

# Understanding the challenges in creating Motion capture Sign Language Dataset: A Technical Perspective

Sammy Sedrati, Oussama El-Ghoul  
*Mada Center*

## Abstract –

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

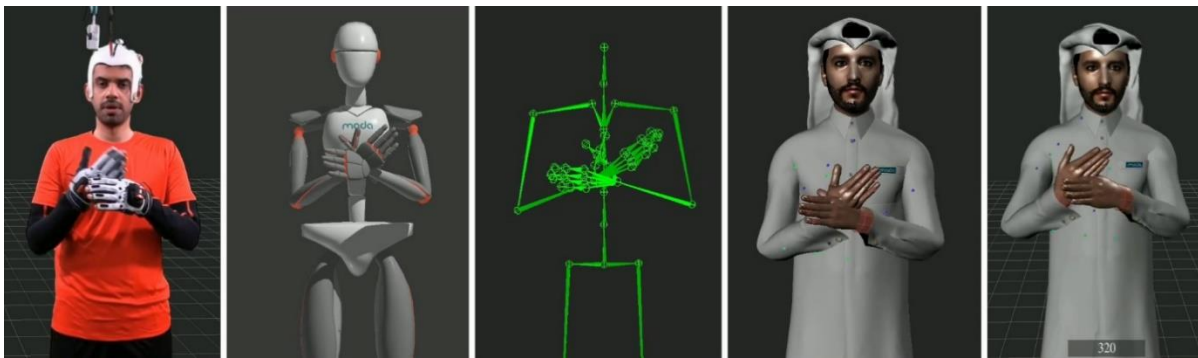
## Introduction

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

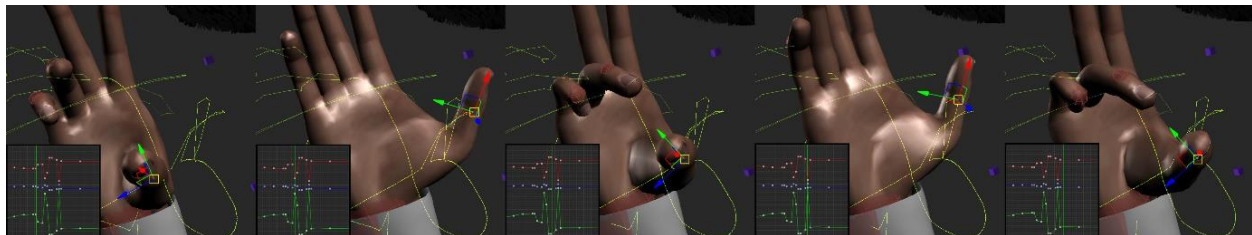
A signer, wearing a mocap suit, translates in SL words/sentences recorded by a computer equipped with mocap's native software. Every motion performed by the signer will be reflected on a dummy skeleton and saved in a format readable by editing software for treatment and clean-up. The cleaning consists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



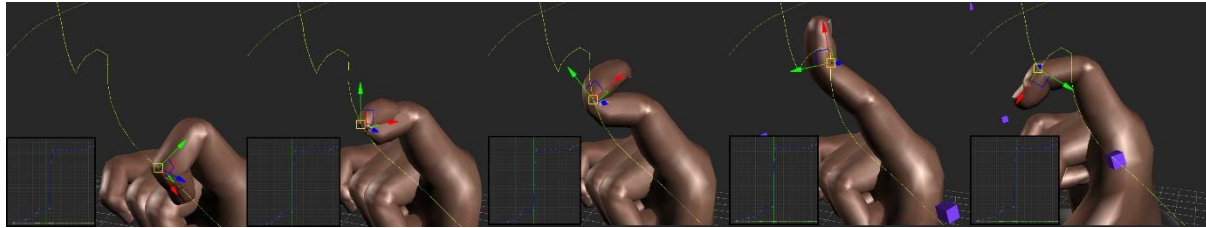
**Figure 1. Cleaning process from Left to right: Signer with mocap suite – real-time capture on native software – imported data in editing software – matching motion with the avatar – cleaned motion on avatar**

## 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.

## Conclusion and perspectives

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.

## Acknowledgment

We 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.