RECONOCIMIENTO DE LENGUAJE DE SEÑAS MEDIANTE DEEP LEARNING

FRANCO EXEQUIEL SCHÜLER ALLUB

**Contenido**

[Capítulo 1 : Introducción 4](#_Toc96448774)

[1.1 Motivación 4](#_Toc96448775)

[1.2 Objetivo 4](#_Toc96448776)

[1.3 Estructura 4](#_Toc96448777)

[1.4 Glosario de términos 4](#_Toc96448778)

[Capítulo 2 : Estado del Arte 5](#_Toc96448779)

[2.1 El Lenguaje de Señas 5](#_Toc96448780)

[2.1.1 Conceptos básicos 5](#_Toc96448781)

[2.1.2 Iconicidad y modalidad manual 7](#_Toc96448782)

[2.2 Análisis de la situación actual en Nicaragua 8](#_Toc96448783)

[2.2.1 Visión general 8](#_Toc96448784)

[2.3 Sistemas de reconocimiento de gestos y señas 8](#_Toc96448785)

[2.4 Mediapipe 8](#_Toc96448786)

[2.5 Redes Neuronales Recurrentes 8](#_Toc96448787)

[Capítulo 3 : Análisis y Diseño del Sistema 10](#_Toc96448788)

[3.1 Diseño conceptual 10](#_Toc96448789)

[3.2 Requisitos 10](#_Toc96448790)

[3.3 Casos de uso 10](#_Toc96448791)

[Capítulo 4 : Implementación del Sistema 11](#_Toc96448792)

[4.1 Implementación del modelo Holistic 11](#_Toc96448793)

[4.2 Extracción de características 11](#_Toc96448794)

[4.3 Creación de los datasets 11](#_Toc96448795)

[4.4 Entrenamiento (provisional) 11](#_Toc96448796)

[Capítulo 5 : Experimentación 12](#_Toc96448797)

[5.1 Entorno de experimentación 12](#_Toc96448798)

[5.2 Conjunto de pruebas 1 12](#_Toc96448799)

[5.3 Conjunto de pruebas 2 12](#_Toc96448800)

[5.4 … 12](#_Toc96448801)

[Capítulo 6 : Gestión del Proyecto 13](#_Toc96448802)

[6.1 Metodología 13](#_Toc96448803)

[6.2 Planificación 13](#_Toc96448804)

[6.3 Presupuesto 13](#_Toc96448805)

[6.4 Entorno socioeconómico 13](#_Toc96448806)

[6.5 Entorno regulador 13](#_Toc96448807)

[6.6 Gestión de riesgos 13](#_Toc96448808)

[Capítulo 7 : Conclusiones 14](#_Toc96448809)

[Capítulo 8 : Trabajos Futuros 15](#_Toc96448810)

**Índice de ilustraciones**

[Ilustración 1. Representación simbólica de ‘cat’ (gato). Extraído de [1]. 6](#_Toc97370856)

[Ilustración 2. Las 7 formas básicas de Battison. Extraído de [1]. 7](#_Toc97370857)

[Ilustración 3. Ejemplo de colocación de sensor EMG. Extraído de [13] 10](#_Toc97370858)

# Capítulo 1 : Introducción

## 1.1 Motivación

## 1.2 Objetivo

## 1.3 Estructura

## 1.4 Glosario de términos

# Capítulo 2 : Estado del Arte

En este capítulo se presenta la revisión bibliográfica correspondiente al lenguaje de señas. Se tratan temas esenciales para establecer un contexto previo, comprender la importancia del lenguaje de señas, exponer qué implicaciones tiene en la vida de las personas con dificultades auditivas o de habla en la sociedad nicaragüense y definir las principales soluciones dadas a este problema. Finalmente, se exponen las herramientas utilizadas para este proyecto.

## 2.1 El Lenguaje de Señas

Antes de abordar cualquier tema relacionado con la presentación de soluciones para el problema del lenguaje de señas de manera general, resulta crucial entender qué es el lenguaje de señas, qué supone para las personas que precisan de él y qué componentes principales se necesitan para representarlo.

