



Universidad Autónoma de Querétaro
Facultad de Informática

**Entrenador virtual basado en avatares e Inteligencia Artificial para el
aprendizaje de la Lengua de Señas Mexicana**

Tesis

Que como parte de los requisitos para obtener el Grado de
Maestro en Ciencias de la Computación

Presenta:

Ing. Felipe de Jesús Rivera Cervantes

Dirigido por:

Dra. Diana Margarita Córdova Esparza

Co-dirigido por:

Dr. Juan Ramón Terven Salinas

Querétaro, Qro., a 4 de junio de 2026

La presente obra está bajo la licencia:
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>



CC BY-NC-ND 4.0 DEED

Atribución-NoComercial-SinDerivadas 4.0 Internacional

Usted es libre de:

Compartir — copiar y redistribuir el material en cualquier medio o formato

La licenciante no puede revocar estas libertades en tanto usted siga los términos de la licencia

Bajo los siguientes términos:



Atribución — Usted debe dar [crédito de manera adecuada](#), brindar un enlace a la licencia, e [indicar si se han realizado cambios](#). Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante.



NoComercial — Usted no puede hacer uso del material con [propósitos comerciales](#).



SinDerivadas — Si [remezcla, transforma o crea a partir](#) del material, no podrá distribuir el material modificado.

No hay restricciones adicionales — No puede aplicar términos legales ni [medidas tecnológicas](#) que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia.

Avisos:

No tiene que cumplir con la licencia para elementos del material en el dominio público o cuando su uso esté permitido por una [excepción o limitación](#) aplicable.

No se dan garantías. La licencia podría no darle todos los permisos que necesita para el uso que tenga previsto. Por ejemplo, otros derechos como [publicidad, privacidad, o derechos morales](#) pueden limitar la forma en que utilice el material.



Universidad Autónoma de Querétaro
Facultad de Informática
Maestría en Ciencias de la Computación

**Entrenador virtual basado en avatares e Inteligencia Artificial para el
aprendizaje de la Lengua de Señas Mexicana**

Tesis

Que como parte de los requisitos para obtener el Grado de
Maestro en Ciencias de la Computación

Presenta:

Ing. Felipe de Jesús Rivera Cervantes

Dirigido por:

Dra. Diana Margarita Córdova Esparza

Co-dirigido por:

Dr. Juan Ramón Terven Salinas

Dra. Diana Margarita Córdova Esparza
Presidente

Dr. Juan Ramón Terven Salinas
Secretario

Dr. Julio Alejandro Romero González
Vocal

Dr. Mauricio Arturo Ibarra Corona
Suplente

Dr. Fidel González Gutiérrez
Suplente

Centro Universitario, Querétaro, Qro.

Junio 2026

México

Dedicatorias

Este trabajo está dedicado al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT), por la confianza depositada en los investigadores mexicanos y por el apoyo financiero otorgado durante los dos años de realización de este trabajo, el cual hizo posible su desarrollo y culminación.

Asimismo, se dedica a la Universidad Autónoma de Querétaro, en particular a la Facultad de Informática, por el acompañamiento académico brindado durante la formación de posgrado y el respaldo institucional al desarrollo de esta investigación.

Agradecimientos

Quiero expresar mi más sincero agradecimiento a todas las personas que estuvieron involucradas en la culminación de mi trabajo de maestría.

En particular, agradezco a la Dra. Diana Margarita Córdova Esparza, directora de esta tesis, y al Dr. Juan Ramón Terven Salinas, codirector de la misma, por su guía, orientación y valiosos consejos, los cuales contribuyeron de manera significativa a que esta investigación tomara un rumbo claro y estructurado.

A mis profesores de la maestría, por los conocimientos y herramientas que me brindaron a lo largo de estos dos años, los cuales fueron fundamentales para el desarrollo de este proyecto y han tenido un impacto significativo en mi formación profesional y personal.

A mis compañeros de la maestría, quienes con el paso del tiempo se convirtieron en grandes amigos y en valiosos colegas. La convivencia, el intercambio de ideas y las experiencias compartidas a lo largo del posgrado contribuyeron de manera significativa a este proceso formativo.

A mi familia, por su apoyo incondicional, comprensión y acompañamiento constante a lo largo de esta etapa de formación. Su respaldo fue fundamental para alcanzar esta meta académica.

Por último, agradezco a las personas que me brindaron apoyo y acompañamiento significativo durante esta etapa, y cuya presencia fue importante a lo largo del proceso de investigación.

Resumen

El presente trabajo tiene como objetivo el desarrollo de un entrenador virtual basado en técnicas de inteligencia artificial para el aprendizaje de la Lengua de Señas Mexicana (LSM), con el propósito de contribuir a la inclusión de las personas sordas y facilitar el acceso a herramientas tecnológicas de apoyo al aprendizaje.

