa/ (accessed Nov. 30, 2021).

[18] “Mada Center,” GitHub. https://github.com/madainnovation (accessed Nov. 30, 2021).

Fostering Accessibility during the journey at Museums

QR-Code solution to make Museum artefact description accessible in Sign Language

Al Dana Al Mohannadi,
Shahbaz Ahmed
Mada Center

Museums form the cultural backbone of any city. Ensuring that they are accessible to People with Disabilities means that everyone can participate in the interaction with or even the curation of collections, whether they be displayed in a physical or virtual setting. Using different digital tools, barriers that have been traditionally associated with museum experiences can be easily broken down. Mathaf: Arab Museum of Modern Art – Mathaf and Mada collaborated on creating Sign language Interpreted content and having QR Code for each piece of art that will open a link containing a video interpreting in Sign Language. Mathaf provided the content which was then SL translated & recorded in Mada Studio. Such an initiative will support people with hearing impairment to have an independent museum experience.

Keywords
Sign Language, Deaf, Accessible Museum

Introduction

The concept of digital access focuses on ensuring that people with disabilities have equal access to all grades of life via ICT. This includes eliminating all barriers to the use of ICT products, services, and applications by persons with disabilities that facilitate daily life activities, increase work productivity, facilitate the exchange of information, and improve social life. If these services are not comprehensive enough to be provided to all, including persons with disabilities, they may indeed become tools for isolating certain segments of society from others, thereby preventing digital transformation, and ensuring that everyone contributes to it in a real and tangible way. New media pose new interaction requirements and new considerations in terms of accessibility. The need to address the requirements of a diverse population in the museum in terms of accessibility is extremely challenging. More specifically, people with hearing disabilities are considered among the ones that face barriers to understanding both written and oral information.

Museums can be a hostile place for visitors with a disability, with buildings that are hard to navigate in a wheelchair, and exhibits presented with few concessions to those with sensory or cognitive impairments. With the right solutions, expert recommendations and user feedbacks, accessibility can be improved to meet the growing need for accessible tourism spots in Qatar. Out of its commitment to perform the role as ICT Accessibility strategic enabler, and in line with Mada's efforts to enable equal basis for PwDs and the elderly to take part in cultural life via ICT, Mada collaborated with Mathaf in developing an innovative Sign Language Video Interpretation project to provide full user experience for visitors with hearing impairment. The initiative seeks to allow hearing-impaired people to visit the Mathaf facilities and gain access to all information in the artefacts that is currently in the pilot phase. With such collaboration with Mathaf, Mada aims to enhance the user experience as visitors are guided through the various exhibits in a way that maximizes the learning experience. This makes museums more efficient, and the experience of the visitor more enjoyable.

Mada Use Case

Real-time sign language interpretation using Augmented Reality in Museums

Summary

Seeing that people with hearing disabilities in the Arabic speaking world rely heavily on the use of sign language to communicate, there is a critical need for the development of technology to provide sign language interpretation where human interpreters cannot always be present. As such, a solution is required to provide people with hearing disabilities access to static and rich multimedia content (e.g., video, audio announcements, text, graphics, signage) through Augmented Reality. The solution will enable people with hearing disabilities to access key information and partake in different activities more independently and equitably. This is particularly relevant in Museums, where both the user journey and displays rely heavily on the use of signage and multimedia displays to convey key information.

Target Users

- **Museum visitors with Hearing Disabilities**
- **Sign language users**

User Journey

Maryam is 24-year-old. She is deaf and uses sign language to communicate. Maryam is looking to use her smartphone to access information on static and digital signage in her everyday life.

Issue Statement

As a sign language user, Maryam is unable to read the information found in static or text and rich multimedia-based mediums (e.g., audio, video, graphics, etc.) during her visit to museums. The inability to access crucial museum-related information such as written display descriptions put people with hearing disabilities at a tremendous disadvantage to others.

Sign Language Video Tours

Most, if not all museums provide audio or guided tours to their visitors. Audio tours are provided through an audio device that contains audio descriptions of each exhibit, and guided tours are led by a staff member. Neither of these methods is beneficial to deaf visitors (unless there is another guide who signs in SL – which is unheard of). Exhibits do have text descriptions, but most times there is not much patience in visitors to read all the descriptions.

To provide greater convenience to deaf visitors, and to encourage them to visit museums to enlighten themselves, Mada teamed up with certified sign language interpreters in Qatar, along with Mathaf Arts Team to create video tours that will have audio, Qatari SL, and captions to enhance deaf visitors' experience. Each exhibit will have its video which will be accessed via a QR code displayed on the wall. Scanning the QR code via a phone or tablet will launch a video explaining the exhibit the deaf visitor is at.

Process Implementation

Mada received the narration script generated by the Mathaf curators' team which was then proofread and used for recording from Mada Studio. Qatari Sign Language specialists were hired by Mada to convert the script to signs and were recorded at the Mada Studio. The recorded videos were later matched with the Artifacts and a QR code was generated for each art piece. The first phase of the project targeted the permanent collection at Mathaf. Once finalized, the document was shared with the Mathaf team for final inspection of QR Codes before sending for the production team to print and install QR Codes next to the art pieces. Mada along with the specialist team and end-users visited Mathaf to test all the installed QR Codes and verified the functionality.

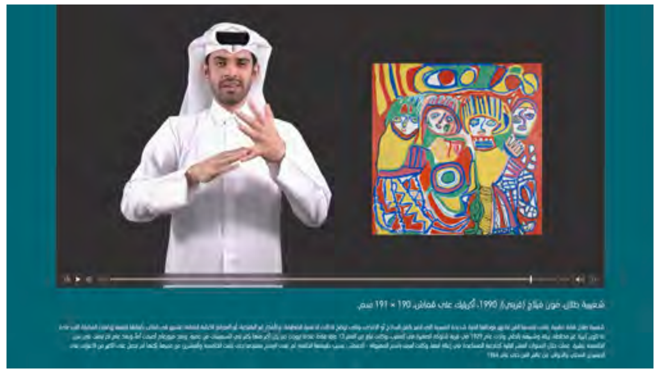


Figure 1.
Sign Language QR Code video

Conclusion

A very common approach is that of the use of pre-recorded videos of human signers. Although this solution generally produces natural results, it is not viable as these videos cannot be easily updated and enriched with new information and thus become obsolete. Re-recording human signers eventually ends up being a tedious, time-consuming task and is also an expensive process, whereas "stitching" together clips of signs to produce words and sentences often leads to non-natural results due to discontinuities that certainly appear in the video editing. As virtual signing avatars are developed, this matter can be rectified in the near future.

SpeakLiz by Talov

Toward a Sign Language Recognition mobile application

Supported by Mada Innovation Program

Hugo Jacome Andrade
Talov

Shahbaz Ahmed
Mada Center

The present article discusses Sign language recognition which is part of one of the most challenging Artificial Intelligence (AI) algorithms: camera-based gesture recognition. It is a relatively novel practical application in contrast with its old and theoretical presence in the general AI field. Probably due to the required computing power for a device to track complex gestures (like sign languages) just with a camera along with non-ideal light, focus, and stability conditions or even many of them combined. Nevertheless, reaching acceptable results on mobile devices with this recognition task is even more elusive than doing it on computers (actually, most research projects in this field are computer-oriented).