### 2.1.1 Definición y conceptos básicos

El lenguaje de señas es el medio de comunicación que utilizan las personas sordas o con dificultades para la escucha. También es usado por aquellas personas que presentan dificultades para el habla. Cada país tiene su propio lenguaje de señas y cada seña posee su propio sistema gramatical y fonológico.

Valli, C. y Lucas, C. en su libro *Linguistics of American sign language: An introduction* [1] definen el lenguaje, de forma general, como un complejo sistema de comunicación interpersonal gobernado por reglas. Sin estas reglas, las personas serían incapaces de comprender la información que otras personas desean comunicar; es decir, la comunicación en sí sería inviable. Además, el lenguaje es un sistema basado en símbolos. En los lenguajes escritos, por ejemplo, una letra es un símbolo mientras que en los lenguajes hablados los símbolos son las palabras y, sin embargo, ambos están fuertemente relacionados. Puede intuirse pues, que los lenguajes de señas no son completamente distintos al resto de lenguajes.

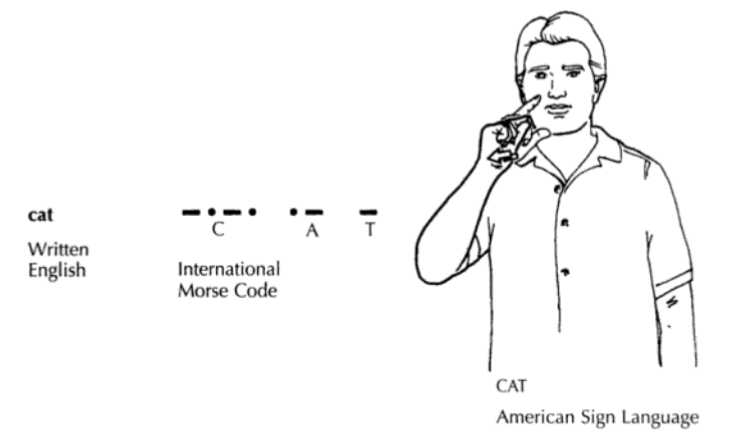


Ilustración . Representación simbólica de ‘cat’ (gato). Extraído de [1].

Estos símbolos pueden analizarse prestando atención a los estudios sobre la estructura de la simbología del lenguaje de señas introducidos por Robbin Battison en 1978 en [2]. Según Battison, la formación de señas está basada en dos condiciones: la Condición de Simetría y la Condición de Prevalecimiento. La Condición de Simetría establece que en la formación de una seña con dos manos implicadas en la cual ambas manos presentan movimiento, ambas tendrán la misma forma y tipo de movimiento. La Condición de Prevalecimiento, en cambio, establece que en la formación de una seña con dos manos implicadas, si las manos presentan una forma diferente, entonces solo una de ellas podrá moverse. De esta manera, Battison descubrió que la mano que no se mueve se encarga de representar 7 formas básicas: B, A, S, O, C, 1 y 5. La conclusión que podemos extraer de esta información es que la estructura de las señas no es arbitraria, si no que estas pueden clasificarse en diferentes categorías siguiendo un patrón.

Las dos condiciones introducidas por Battison implican que las señas pueden clasificarse en dos tipos principales: estáticas o dinámicas. Las señas estáticas se pueden formar con una o ambas manos sin presentar ningún tipo de movimiento. Son las que se utilizan, por ejemplo, para representar el abecedario o los números. Las señas dinámicas se pueden formar con una o ambas manos en las que alguna presente movimiento. Estas últimas suponen un grado más alto de complejidad puesto que son capaces de representar situaciones, relaciones, continuidad en el tiempo, espacio, etc. Este trabajo se centrará en el reconocimiento de señas dinámicas con la finalidad de crear palabras y sentencias compuestas.