Para ello, se construyó un corpus de videos organizado en doce lecciones temáticas, compuesto por 335 señas de la Lengua de Señas Mexicana (LSM), a partir de grabaciones realizadas por intérpretes profesionales cuya lengua materna es la lengua de señas.

La metodología empleada combina un enfoque cuantitativo con el desarrollo de un sistema interactivo que utiliza una cámara convencional y técnicas de visión por computadora. A partir de los videos, se extrajeron 48 puntos espaciales mediante la herramienta MediaPipe, correspondientes a 21 puntos de la mano derecha, 21 puntos de la mano izquierda y 6 puntos del rostro. Cada uno de estos puntos está definido por coordenadas bidimensionales (x, y). Los puntos obtenidos fueron normalizados y comparados utilizando el algoritmo *Dynamic Time Warping* (DTW), lo que permitió evaluar la similitud entre las señas realizadas por un intérprete y aquellas ejecutadas por los usuarios durante el proceso de aprendizaje.

El sistema desarrollado permite al usuario practicar señas de manera interactiva y visualizar los puntos clave durante la ejecución del gesto. Posteriormente, el sistema proporciona una retroalimentación automática sobre la precisión del movimiento, la cual es evaluada a partir de la distancia calculada con el algoritmo DTW. Durante la etapa de diseño se exploró el uso de avatares virtuales generados automáticamente; sin embargo, debido a las limitaciones observadas en la representación precisa de los movimientos finos de las manos, se optó por emplear videos reales de intérpretes con el fin de garantizar la fidelidad de las señas presentadas.

Para validar el funcionamiento del sistema, se realizaron pruebas con usuarios reales, así como experimentos adicionales bajo distintas condiciones de captura, considerando variaciones en iluminación y resolución, con el objetivo de analizar la robustez del método propuesto. Asimismo, se llevó a cabo una evaluación de usabilidad mediante un cuestionario aplicado a los participantes, cuyos resultados fueron analizados estadísticamente.

Los resultados obtenidos indican que el entrenador virtual facilita el aprendizaje de señas, presenta una percepción positiva por parte de los usuarios y demuestra el potencial del uso de técnicas de inteligencia artificial como apoyo en la enseñanza de la Lengua de Señas Mexicana.

Abstract

This work aims to develop a virtual training system based on artificial intelligence techniques for learning Mexican Sign Language (LSM), with the purpose of contributing to the inclusion of Deaf people and facilitating access to technological tools that support learning.

To this end, a video corpus organized into twelve thematic lessons was constructed, comprising 335 Mexican Sign Language (LSM) signs recorded by professional interpreters whose native language is sign language.

The proposed methodology combines a quantitative approach with the development of an interactive system that operates using a conventional camera and computer vision techniques. From the videos, 48 spatial keypoints were extracted using the MediaPipe framework, corresponding to 21 keypoints from the right hand, 21 from the left hand, and 6 from the face. Each keypoint is defined by two-dimensional coordinates (x , y). The extracted points were normalized and compared using the Dynamic Time Warping (DTW) algorithm, allowing the evaluation of similarity between signs performed by an interpreter and those executed by users during the learning process.

The developed system enables users to practice signs interactively and visualize keypoints during gesture execution. Subsequently, the system provides automatic feedback on movement precision, which is evaluated based on the distance calculated using the DTW algorithm. During the design stage, the use of automatically generated virtual avatars was explored; however, due to observed limitations in accurately representing fine hand movements, real interpreter videos were used to ensure the fidelity of the signs presented.

To validate system performance, experiments were conducted with real users, along with additional tests under different capture conditions, including variations in lighting and resolution, to analyze the robustness of the proposed method. In addition, a usability evaluation was carried out using a questionnaire administered to participants, and the results were statistically analyzed.

The findings indicate that the virtual trainer facilitates sign learning, is positively perceived by users, and demonstrates the potential of artificial intelligence techniques as effective support tools for teaching Mexican Sign Language.