Keywords
Sign Language Recognition

Everything in AI starts with the dataset (image, text, sound, or document). Regardless of the many advances in AI reliability with less and fewer amounts of data, it is still a utopia to find an AI that can even get close to the typical human ability of one-shot learning. So, that implies that the datasets for feeding this gesture recognition tool must consider a large amount of data with many variables (slight differences in the performance of the gesture from one person to another, speed) [1]. Also, sign languages are too complex in their diversity across different countries and regions [2]. The information contained in facial gestures tiny finger movements; so, a very detailed curating process within a dataset becomes essential to have the best possible input for the later AI structures that will process that data.

Understanding Sign Language Recognition

To begin with, understanding sign language recognition from the tech side requires a deep comprehension of some computing and AI trends [3]. It becomes a complex task on mobile devices due to the required time local processing and the Internet connection with a server that processes the data and sends labels or similar outputs back to the mobile) [4]. In that case, the level of computational resources optimization must be huge to have a helpful tool considering the best possible user experience (UX). Otherwise, developers can simply be getting closer to something able to work although at the same time, burning the mobile's system-on-a-chip and draining its battery in a matter of seconds, resulting in becoming useless for real-life applications.

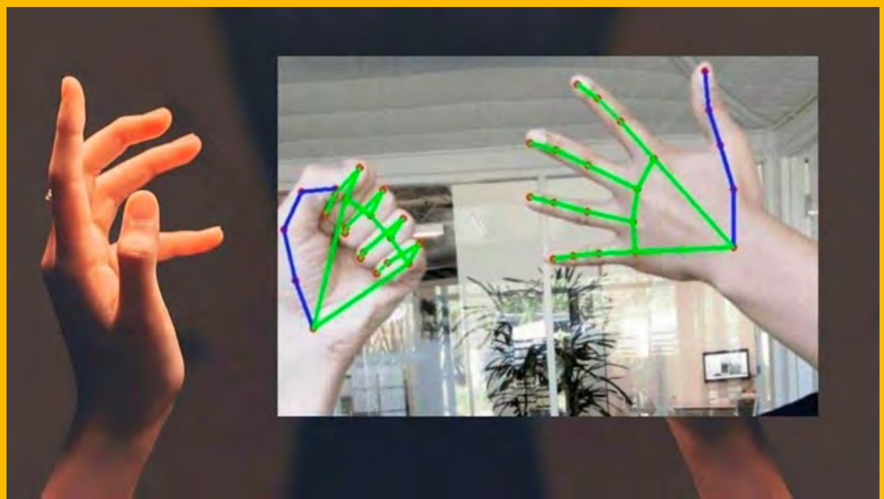


Figure 1.
Sign Language Recognition Algo – (Kumar, A, 2016)

Artificial Intelligence is a vast umbrella discipline that involves all the aspects of giving machines the capacity to be smart [5]. Here is where convolutional neural networks (CNN) conform to one the most common ways for a machine to learn because they are intended to perform some tasks emulating how the human brain's neural networks identify patterns and classify information. When it comes to Sign Language Recognition, there are plenty of techniques with different approaches. However, some of the most common are Object detection and Video analysis, depending on the kind of data to analyze (stationary gestures or those including complex movements).

Therefore, it is essential to consider which kind of CNN could best perform when the goal is to analyze or track data from human movement. It can be an excellent choice to look for those CNN designed to deal with data that may vary in some way over time. There could be many options like extended short-term memory networks (LSTM) [6] spiking convolutional recurrent neural networks (SCRNN),

among others. It is essential to notice that the neural network choosing decision is dependent again on the type of dataset is being used within the project. It is imperative to analyze several aspects from the beginning of the project so that researchers and developers can avoid many misconceptions and cause time delays or unnecessary resources spending.

Nonetheless, the other crucial part of this journey is to provide the right and specific UX for users with hearing impairment, which includes working on the tool's reliability not only from the tech side but also from its usefulness side. It is commonly known that many AI tools provide poor outputs in terms of syntax complexity, unable to go beyond just individual words. Applying correct natural language processing (NLP) becomes imperative to close the syntax complexity gaps between the two sides. The grammar structures of sign languages could be different from spoken languages. That is why the NLP is a vital part of the puzzle even in the double way (not only considering the sign language recognition but also the sign language generation with avatars or any similar option) [7]. This kind of technology intends to democratize and make it cheaper to access the possibility of having at least a digital assistant for better understanding the surroundings in daily life tasks.

Not everyone can afford sign language interpretation services for more than a couple of hours in the best of cases. Even considering the case of unlimited resources to pay for these services permanently, there is still a little big problem: privacy [8]. There are so many moments where a sign language user and a person who does not know sign language would love to communicate just between them without the presence of nobody else; that moment is when this kind of accessible technology becomes so necessary. The rest of the events, like conferences, TV broadcastings... It will continue requiring human sign language interpretation services.

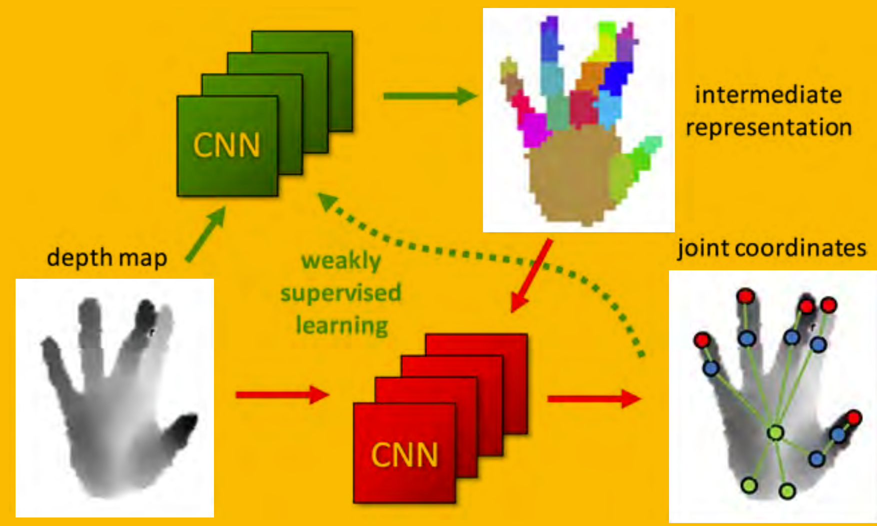


Figure 2.
Convolutional neural networks (CNN) Processing – (Maro, J, 2020)

SpeakLiz by Talov

Supported by Mada Innovation Program

Briefly, all this sign language recognition approach sounds challenging, and that cannot be denied: is highly challenging, but at the same time, the positive reaction in the users while testing this tool is a huge motivation to continue developing the mission of the company Talov (co-founded by Carlos Obando and Hugo Jacome), bringing the world closer to everyone in a more accessible way while using cutting edge technology running just in the pocket and in the hand of every user [9]. This kind of AI to build a mobile sign language recognition tool is tough because building this type of technology requires creating many things from scratch, even from the datasets. The goal is to update SpeakLiz (designed for people with hearing impairment) with better features, especially for the camera-based sign language recognition.