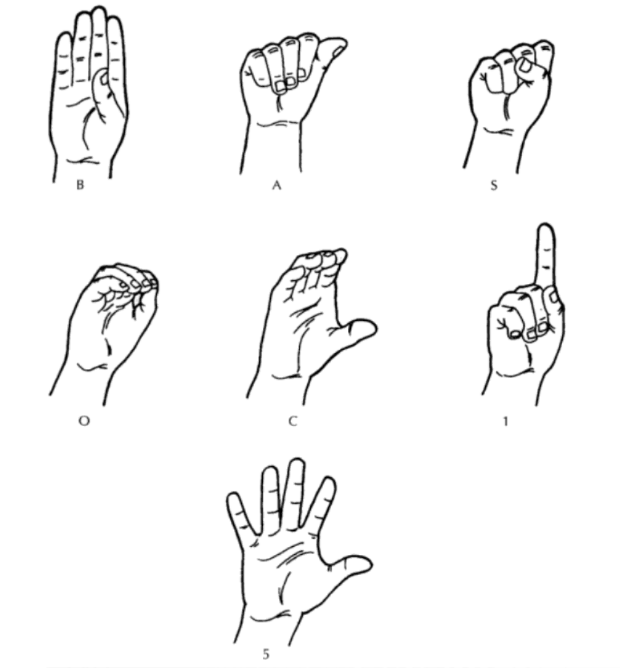


Ilustración . Las 7 formas básicas de Battison. Extraído de [2].

Bellugi, U. y Fischer, S. definen en [3] el lenguaje de señas como un lenguaje que implica aspectos tanto visuales como espaciales, y que se representa haciendo uso de las manos, la cara, el torso, la cabeza y otras partes del cuerpo. Este lenguaje se articula sobre el cuerpo o en el espacio que rodea al cuerpo y rara vez se realiza fuera de este espacio. Por contrapartida, los lenguajes hablados son inherentemente orales y no precisan de una representación corporal más allá de las palabras.

### 2.1.2 Iconicidad y modalidad manual

El lenguaje de señas, de manera general, se reproduce utilizando lo que se conoce como modalidad manual, puesto que proporciona una gran variedad de posibilidades más de iconicidad (o grado de semejanza entre una seña y el concepto que representa) que la modalidad oral. Goldin-Meadow, S. y Brentari, D. hablan sobre esto [4] y ponen como ejemplo la representación de un gato. Resulta posible representar icónicamente a un gato usando las manos (‘dibujando’ la forma de un gato con sus bigotes) o usando el lenguaje oral (haciendo el sonido de un gato). No obstante, expresar relaciones más complejas que involucren al gato, por ejemplo que el gato esté sentado debajo de un objeto, es una tarea difícil si se utilizan estas aproximaciones. Una relación de este tipo puede ser fácilmente representada con gestos. Un ejemplo puede ser utilizar la mano derecha como representación del gato situándola por debajo de la mano derecha, la cual representaría el objeto en cuestión.

De hecho, en los lenguajes de señas es posible la representación de diversas piezas de información en paralelo utilizando la comunicación tanto manual (manos y brazos) como no manual (cara y torso). Estos aspectos son altamente compatibles con la simultaneidad y secuencialidad del lenguaje de señas [5].

## 2.2 Análisis de la situación actual en Nicaragua

En esta subsección se analizarán los entornos sociales, educativos y económicos en Nicaragua en la actualidad con la finalidad de poner en contexto la solución que propone este trabajo y el alcance del mismo.

### 2.2.1 Visión general

Según la OMS [6], se estima que aproximadamente el 15% de la población mundial presenta algún tipo de discapacidad. Este elevado número aumenta considerablemente año a año debido, en gran medida, al incremento de dolencias crónicas, las poblaciones envejecidas, el empobrecimiento de los países y los conflictos armados; por consiguiente, es muy probable que todas las personas experimenten algún tipo de discapacidad a lo largo de su vida, ya sea temporal o crónica.

Estas personas a menudo sufren discriminaciones a la hora de acceder a servicios de salud o simplemente reciben servicios muy deficientes. Resulta crucial el desarrollo de un sistema que elimine las barreras impuestas por las condiciones de discapacidad de las personas, brindándoles así los derechos que merecen.