Índice general

Dedicatoria	I
Agradecimientos	II
Resumen	III
Abstract	IV
1. Introducción	1
1.1. Antecedentes	1
1.2. Justificación	6
1.3. Planteamiento del problema	8
1.4. Preguntas de investigación	9
1.5. Hipótesis	9
1.6. Objetivos	10
1.6.1. Objetivo General	10
1.6.2. Objetivos Específicos	10
2. Marco Teórico	11
2.1. Reconocimiento de lengua de señas mediante visión por computadora	11
2.2. Representación mediante puntos clave y uso de MediaPipe	12
2.3. Comparación de secuencias mediante Dynamic Time Warping	12
2.4. Retroalimentación en sistemas de aprendizaje asistido	12
2.5. Evaluación de usabilidad en aplicaciones educativas	13
3. Metodología	14
3.1. Constitución y segmentación del corpus de datos	14
3.1.1. Modelos de referencia (Intérprete)	15
3.1.2. Conjunto de validación y análisis de robustez (Usuario)	16
3.1.3. Curaduría manual y optimización de archivos	17
3.2. Generación de animaciones a partir de video	18
3.3. Extracción de puntos de las manos y la cara	19
3.3.1. Representación temporal de las señas	20
3.4. Normalización y comparación de los puntos	22

3.4.1.	Condiciones de iluminación y resolución para el análisis de robustez	23
3.5.	Comparación cuadro a cuadro mediante distancia euclidiana	24
3.6.	Alineación temporal mediante Dynamic Time Warping (DTW)	25
3.7.	Definición del umbral de aceptación basado en DTW	26
3.8.	Evaluación de robustez ante ruido	26
3.9.	Comparación de los puntos normalizados	26
4.	Resultados	27
4.1.	Evaluación de animaciones generadas para las señas	27
4.2.	Comparación de trayectorias mediante DTW y distancia euclidiana	28
4.2.1.	Análisis mediante mapas de calor con DTW	28
4.2.2.	Comparación con distancia euclidiana	30
4.3.	Desempeño del algoritmo DTW por lecciones	32
4.3.1.	Optimización del almacenamiento de los recursos multimedia	33
4.4.	Evaluación del desempeño computacional del método DTW	34
4.5.	Análisis de robustez del método DTW	35
4.5.1.	Efecto de variaciones en iluminación y resolución	35
4.5.2.	Efecto de oclusiones temporales (Pérdida de rastreo)	36
4.6.	Evaluación del sistema en entornos reales	39
4.7.	Funcionamiento de la aplicación desarrollada	39
4.7.1.	Menú principal de la aplicación	40
4.7.2.	Selección de una lección con señas	40
4.7.3.	Captura de señas	41
4.7.4.	Visualización de puntos clave	42
4.7.5.	Retroalimentación	43
4.8.	Justificación del umbral de aceptación para la distancia DTW	44
4.8.1.	Transformación de la distancia DTW a porcentaje de similitud	44
4.9.	Evaluación de usabilidad mediante encuestas	46
4.10.	Resultados del cuestionario de evaluación con usuarios	46
4.10.1.	Confiabilidad interna – Alfa de Cronbach	47
4.11.	Análisis inferencial mediante prueba t	47
4.11.1.	Facilidad de uso	48
4.11.2.	Utilidad percibida	48
4.11.3.	Satisfacción	49
4.11.4.	Eficiencia del aprendizaje	49
4.11.5.	Interpretación general de resultados	50

4.12. Conclusión del análisis cuantitativo	50
5. Conclusiones y Trabajo Futuro	51
5.1. Conclusiones	51
5.2. Trabajo Futuro	53
A. Cuestionario de Validación de la Aplicación de Lengua de Señas	58

Índice de figuras

1.1. Distribución porcentual de la población con discapacidad según causa, por tipo de actividad cotidiana	7
3.1. Muestra del corpus de referencia: estandarización visual de las lecciones de números (1-11), días de la semana y meses del año ejecutadas por el intérprete profesional	16
3.2. Muestra del conjunto de validación: réplica de las lecciones del corpus por un usuario no experto, utilizada para la evaluación de la robustez algorítmica y el cálculo de distancias DTW	17
3.3. Interfaz de la plataforma DeepMotion mostrando el conjunto de videos cargados	18
3.4. Ejemplos de modelos de avatar generados a partir de los videos del intérprete .	19
3.5. Visualización de los puntos clave obtenidos mediante MediaPipe	20
3.6. Ejemplos ilustrativos de secuencias temporales correspondientes a cuatro señas de la Lengua de Señas Mexicana. Cada seña se muestra como una secuencia de cuatro frames consecutivos (de izquierda a derecha)	21
3.7. Esquema del proceso de normalización: las trayectorias originales se transforman a un espacio normalizado	23
3.8. Ejemplo ilustrativo de las condiciones de iluminación y resolución empleadas para el análisis de robustez, utilizando como referencia la ejecución de la seña correspondiente al número 5	24
3.9. Cálculo de la distancia euclidiana entre dos puntos homólogos en el espacio normalizado	25
3.10. Ejemplo del alineamiento de dos secuencias de señas utilizando el algoritmo Dynamic Time Warping (DTW)	25
4.1. Mapa de calor de la distancia DTW entre los puntos de las manos y el rostro del intérprete y el usuario	29
4.2. Mapa de calor de la distancia DTW considerando únicamente los fotogramas centrales de las secuencias del intérprete y del usuario, con el fin de igualar la cantidad de frames para la comparación visual	30
4.3. Mapa de calor de la distancia euclidiana entre los puntos de las manos y el rostro del intérprete y el usuario, mostrando una escala de valores menor en comparación con DTW debido a la naturaleza punto a punto de esta métrica . .	31