Figure 3.
SpeakLiz by Talov

The SpeakLiz app by Talov has features like the Sign Language Recognition feature, which transforms classic sign language into a modern form of communication, by converting it into voice and text in real-time. The Surrounding Sound feature keeps the users aware of their surroundings; it detects acoustic parameters to identify specific sounds around them. These include emergency vehicles, alarms, ring bells, animal sounds, female and male voices, kids' voices, even some types of music. Once such sounds are detected, the users are alerted by vibration and visual signals. The Listening feature allows deaf users to "read" what other people are saying. SpeakLiz understands up to 48 languages (Android) and 35 (iOS) in a human voice. When a person speaks to a SpeakLiz user, their sound is converted into the text format in real-time. The user can read what is being said and respond accordingly. The deaf user can now perform daily activities without any impairment. Speaking is the opposite process of listening. Using this feature, deaf user can type their thoughts. The text can then be converted into speech with the press of a button (emoticons included).

Mada partners with Seedstars to promote the ICT Accessibility Solutions with the ultimate goal of supporting Accessibility startups in Qatar and beyond to benefit and improve the lives of PWD. Seedstars and Mada share a vision to support innovation by working with the best startups in the field and offering them suitable funding and subject matter expertise [10].
In 2020, Talov won the Digital ICT Accessibility Awards by Mada- Seedstars. Recently, Talov opened

free access to its SpeakLiz and Vision technology for the hearing and visually impaired in Ecuador. Also, the team won the first prize in the Ecuadorian largest Entrepreneurship contest, was selected as one of the world's Top 200 most innovative startups at Web Summit thanks to the SpeakLiz app.

One more thing, Talov has demonstrated that many of these technologies, including gesture recognition, works at accessibility, which probably can be treated as one of the most complex scenarios for the AI algorithms inference due to the large number of factors to be considered. Consequently, it is time to take advantage of this tool to apply it beyond accessibility. Many other industries can benefit from something like this. Indeed, there are a lot of other industries that are not even aware of how gesture recognition and other technologies can positively impact their operations.

A common approach to this sign language recognition field often is the usage of remote/online servers or instances with significant power of processing. However, the challenge is that even with a lightning-fast internet connection, there is always latency. For a real user, this factor can be the difference between experience or not an incident. Talov is doing further research to match the current real-time performance with the possibility of learning from each user experience new patterns layers above the central model. It probably could require using some cloud instances for specific on-demand model updates and then deploying all those new CNN weights again to a local environment. It can continue offering an AI inference real-time performance independent from remotely.

References

[1] S. Ameur, A. B. Khalifa, and M. S. Bouhlel, "A novel hybrid bidirectional unidirectional LSTM network for dynamic hand gesture recognition with leap motion," Entertain. Comput., vol. 35, p. 100373, 2020.

[2] A. Othman and M. Jemni, "Designing high accuracy statistical machine translation for sign language using parallel corpus: case study English and American Sign Language," J. Inf. Technol. Res. JITR, vol. 12, no. 2, pp. 134–158, 2019.

[3] A. Lahiri, A. Othman, D. A. Al-Thani, and A. Al-Tamimi, "Mada Accessibility and Assistive Technology Glossary: A Digital Resource of Specialized Terms," in ICCHP, 2020, p. 207.

[4] M. Jemni, K. Khribi, A. Othman, O. Elghoul, and K. Jaballah, "AlecsoApps: Toward empowering mobile applications development in the Arab world," in State-of-the-art and future directions of smart learning, Springer, 2016, pp. 87–93.

[5] M. J. Garbade, "Clearing the confusion: Ai vs machine learning vs deep learning differences," Data Sci., vol. 14, 2018.

[6] H. Cooper, B. Holt, and R. Bowden, "Sign language recognition," in Visual analysis of humans, Springer, 2011, pp. 539–562.

[7] A. Amir et al., "A low power, fully event-based gesture recognition system," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 7243–7252.

[8] L. Zhang, G. Zhu, L. Mei, P. Shen, S. A. A. Shah, and M. Bennamoun, "Attention in convolutional LSTM for gesture recognition," Adv. Neural Inf. Process. Syst., vol. 31, 2018.

[9] L. Pigou, S. Dieleman, P.-J. Kindermans, and B. Schrauwen, "Sign language recognition using convolutional neural networks," in European Conference on Computer Vision, 2014, pp. 572–578.

[10] D. Al Thani, A. Al Tamimi, A. Othman, A. Habib, A. Lahiri, and S. Ahmed, "Mada Innovation Program: A Go-to-Market ecosystem for Arabic Accessibility Solutions," in 2019 7th International conference on ICT & Accessibility (ICTA), 2019, pp. 1–3.

Understanding the challenges in creating Motion capture Sign Language Dataset

A Technical Perspective

Sammy Sedrati, Oussama El-Ghoul
Mada Center

Sign language (SL) is a visual language with a syntactic system different from spoken language. Deaf and hard-of-hearing individuals use it as a primary communication channel. The lack of translated information from Arabic text to sign language represents a barrier. It creates a gap between the deaf and hearing communities, which is even wider within deaf communities with low literacy skills. To fill this gap, in 2021, as part of the Jumla Sign language project, Mada Center developed and launched "Bu Hamad," the first virtual and digital character (Avatar) using the Qatari sign language to serve this community and help translate Arabic texts into sign language. For the digital performance and to bring the avatar to life, a motion capture (mocap) system has been used to record the performance of professional signers translating common and mostly used Arabic words and sentences. These recordings were then converted into digital animation data to be further quantified and optimized as a training dataset for Machine learning (ML). The usage of mocap is currently the most effective way to achieve a broader acceptance amongst the deaf and hard-of-hearing individuals by synthesizing the natural motion of an actual signer. This article summarizes the process and the technical challenges faced in bringing the raw data from the mocap system to the avatar in a refined set of motions to build the dataset.

Keywords
Sign Language, Avatar, Motion capture, Jumla, Machine Learning, Artificial Intelligence, Dataset

Mocap is digitally recording the motion of objects or persons [1]. In this latter case, the recording is done with the help of sensors that are placed inside a specially made suit ("Inertial Mocap") or by attaching markers on a person's body and recording his motion with unique cameras ("Optical Mocap"). There are other mocap technics where only specialized cameras with depth sensors capture an actor's performance. Mocap is more suited to record subtle movements of hands, fingers, and arms for SL rather than the traditional keyframing animation that proved to be very time-consuming and practically impossible for the creation of a sign language database. Despite the satisfactory visual results it may achieve, Mocap often requires editing and cleaning before it can be used. Mada Center has used this technic to animate "BuHamad," an avatar automatically generating Qatari sign language. This task is part of the Jumla Sign Language research project supported by the Mada Innovation Program [2].

Understanding the cleaning process

A signer, wearing a mocap suit, translates in SL words/sentences recorded by a computer equipped with mocap's native software. Every motion performed by the signer will be reflected on a dummy skeleton and saved in a format readable by editing software for treatment and clean-up. The cleaning consists of optimizing raw data, such as reducing keyframes, adjusting the placement of body parts, mainly hand and arms, and removing noise.