## 2.3 Sistemas de reconocimiento de gestos y señas

Como ya se ha comentado, existe una gran cantidad de lenguajes de señas alrededor del mundo y cada uno, a su vez, posee su propia estructura, gramática y forma y posición de las señas, convirtiendo el reconocimiento de lenguaje de señas en un problema de gran complejidad en el campo de la visión computacional. En esta sección se presentan algunas de las soluciones propuestas para dicho problema, así como sus ventajas e inconvenientes principales. La tarea de detección de lenguaje de señas usualmente requiere un proceso con varios pasos [7]:

1. Seguimiento o *tracking* de las manos, la cara y la pose del intérprete
2. Segmentación de las manos, cara y pose en diversos puntos de anclaje y creación de modelos con la forma de estos.
3. Reconocimiento de las señas como sentencias.
4. Traducción de las sentencias a texto escrito.

Por otra parte, algunos artículos publicados proponen que los sistemas desarrollados en las últimas décadas pueden clasificarse en dos grandes aproximaciones: detección por sensores y detección visual [8] [9].

### 2.3.1 Sistemas de detección por sensores

Los sistemas de detección mediante sensores basan su funcionamiento en la utilización de sensores que se colocan en alguna parte del cuerpo del intérprete y obtienen diversas métricas relativas al espacio, la velocidad o la aceleración. A menudo, se utilizan de forma complementaria con algoritmos de aprendizaje automático como SVM (*Support Vector Machine*) [21], árboles de decisión [22], redes de neuronas artificiales [23], etc. que ayudan a la obtención de predicciones con resultados notables. Al contrario como se verá en los sistemas de detección visual, estos sistemas no requieren de ninguna técnica de preprocesado y segmentación de los datos [8]. Sin embargo, requieren de algún medio físico como guantes o electrodos los cuales en ciertos casos pueden llegar a ser demasiado intrusivos para el usuario y limitan la efectividad de la solución final a unas condiciones muy específicas . En otras palabras estos sensores, en principio, serían considerablemente incómodos de utilizar y no serían capaces de extrapolar la información a una situación real. Algunos sensores utilizados en los últimos años se exponen a continuación.

Los sensores inerciales [11] han sido usados en numerosos trabajos, colocados en el cuerpo del usuario o utilizados de manera complementaria con sensores EMG. Los sensores EMG son capaces de detectar movimientos musculares del usuario mientras que los inerciales son pequeños y permiten obtener medidas relativas a la aceleración y la velocidad. Suelen colocarse en los músculos del brazo, concretamente en la zona del antebrazo como puede verse en la Ilustración 3. Estas soluciones suelen ser bastante efectivas dado que son considerablemente precisas y no resultan demasiado intrusivas para el usuario [12], a pesar de que otros factores externos como la temperatura, el nivel de hidratación, el flujo de sangre, etc. influyen considerablemente en las mediciones. Algunos trabajos han tratado de lograr la detección de movimientos de manos y dedos [13] [14] [16] haciendo uso de sensores EMG, obteniendo resultados prometedores.

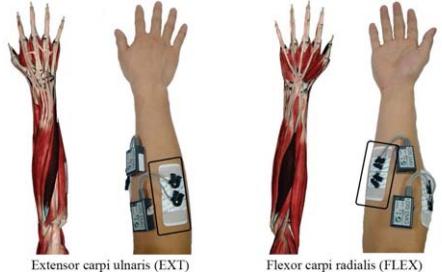


Ilustración . Ejemplo de colocación de sensor EMG. Extraído de [13]

Otros, han hecho uso de productos comerciales. Un ejemplo de ello es el *Myo ArmBand*, un dispositivo que se coloca justo debajo del codo y permite al usuario interaccionar con el entorno vía Bluetooth [15]. Este dispositivo puede llegar a costar hasta 200$ y, a pesar de su eficacia, no es un producto accesible para todo el mundo.

Por otro lado, los sistemas basados en guantes hápticos también han supuesto un gran avance este campo. Estos dispositivos tienen numerosas aplicaciones en robótica, realidad virtual e interacción persona-ordenador [11]. Muchos investigadores han hecho uso de guantes comerciales debido que el diseño del hardware requiere muy poco tiempo. En los últimos años se ha hecho un amplio uso de guantes comerciales de distintas compañías, algunos de los más utilizados son los *CyberGlove* [24] con un coste de entre 13.000 y 18.000 dólares la unidad [11]. Sin embargo, estos guantes comerciales a menudo suponen un gran coste y los investigadores en general optan por construir los guantes ellos mismos dado que, a pesar de requerir más tiempo de diseño y creación y ser menos precisos, son mucho más baratos. Se han encontrado trabajos que presentan soluciones asequibles a estos problemas, fabricando guantes basados en sensores capacitivos [17] [18] ,utilizando guantes coloreados [19] o guantes basados en sensores inerciales [20].



