

# SpeakLiz by Talov: Toward a Sign Language Recognition mobile application

## – Supported by Mada Innovation Program

Hugo Jacome Andrade<sup>1</sup>, Shahbaz Ahmed<sup>2</sup>  
*Talov<sup>1</sup>, Mada Center<sup>2</sup>*

### Abstract –

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

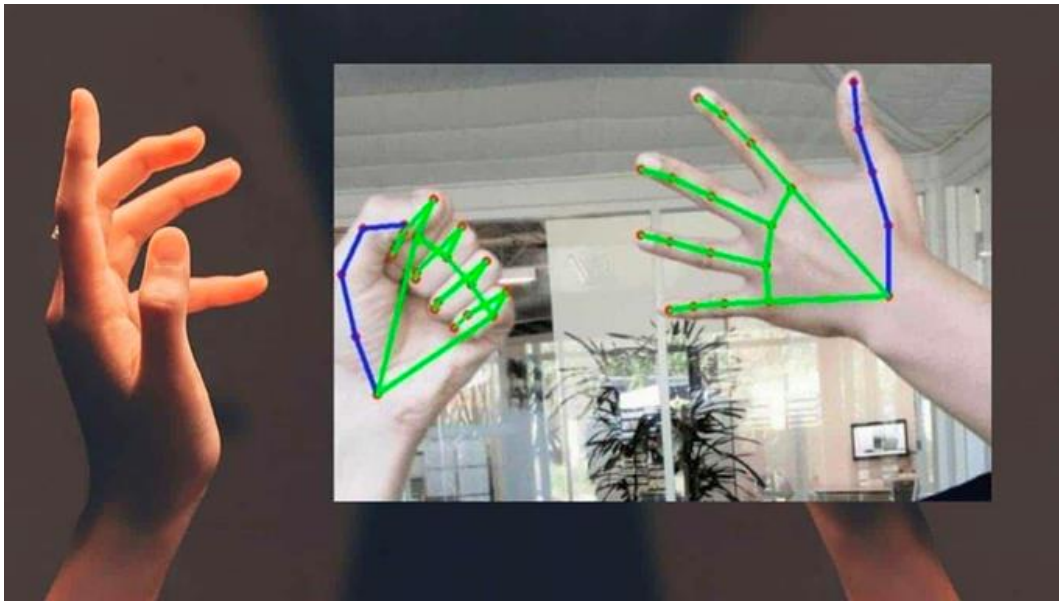
### Introduction

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)**

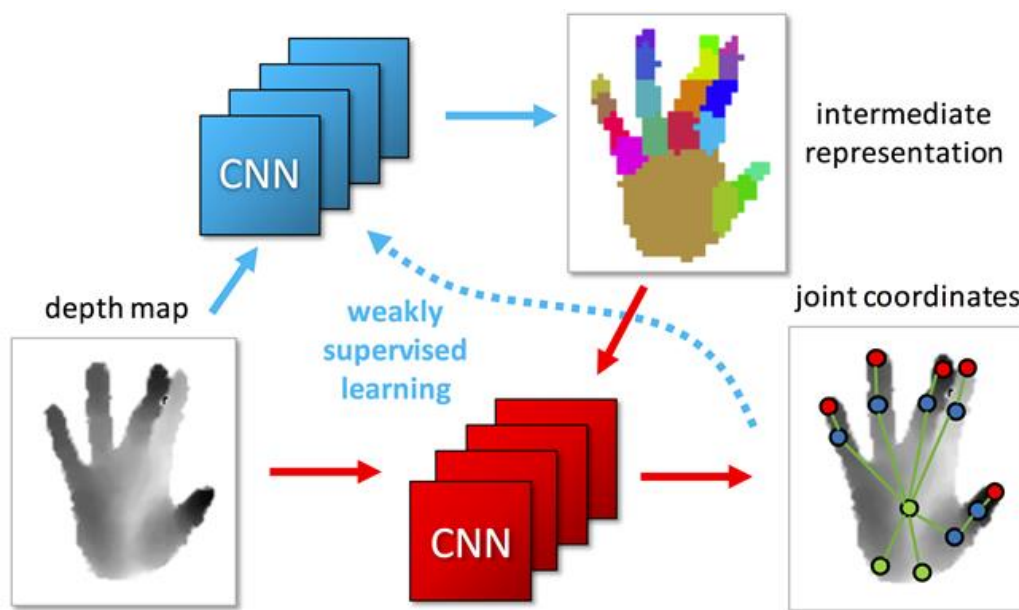
Artificial Intelligence is a vast umbrella discipline that involves all the aspects of giving machines the capacity to be smart [5]. Here is where convolutional neural networks (CNN) conform to one the most common ways for a machine to learn because they are intended to perform some tasks emulating how 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)**

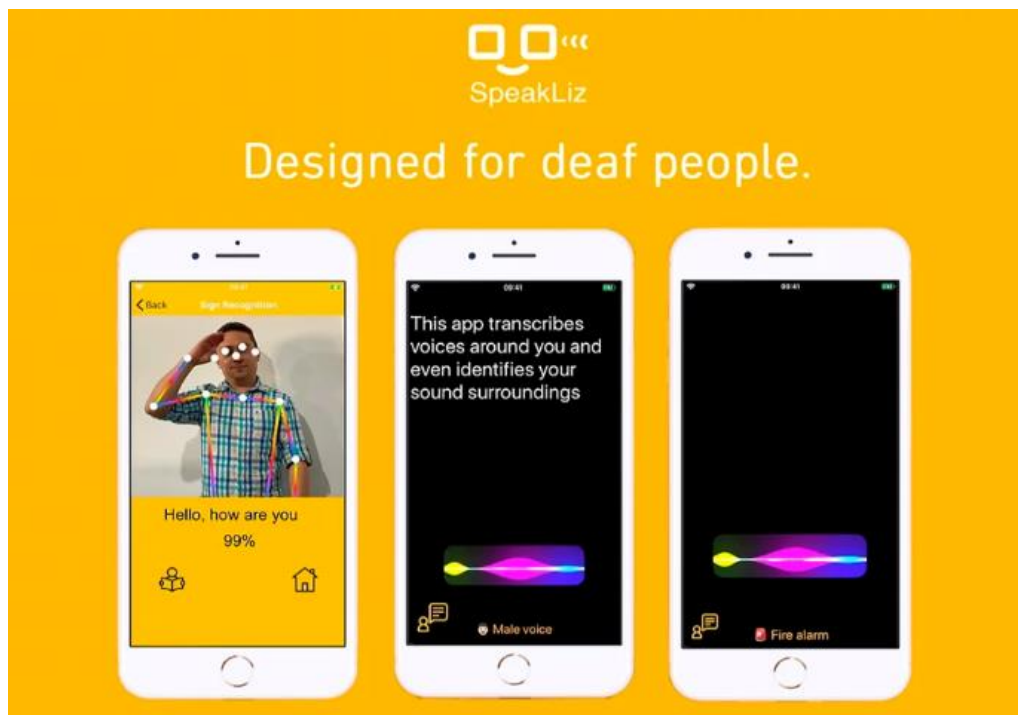
### **SpeakLiz by Talov: supported by Mada Innovation Program**

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.

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).



**Figure 3: SpeakLiz by Talov**

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, and won the second prize in the History Channel is an idea to Change History contest thanks to the Vision 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.

## Conclusion

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.

## Reference

- [1] S. Ameer, A. B. Khalifa, and M. S. Bouhlef, "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.