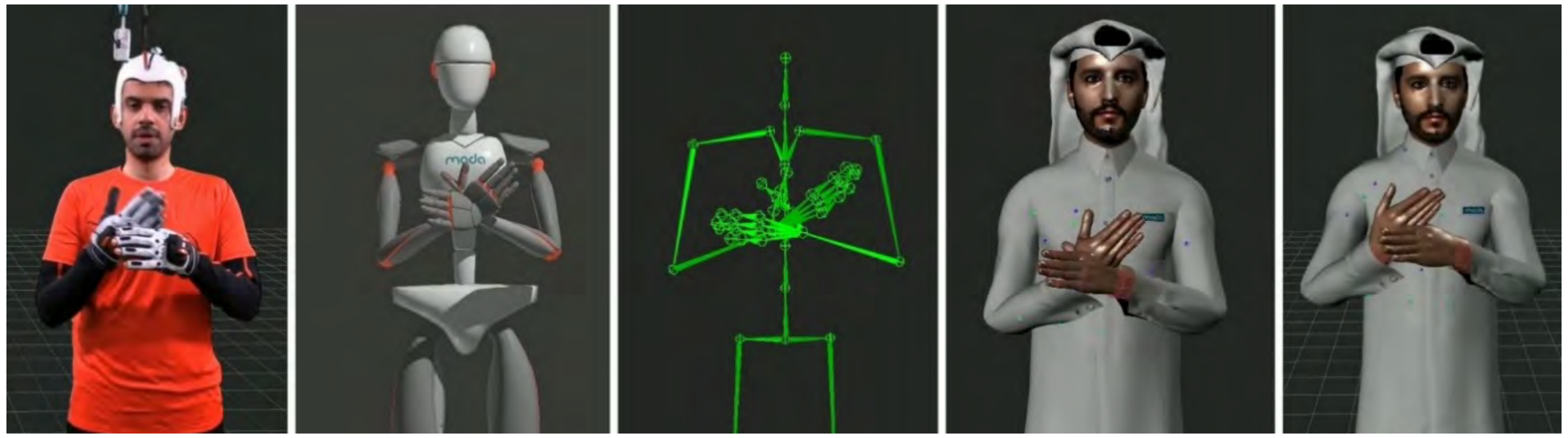


Figure 1.
Cleaning process from Left to right: Signer with mocap suite – real-time capture on native software – imported data in editing software – matching motion with the avatar – cleaned motion on avatar

Motion capture cleaning phase

Some of the most recurring tasks in cleaning are related to the zeroed keyframes (recorded motions in initial/resting state). It is due to a brief signal loss during the mocap recording session. Although some built-in filters in editing software may automatically fix this, there are instances where those filters may delete some needed motions. In this case, manually fixing those keyframe animations is the solution. Another recurring issue is when one or more animated rotations are offset, where in most cases a long arc is drawn instead of a short one. It is mainly due to the way the capturing software and its algorithm is dealing with the rotation of a joint while inheriting the motion of its parent joint (example: base finger joint inheriting wrist motion).

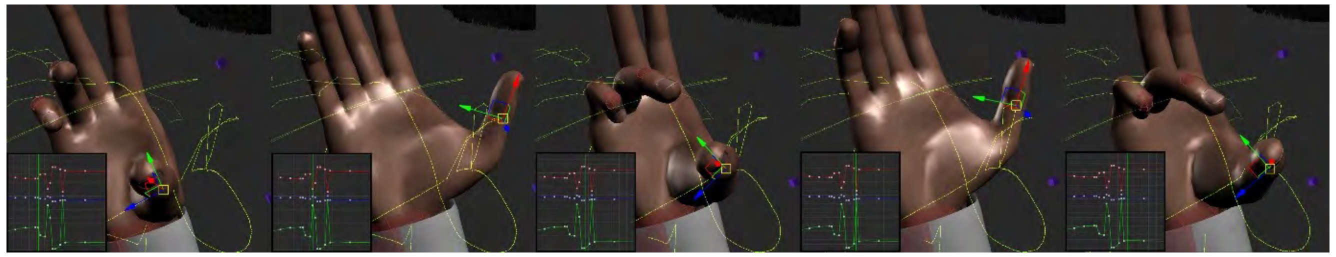


Figure 2.
Zeroed thumb's base (Metacarpal Joint)

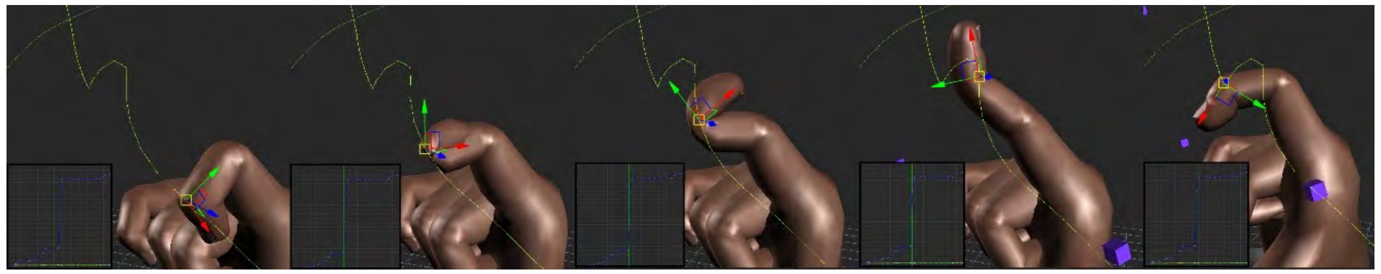


Figure 3.
Interphalangeal Joint inheriting motion from the wrist's accelerating movement

Methodology and approach

Mocap editing falls under 3D animation, but unlike the traditional 3D animation that everyone knows, mocap does not follow all the core principles of traditional animation. These core principles were developed by Disney's animators in the 1930s and are still followed nowadays by big-name animation studios [3]. "Straight ahead and pose to pose" is one of those shared principles between mocap editing and traditional animation with an emphasis to "pose to pose" editing [4]. For" Bu Hamad" avatar motion cleaning, two methods were used to achieve the final goal:

- **Manual editing (Pose to pose)** is creating a series of bookmarks in the timeline by storing positions of the avatar and performing cleaning on those bookmarks. The original keyframes from the mocap either procedurally drive motions between bookmarks. This approach has the advantage of storing those cleaned bookmarks (poses) for future use when facing similar avatar posture in different words or sentences.
- **Automatic editing (Scripts and plugins)** There is a fair amount of data to be treated some of the repetitive tasks like motion trimming and clipping, fixing initial joints' rotations. These tasks must be automated by scripts and plugins; as such their development and regular updates and maintenance are critical to cutting production time.

Editing and mocap cleaning are time-consuming and laborious tasks. While existing motion editing methods accomplish modest changes, more extensive edits and intricate cleaning can require the artist to "re-animate" parts of the motion manually to achieve the best visual representation of the original motion. However, with the manual and automatic approach introduced in this article production time could be sped up considerably. With these methods, there are opportunities for further study in enhancing the cleaning process by blending between these two approaches and introducing new cleaning methods with the usage of artificial intelligence for avatar posture recognition and pose to pose automatic fixtures.

We at Mada highly appreciate and are grateful for the team of the Mada Innovation Program and contributors from the Qatari deaf community, The Qatari Center of Social Cultural for the Deaf and Audio Education Complex. Mocap cleaning tasks and the Jumla sign language project as a whole have seen the light thanks to their support and contribution.