Ilustración . CyberGlove. Extraído de [25]

### 2.3.2 Sistemas de detección visual

El reconocimiento de gestos es un reto interdisciplinario que involucra visión computacional, procesado de imágenes y aprendizaje automático (*Machine Learning*), entre otros. Para lograr que un sistema de este tipo funcione, se tienen que lograr las siguientes características [10] :

1. **Robustez**. Actualmente los datos pueden suponer un gran coste y, además, pueden estar incompletos. El sistema debe ser capaz de comportarse adecuadamente sin importar el estado de los datos.
2. **Eficiencia computacional**. A menudo los sistemas basados en visión computacional requieren predicciones en tiempo real, por lo que los algoritmos de aprendizaje automático utilizados para dichas predicciones deben ser eficientes.

### 2.3.3 Comparación y conclusiones

## 2.4 Mediapipe

## 2.5 Redes Neuronales Recurrentes

# Capítulo 3 : Análisis y Diseño del Sistema

## 3.1 Diseño conceptual

## 3.2 Requisitos

## 3.3 Casos de uso

# Capítulo 4 : Implementación del Sistema

## 4.1 Implementación del modelo Holistic

## 4.2 Extracción de características

## 4.3 Creación de los datasets

## 4.4 Entrenamiento (provisional)

# Capítulo 5 : Experimentación

## 5.1 Entorno de experimentación

## 5.2 Conjunto de pruebas 1

## 5.3 Conjunto de pruebas 2

## 5.4 …

# Capítulo 6 : Gestión del Proyecto

## 6.1 Metodología

## 6.2 Planificación

## 6.3 Presupuesto

## 6.4 Entorno socioeconómico

## 6.5 Entorno regulador

## 6.6 Gestión de riesgos

# Capítulo 7 : Conclusiones

# Capítulo 8 : Trabajos Futuros

# Referencias

[1] Valli, C., & Lucas, C. (2000). *Linguistics of American sign language: An introduction*. Gallaudet University Press.

[2] Battison, R. (1978). Lexical borrowing in American sign language.

[3] Bellugi, U., & Fischer, S. (1972). A comparison of sign language and spoken language. *Cognition*, *1*(2-3), 173-200.

[4] Goldin-Meadow, S., & Brentari, D. (2017). Gesture, sign, and language: The coming of age of sign language and gesture studies. *Behavioral and Brain Sciences*, *40*.

[5] Krebs, Wilbur, R. B., Alday, P. M., & Roehm, D. (2019). The Impact of Transitional Movements and Non-Manual Markings on the Disambiguation of Locally Ambiguous Argument Structures in Austrian Sign Language (ÖGS). *Language and Speech*, *62*(4), 652–680.

[6] *Discapacidad y salud*. (2021, 24 noviembre). Organización Mundial de la Salud. Recuperado 21 de febrero de 2022, de <https://www.who.int/es/news-room/fact-sheets/detail/disability-and-health>

[7] Ronchetti, F. (2018). Reconocimiento de gestos dinámicos y su aplicación al lenguaje de señas. In XX Workshop de Investigadores en Ciencias de la Computación (WICC 2018, Universidad Nacional del Nordeste).

[8] Cheok, M. J., Omar, Z., & Jaward, M. H. (2019). A review of hand gesture and sign language recognition techniques. *International Journal of Machine Learning and Cybernetics*, *10*(1), 131-153.

[9] Elakkiya, R. (2021). Machine learning based sign language recognition: A review and its research frontier. *Journal of Ambient Intelligence and Humanized Computing*, *12*(7), 7205-7224.

[10] Murthy, G. R. S., & Jadon, R. S. (2009). A review of vision based hand gestures recognition. *International Journal of Information Technology and Knowledge Management*, *2*(2), 405-410.