References

[1] T. H. Ribeiro and M. L. H. Vieira, "Motion Capture Technology—Benefits and Challenges," Int J Innov Res Technol Sci Int J Innov Res Technol Sci, vol. 48, no. 1, pp. 2321–1156, 2016.

[2] D. Al Thani, A. Al Tamimi, A. Othman, A. Habib, A. Lahiri, and S. Ahmed, "Mada Innovation Program: A Go-to-Market ecosystem for Arabic Accessibility Solutions," in 2019 7th International conference on ICT & Accessibility (ICTA), 2019, pp. 1–3.

[3] F. Thomas, O. Johnston, and F. Thomas, The illusion of life: Disney animation. Hyperion New York, 1995.

[4] T. White, Animation from pencils to pixels: Classical techniques for digital animators. Routledge, 2012.

Toward an evaluation model for signing avatars

Signing avatars can make a significant impact on the lives of deaf people by making information accessible anytime and anywhere. With technological development, sign language avatars can be the cost-effective communication solution that will remove the barriers between deaf people and the world. However, most researchers are not part of the deaf community. To this day, the deaf community has little to no knowledge about the signing avatar technology. Thus, researchers have created and used evaluation methods to involve deaf people and feedback to develop and improve sign language avatars based on their needs and requirements. In the article, evaluation methods and tools used to assess signing avatars' functionality, acceptability, and shortcomings were presented and discussed.

Shada Bennbaia
Mada Center

Keywords
Sign Language Processing, Evaluation Methodology, Signing Avatar

Signing avatars are virtual signers that produce sign language to improve the communication and information accessibility for a deaf individual. They are not intended to replace human sign language interpreters but to co-exist and support them in different domains. For instance, interpreters are required when the sign language translation is highly sensitive and must be as accurate as possible, such as in doctor appointments. At the same time, the avatar can serve the purpose of translating standardized text or automatically translating dynamic content, such as announcements and websites [1]. To read and write without auditory cues is a challenging task for a deaf person which causes many of them to leave school low level of reading/writing abilities. It limits their access to all written content. One of the early adopted solutions was the use of recorded human signers. However, this solution requires high production costs, and the recorded videos cannot be modified or even anonymized after production. As opposed to using virtual characters to produce sign language where the appearance can be completely customized, the animations can be dynamic and easily adjusted when needed, can have an interactive behavior, and the production of new content is relatively cost-effective [2].

Any new technology often faces the challenge of acceptability within the targeted users. Moreover, most of the developers of signing avatars are hearing researchers, which can make deaf people skeptical about the technology for historical reasons [2]. Accepting the sign language avatar within the deaf community is crucial for the successful implementation of the technology. Previous works tried to involve deaf people in developing and evaluating signing avatars but an efficient effort to clarify the overall acceptance is not made yet. It is essential to identify the issues that cause the rejection of the technology to solve them. Researchers used different approaches to assess the needs of the deaf individuals, incorporate them in the development process and use their feedback to evaluate the signing avatar's performance. Comprehensibility of signing avatars is a significant factor that affects acceptance and performance. Assessing comprehensibility is a challenging and not straightforward task. No unified methodology exists to test sign language avatar comprehensibility [3]. In the ViSiCAST/eSIGN projects, comprehensibility tests have been carried out. One limitation is that these tests primarily used few participants [4].

In the following sections, focus groups, as one of the most known methods to collect qualitative data, are presented with examples followed by performance metrics used to quantify sign language avatar performance. This task is part of the Jumla Sign Language project supported by the Mada Innovation Program [5].

Focus groups

Focus groups are well-known tools used to evaluate human-computer interaction and extract empirical data in research and analysis. The aim of focus groups in evaluating sign language avatars is to elicit deaf people's opinions and feedback toward the signing avatar in the research and development phase to improve the technology and better address the users' needs. Through interactions and comments, this method enables researchers and developers to get in-depth information from the focus group participants about their preferences and how to prioritize them and what issues they consider important. Usually, the focus group is controlled discussions led by an expert with 3-10 potential end-users, where the expert presents the product and guides the discussion.

Balch et al. [6] were the first to use the focus group method with deaf people where they found that this method is very productive in terms of extracting information and understanding the needs and requirements of participants. Kipp et al. used focus group methods to evaluate the acceptability and comprehensibility of sign language avatars [2]. They provided a few recommendations to ensure the quality of the focus group sessions: a) use visual materials such as images, icons, and videos; b) ensure the environment is sign language friendly; C) utilize interactive tools such as voting and open discussions, and the focus group session with online assessments. Kipp et al. [2] conducted a focus group study, complemented by an online questionnaire, to evaluate the German deaf community's opinion about sign language avatars.

Two focus groups with 5 and 3 deaf participants. Different sign language avatars were presented to the focus group, and they were asked to rate them and give their feedback on the signing avatar's performance. The deaf participants mainly criticized the appearance and the animation of the signing avatars and described it as unnatural, robotic, and emotionless. These criticisms showed that the nonmanual features of the SL avatars such as the mouth patterns, face expressions, and natural body movements, are essential for deaf people to accept the signing avatar technology. To quantify the focus group results, the deaf participants were asked to vote on the most critical avatar feature. Facial expressions got the most votes, followed by natural body movement, emotions, appearance, and comprehensibility.

Figure 1.
Focus group from the Deaf community

Researchers around the world have used the focus group method to test signing avatars that produce different signing languages such as American Sign Language, Swiss-German Sign Language, Japanese Sign Language, Brazilian Sign Language, Turkish Sign Language, British Sign Language, and many more [7]–[14]. Evaluating signing avatars through a focus group from the deaf community proved a necessary step to incorporate the end-user in the development process and increase their understanding and acceptability of the technology. Moreover, it is also an essential step for researchers to understand the end-user since most of the researchers are not from the deaf community and provide them with a communication tool that satisfies their needs and requirements.

Performance metrics

—

Even though focus groups and questionnaires provide a good evaluation of the animated avatars, the development process requires detailed, quantified performance metrics to measure and benchmark the development and evolution of these signing avatars. One method to quantify focus group study is to let the participants objectively measure the accuracy of the sign language translation. For example, with one of the first known sign language avatars, TESSA [7]–[14], the group was asked to indicate whether the sign language phrase produced by the avatar is accurate and easy to understand and, if not, what can be the reason. Through this simple method, they were able to identify the phrases identification average accuracy to be 61%, where 70% of the identification errors were identified to be due to unclear signs.

Figure 2.
Focus group from the Deaf community

Using machine translation methods to translate and produce sign language requires performance metrics to precisely evaluate the translation process. San-Segundo et al. [15] used the scoring metrics of BiLingual Evaluation Understudy (BLEU), which calculates the statistical difference between the machine translation and original translation, and Sign Error Rate (SER), which is the percentage of wrong produces signs. However, in this study, the SER was reported to be 31.6% and the BLEU to be 0.5780. Similarly, Patel et al. [16] evaluated the performance of their machine translation avatar system by statistically measuring the accuracy of voice recognition, grammar search and sign synthesis which in total achieved an average translation accuracy of about 77%. Moreover, they used the processing time to benchmark their work faster at publishing the study at 0.85 seconds. In another work by Oh et al. [17], they used the word translation, the correctly machine-translated word ratio, to evaluate their signing avatar for a weather forecast system.

Furthermore, System Usability Scale (SUS) has been used to evaluate the usability of the signing avatar systems by the deaf user. El-Gayyar et al. [18] used this performance metric with a focus group of 5 deaf people to evaluate an Arabic signing avatar application within a limited domain. They achieved a 79.8% average SUS score, indicating that the developed application is acceptable.

It is important to note that to evaluate a sign language avatar system, comprehensive testing should be conducted considering all sign language production characteristics such as accurate translation, non-manual signals, spatial representation, and avatar appearance and naturality, using both qualitative studies and performance metrics.

Producing sign language through animated avatars is a challenging task due to the complex nature of sign language. Evaluating signing avatar systems is not only essential to measure the development progress but also to increase the engagement and acceptability of the deaf community as the signing avatar is a solution that can make their daily lives more accessible and more independent. For that reason, complimentary evaluation methods should be combined to efficiently test and evaluate the sign language avatar's performance, comprehensibility, and acceptability.

—

References

[1] S. Ebling, "Evaluating a Swiss German Sign Language Avatar among the Deaf Community," Proc. Third Int. Symp. Sign Lang. Transl. Avatar Technol., no. October, 2013.

[2] M. Kipp, Q. Nguyen, A. Heloir, and S. Matthes, "Assessing the deaf user perspective on sign language avatars," in ASSETS'11: Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility, 2011, pp. 107–114. doi: 10.1145/2049536.2049557.

[3] M. Huenerfauth, L. Zhao, E. Gu, and J. Allbeck, "Evaluation of American sign language generation by native ASL signers," in ACM Transactions on Accessible Computing, 2008, vol. 1, no. 1. doi: 10.1145/1361203.1361206.

[4] J. R. Kennaway, J. R. W. Glauert, and I. Zwitserlood, "Providing signed content on the Internet by synthesized animation," ACM Trans. Comput.-Hum. Interact., vol. 14, no. 3, 2007, doi: 10.1145/1279700.1279705.

[5] D. Al Thani, A. Al Tamimi, A. Othman, A. Habib, A. Lahiri, and S. Ahmed, "Mada Innovation Program: A Go-to-Market ecosystem for Arabic Accessibility Solutions," in 2019 7th International conference on ICT & Accessibility (ICTA), 2019, pp. 1–3.

[6] G. I. Balch and D. M. Mertens, "Focus group design and group dynamics: Lessons from deaf and hard of hearing participants," Am. J. Eval., vol. 20, no. 2, 1999, doi: 10.1177/109821409902000208.

[7] M. J. Davidson, "PAULA : A Computer-Based Sign Language Tutor for Hearing Adults," 2006.

[8] S. Ebling and J. Glauert, "Building a Swiss German Sign Language avatar with JASigning and evaluating it among the Deaf community," Univers. Access Inf. Soc., vol. 15, no. 4, pp. 577–587, 2016, doi: 10.1007/s10209-015-0408-1.

[9] I. Zwitserlood, M. Verlinden, J. Ros, and S. Van Der Schoot, "SYNTHETIC SIGNING FOR THE DEAF: eSIGN."

[10] T. Uchida et al., "Sign language support system for viewing sports programs," in ASSETS 2017 - Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility, 2017, pp. 339–340. doi: 10.1145/3132525.3134768.

[11] T. Uchida et al., "Systems for supporting deaf people in viewing sports programs by using sign language animation synthesis," ITE Trans. Media Technol. Appl., vol. 7, no. 3, pp. 126–133, 2019, doi: 10.3169/mta.7.126.

[12] J. R. F. Brega, I. A. Rodello, D. R. C. Dias, V. F. Martins, and M. de P. Guimarães, "A virtual reality environment to support chat rooms for hearing impaired and to teach Brazilian Sign Language (LIBRAS)," 2014.

[13] C. Eryigit, H. Köse, M. Kelepir, and G. Eryigit, "Building machine-readable knowledge representations for Turkish sign language generation," Knowl.-Based Syst., vol. 108, 2016, doi: 10.1016/j.knosys.2016.04.014.

[14] S. Cox et al., "TESSA, a system to aid communication with deaf people," in Annual ACM Conference on Assistive Technologies, Proceedings, 2002, pp. 205–212. doi: 10.1145/638286.638287.

[15] R. San-Segundo et al., "Speech to sign language translation system for Spanish," Speech Commun., vol. 50, no. 11–12, pp. 1009–1020, 2008, doi: 10.1016/j.specom.2008.02.001.

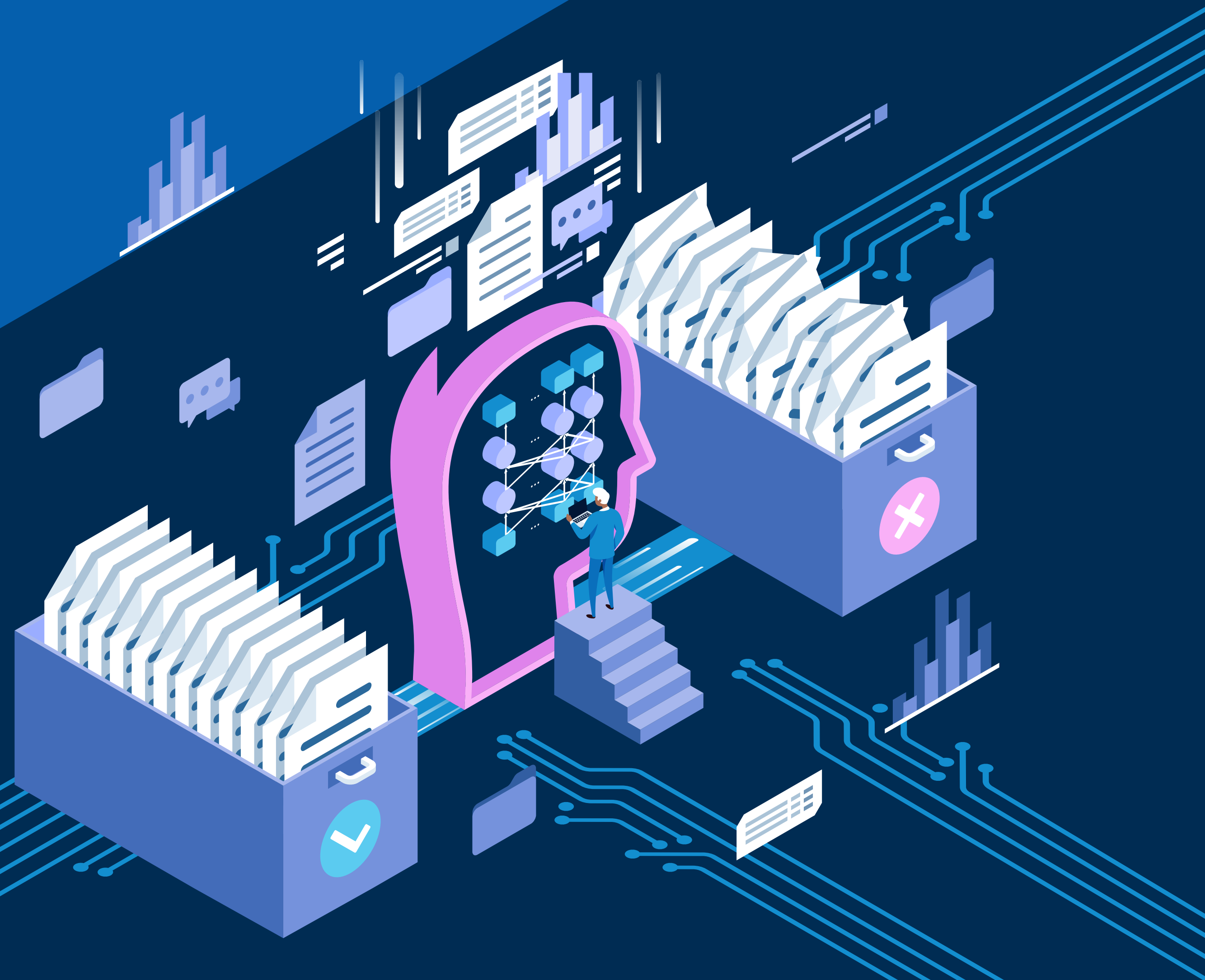
[16] B. D. Patel, H. B. Patel, M. A. Khanvilkar, N. R. Patel, and T. Akilan, "ES2ISL: An Advancement in Speech to Sign Language Translation using 3D Avatar Animator," in Canadian Conference on Electrical and Computer Engineering, 2020, vol. 2020-Augus. doi: 10.1109/CCECE47787.2020.9255783.