[11] Kudrinko, K., Flavin, E., Zhu, X., & Li, Q. (2020). Wearable sensor-based sign language recognition: a comprehensive review. *IEEE Reviews in Biomedical Engineering*, *14*, 82-97.

[12] Chen, X., Zhang, X., Zhao, Z. Y., Yang, J. H., Lantz, V., & Wang, K. Q. (2007, October). Hand gesture recognition research based on surface EMG sensors and 2D-accelerometers. In *2007 11th IEEE International Symposium on Wearable Computers* (pp. 11-14). IEEE.

[13] Yang, Y., Duan, F., Ren, J., Liu, Z., Zhu, C., Soo, Y., & Mun, K. (2019, July). A multi-gestures recognition system based on less sEMG sensors. In *2019 IEEE 4th International Conference on Advanced Robotics and Mechatronics (ICARM)* (pp. 105-110). IEEE.

[14] Leone, F., Gentile, C., Ciancio, A. L., Gruppioni, E., Davalli, A., Sacchetti, R., ... & Zollo, L. (2019). Simultaneous sEMG classification of hand/wrist gestures and forces. *Frontiers in Neurorobotics*, *13*, 42.

[15] Rawat, S., Vats, S., & Kumar, P. (2016, November). Evaluating and exploring the MYO ARMBAND. In *2016 International Conference System Modeling & Advancement in Research Trends (SMART)* (pp. 115-120). IEEE.

[16] Benatti, S., Casamassima, F., Milosevic, B., Farella, E., Schönle, P., Fateh, S., ... & Benini, L. (2015). A versatile embedded platform for EMG acquisition and gesture recognition. *IEEE transactions on biomedical circuits and systems*, *9*(5), 620-630.

[17] Abhishek, K. S., Qubeley, L. C. F., & Ho, D. (2016, August). Glove-based hand gesture recognition sign language translator using capacitive touch sensor. In *2016 IEEE International Conference on Electron Devices and Solid-State Circuits (EDSSC)* (pp. 334-337). IEEE.

[18] Weissmann, J., & Salomon, R. (1999, July). Gesture recognition for virtual reality applications using data gloves and neural networks. In *IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No. 99CH36339)* (Vol. 3, pp. 2043-2046). IEEE.

[19] Iwai, Y., Watanabe, K., Yagi, Y., & Yachida, M. (1996, October). Gesture recognition by using colored gloves. In *1996 IEEE International Conference on Systems, Man and Cybernetics. Information Intelligence and Systems (Cat. No. 96CH35929)* (Vol. 1, pp. 76-81). IEEE.

[20] Fang, B., Lv, Q., Shan, J., Sun, F., Liu, H., Guo, D., & Zhao, Y. (2019, July). Dynamic gesture recognition using inertial sensors-based data gloves. In *2019 IEEE 4th International Conference on Advanced Robotics and Mechatronics (ICARM)* (pp. 390-395). IEEE.

[21] Mohandes, M. A. (2013). Recognition of two-handed arabic signs using the cyberglove. *Arabian Journal for Science and Engineering*, *38*(3), 669-677.

[22] Kuznetsova, A., Leal-Taixé, L., & Rosenhahn, B. (2013). Real-time sign language recognition using a consumer depth camera. In *Proceedings of the IEEE international conference on computer vision workshops* (pp. 83-90).

[23] Mehdi, S. A., & Khan, Y. N. (2002, November). Sign language recognition using sensor gloves. In *Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP'02.* (Vol. 5, pp. 2204-2206). IEEE.

[24] Kevin, N. Y. Y., Ranganath, S., & Ghosh, D. (2004, November). Trajectory modeling in gesture recognition using CyberGloves/sup/spl reg//and magnetic trackers. In *2004 IEEE Region 10 Conference TENCON 2004.* (pp. 571-574). IEEE.

[25] Dipietro, L., Sabatini, A. M., & Dario, P. (2008). A survey of glove-based systems and their applications. *Ieee transactions on systems, man, and cybernetics, part c (applications and reviews)*, *38*(4), 461-482.