[17] J. Oh, S. Jeon, M. Kim, H. Kwon, and I. Kim, "An avatar-based weather forecast sign language system for the hearing-impaired," IFIP Adv. Inf. Commun. Technol., vol. 436, 2014, doi: 10.1007/978-3-662-44654-6_51.

[18] M. M. M. M. El-Gayyar, A. S. A. S. Ibrahim, and M. E. E. Wahed, "Translation from Arabic speech to Arabic Sign Language based on cloud computing," Egypt. Inform. J., vol. 17, no. 3, pp. 295–303, Nov. 2016.

Jumla Sign Language Annotation Tool

an overview



Achraf Othman,
Oussama El Ghoul
Mada Center

This paper describes an ongoing project on developing a new web-based tool for annotating Sign languages. This tool is used to annotate the First Qatari Sign Language dataset called "Jumla Dataset: The Jumla Qatari Sign Language Corpus" with written Arabic text. The annotation of videos in Qatari Sign Language (QSL) takes input from signers to identify the Arabic glosses components toward representing the QSL in a written way with high accuracy, furthermore to the use of the annotation output in the development of computational Sign Language tools. The QSL annotation is based on an input of 4 videos recorded by deaf persons or Sign Language interpreters from different angles (front, left side, right side, and facial view). The output is a JSON file containing all the interpreted sentences given as an entry record. The glosses are annotated for each period and aligned with the Arabic content. Moreover, the presented tool, available as an open-source, provides a management system to classify all records from cameras, motion capture systems, and edited files in addition to the possibility to create components for each gloss annotation terminology depending on the target Sign Language.

Keywords
Annotation Tool, Sign Language Corpus, Qatari Sign Language, Gloss Annotation

Sign languages present the primary way of communication in persons with hearing impairment and the deaf communities used in large groups of persons worldwide. They are considered natural visual languages different than spoken languages [1]. They have a rich grammatical structure with unique syntax and vocabulary [2] compared to spoken languages [3]. However, without a written form, the information cannot be processed by machine and interpreted by humans. For this, sign language can be written in a way that permits anyone to use it to build systems and to develop applications. Sign languages have a spatiotemporal structure different from the linear structure of spoken language. This structure challenges developing the written form to cover syntactic and semantic levels.

Like spoken languages, sign language also has different grammatical levels that require annotation for processing tasks. For each, it can be done automatically or manually according to research progress.

The annotation of sign language allows writing all linguistic information related to the signed language using a text description. Annotation is defined as a special kind of task where we add additional information to a piece of text for its better interpretation, analysis, comprehension, and application [4]. In this paper, we present a new web-based tool developed by Mada to annotate SL. The Jumla Sign Language project is supported by the Mada Innovation Program (MIP) [5]. It provides the first Qatari Sign Language avatar to translate Arabic text to Qatari and Arabic Sign Language.

Literature review

In sign language, the annotation is defined as a critical or explanatory commentary or analysis, comment added to a text, the process of writing such comment or commentary, in computing, metadata added to a document or program. Compared to notation which is defined as a system of characters, symbols, or abbreviated expressions used in an art or science or in mathematics or logic to express technical facts or quantities [6]. In computational sign language, we will refer to the term annotation to describe the written form of sign language. The term notation will be used in linguistics used in the works of Stokoe and Sutton [7].

For the annotation systems used to describe signs in recorded videos, several works can be classified into three categories:

(1) Manual annotation systems: the user must add comments and descriptions for each sign

(2) Automatic annotation: that uses algorithms and machine learning to annotate a video automatically. [Until today, there is no accurate, high-performance system able to reach the same level of manual annotation systems

(3) Supervised annotation systems: they are hybrid systems that allow the user to add manual labels and tags with automatic scripts.

For any type of annotation system, the following sign language components must be included during adding the labels and notes: (1) manual signs: hand shapes, hand orientation, shoulder movements, body movements, body orientation, head movements, eye gaze, facial expression (eyebrows, eyelids, and nose), mouthing (mouth pictures and mouth gestures) and emotion if applicable. It can also have extra grammatical labels like part-of-speech tagging and lemmatization. One of the most cited tools is ELAN [8]. It is a free, multimodal annotation tool for digital audio and video media. It supports multileveled transcription of up to six synchronized video files per annotation document [9]. It was used to annotate the Corpus NGT is a collection of almost 72 hours of dialogues of 92 different signers for whom NGT is the first language [10]. After releasing the software, several research projects used the ELAN tool to annotate videos in different sign languages.

Objectives

The main goal of this work is to provide a free, open-source tool that helps researchers and linguistics annotate videos in different sign languages. Not limited to this and from a research perspective, through the present tool, we try to solve the following issues related to sign language annotation tasks and processing:

(1) Development of an online and collaborative platform and interface with voting systems

(2) Inclusion of manual and non-manual with the integration of manual and automatic labeling techniques (supervised, semi-supervised, and non-supervised)

(3) Integration of multi-level annotation layers. For a sign language input, we can extend the annotation to an intra-linguistic level and an extra-linguistic level.

After recording the raw video, we proceed with tasks to build the corpora and annotate the video. The pre-processing task is essential to crop the video and make it ready for annotation. The annotation task can be performed without the processing phase. The annotation output can enhance the quality of the processing phase output. The meta-data phase enables the annotator to define the types of labels used during the annotation task and can be the output of the pre-processing phase. The meta-data types can be grammatical labels, or any information used during the annotation phases, which helps in the processing phase too. The corpora management model inspires the management model by Dash [4]. We adopt the methodology described in the work of Ide and Suderman, "The Linguistic Annotation Framework" [11], developed by the International Standards Organization (ISO) as a standard ISO TC37 SC4. It identifies the fundamental properties and principles for representing linguistic annotations and leads to the design of an abstract data model. The project is ongoing and still did not implement the complete process of the framework. Until today, an online tool was developed to support annotators to annotate the sign language videos. This task is done as follows:

(1) The webmaster uploads the video to the tool with the sentence in written form

(2) A deaf person accesses the video and segments the video in sign. Each sign will start in a specific frame and end in another frame

(3) A second deaf person add labels related to non-manual signs like facial expression, eye gaze, head movement, body movement, and non-dominant hand

(4) a sign language expert reviews the complete annotation and confirm it

(5) a workshop is organized to review all annotations by a committee and approve the proposed annotation.

The interface contains the following components (Fig 1):

- **Yellow box**
  the sentence in original language (for example in the Arabic language: لم يخبرك بكل ما يعلم (in English: He did not tell you everything he knows)).
- **Red box**
  contains a synchronized preview of 4 recorded videos (front view, left view, right view, and facial view). The annotator can use only one video (for example, front view). In our case, we used four views to ensure annotation accuracy.
- **Blue box**
  the timeline with a progress bar.
- **Orange box**
  contains a list of the sign language components. Each component is added in a separate layer. The tool admin can manage the sign language components (adding, removing) because sign language differs.
- **Green box**
  the annotation zone that contains all labels added by the annotators will be shown from a start frame to an end frame.

For example sentence, we can observe that the sentence was not signed word by word but differently: "معلومة | كامل | خبر | ما فيه" (in English: information | full | news | no). Moreover, to pronunciation labels, non-manual signs to complete the whole meaning of the sentence and the body movement used to place the "no" sign in another space position for reference.

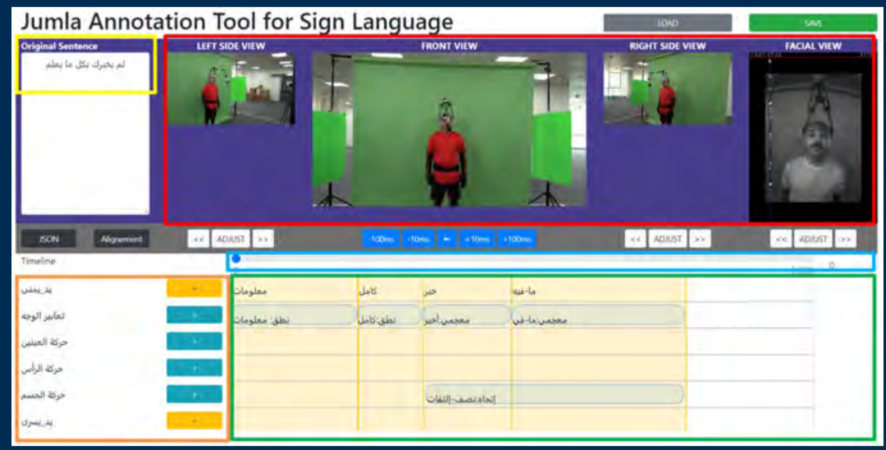


Figure 1.
Jumla annotation tool for sign language.

Annotation tools for sign language must follow a written form. For the Arabic language, we can follow the gloss annotation methodology as applied in American Sign Language [13]. There are several advantages in annotating sign language:

Through sign language annotation tasks, we can understand the structure of sign language sentences and linguistic structure in the absence of sign language grammar.

- Through the annotation tasks, we can identify and retrieve prosody, pronunciation, grammar, meaning, sentence, syntactic, semantic, discourse, and figurative information from the sign language video.
- We can identify the time, location of objects, subjects, and the size of the signing space.
- We can provide a meta-data model for processing purposes.

Moreover, annotation in sign language opened the doors to new research topics, which are sign language segmentation and automatic annotation and not limited to. [12] [13] The benefits of annotated sign language cover several domains such as technology, the commercial sector for machine translation and sign language recognition, ethnographic and humanistic domains. Furthermore, thanks to the fast growth of technology and artificial intelligence, the more we have annotated data, the more research can produce new directions in improving the quality of life and independent living for deaf persons [14].

Sign language annotation tools solve several issues faced in sign language processing from linguistic problems, automatic sign language recognition systems, and the development of statistical machine translation systems. In this work, we completed the annotation of 1600 prerecorded sentences provided for the public. The annotation task takes time to ensure the quality of the output. Moreover, we presented a novel annotation tool for sign language covering several levels of annotation tasks from intra-linguistic to extra-linguistic. The output of the annotation tasks is critical to progress research in computational sign language in general. In addition, this work presents an original contribution to Qatari sign language processing and can be extended to other Arabic and foreign sign languages. Moreover, this can be very useful to enhance and upgrade educational skills within deaf students in primary education.

We at Mada Center highly appreciate and are grateful for the team of the Mada Innovation Program and contributors from the Qatari deaf community, The Qatari Center of Social Cultural for the Deaf and Audio Education Complex. Mocap cleaning tasks and the Jumla sign language project have seen the light thanks to their support and contribution.

References

[1] J. G. Kyle, J. Kyle, B. Woll, G. Pullen, and F. Maddix, Sign language: The study of deaf people and their language. Cambridge university press, 1988.

[2] K. Emmorey, Language, cognition, and the brain: Insights from sign language research. Psychology Press, 2001.

[3] U. Bellugi and S. Fischer, "A comparison of sign language and spoken language," Cognition, vol. 1, no. 2–3, pp. 173–200, 1972.

[4] N. S. Dash, Language Corpora Annotation and Processing. Springer, 2021.

[5] D. Al Thani, A. Al Tamimi, A. Othman, A. Habib, A. Lahiri, and S. Ahmed, "Mada Innovation Program: A Go-to-Market ecosystem for Arabic Accessibility Solutions," in 2019 7th International conference on ICT & Accessibility (ICTA), 2019, pp. 1–3.

[6] "Annotation vs Notation - What's the difference?," WikiDiff, May 05, 2015. //wikidiff.com/annotation/notation (accessed Nov. 30, 2021).

[7] M. Kato, "A study of notation and sign writing systems for the deaf," Intercult. Commun. Stud., vol. 17, no. 4, pp. 97–114, 2008.

[8] O. Crasborn and H. Sloetjes, "Enhanced ELAN functionality for sign language corpora," in 6th International Conference on Language Resources and Evaluation (LREC 2008)/3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora, 2008, pp. 39–43.

[9] O. A. Crasborn and H. Sloetjes, "Using ELAN for annotating sign language corpora in a team setting," 2010.

[10] O. A. Crasborn and I. E. P. Zwitserlood, "The Corpus NGT: an online corpus for professionals and laymen," 2008.

[11] N. Ide and K. Suderman, "The Linguistic Annotation Framework: a standard for annotation interchange and merging," Lang. Resour. Eval., vol. 48, no. 3, pp. 395–418, 2014.

[12] H. Chaaban, M. Gouiffès, and A. Braffort, "Automatic Annotation and Segmentation of Sign Language Videos: Base-level Features and Lexical Signs Classification.," in VISIGRAPP (5: VISAPP), 2021, pp. 484–491.

[13] A. Othman and M. Jemni, "Designing high accuracy statistical machine translation for sign language using parallel corpus: case study English and American Sign Language," J. Inf. Technol. Res. JITR, vol. 12, no. 2, pp. 134–158, 2019.

[14] A. Othman and M. Jemni, "Statistical sign language machine translation: from English written text to American sign language gloss," ArXiv Prepr. ArXiv11120168, 2011.