4.4. Mapa de calor de la distancia DTW obtenido a partir de la comparación de las secuencias del intérprete consigo mismas, donde la diagonal principal indica distancia cero al comparar cada seña con su correspondiente secuencia original	32
4.5. Menú principal de la aplicación	40
4.6. Ventana de visualización de una lección con controles de navegación y práctica	41
4.7. Mensaje de preparación mostrado antes de iniciar la práctica de una seña	42
4.8. Visualización de los puntos clave durante la ejecución de una seña	43
4.9. Mensajes de retroalimentación del sistema basados en el umbral de distancia DTW	44
4.10. Distribución de las respuestas del cuestionario de usabilidad por categoría ($N = 33$)	47

Índice de tablas

1.1. Resumen de los trabajos consultados sobre reconocimiento de Lengua de Señas	5
3.1. Distribución del corpus de datos según su función en la investigación	14
4.1. Media y desviación estándar de las distancias DTW por lección, obtenidas a partir de la comparación entre las ejecuciones del intérprete y de los usuarios .	33
4.2. Comparación de métricas de almacenamiento del corpus de LSM ($N = 335$) . .	34
4.3. Evaluación del desempeño computacional (Tiempo y Memoria)	35
4.4. Resultados del análisis de robustez: Impacto de las variaciones de iluminación y resolución	36
4.5. Resultados del análisis de robustez: Impacto de oclusiones temporales (pérdida de rastreo) en la distancia DTW	37
4.6. Resultados de la prueba t para la categoría Facilidad de Uso	48
4.7. Resultados de la prueba t para la categoría Utilidad Percibida	48
4.8. Resultados de la prueba t para la categoría Satisfacción	49
4.9. Resultados de la prueba t para la categoría Eficiencia del Aprendizaje	49

Capítulo 1

Introducción

1.1. Antecedentes

La lengua de señas es un medio de comunicación único y significativo creado para que las personas sordas y con problemas del habla puedan comunicarse con el mundo exterior y poder expresar pensamientos, sentimientos o necesidades que sirven para reforzar la información que se transmite en la conversación diaria. La lengua de señas implica el uso de diferentes partes del cuerpo, como los dedos, la mano, el brazo, la cabeza, el cuerpo y la expresión facial, para transmitir información Cheok et al. (2019).

Sin embargo, la lengua de señas no es común entre la comunidad de oyentes, y son pocos los que pueden entenderlo. Esto supone una auténtica barrera de comunicación entre la comunidad sorda y el resto de la sociedad, como un problema aún no resuelto del todo hasta el día de hoy. Las personas sordas se encuentran en una situación de discapacidad socialmente desapercibida. Los entornos diseñados desde una cultura oral, en la que el acceso a la comunicación y a la información se transmite mediante las lenguas orales (en formato verbal o escrito), provocan situaciones de desventaja evidentes para el ejercicio de los derechos de estas personas Elakkiya (2020).

Las barreras de comunicación y acceso a la información que encuentran las personas sordas no son visibles a los ojos de la sociedad oyente que no vive ni conoce sus circunstancias, de ahí la importancia que adquieren la toma de conciencia, la accesibilidad universal, las medidas contra la discriminación y las medidas de acción positiva, para lograr la eliminación de las barreras y garantizar el disfrute de sus derechos en las mismas condiciones que sus conciudadanos oyentes Jauregi et al. (2020).

El estado actual de la técnica de reconocimiento del lenguaje de signos (SLR) lleva aproximadamente 30 años de retraso con respecto a los sistemas de reconocimiento del habla debido a múltiples causas. Una de las principales razones es que el procesamiento y el reconocimiento de las señales de vídeo bidimensionales son muy complejos que el procesamiento de las señales de audio unidimensionales.

Además, los elementos léxicos y semánticos del lenguaje de signos aún no se han descubierto del todo, y tampoco existen diccionarios estándares. Asimismo, no existen definiciones comunes para un número tan elevado de signos. El reconocimiento y la clasificación del lenguaje de signos experimentaron un crecimiento en publicaciones de investigación a principios de los años 90. En la mayoría de los casos, los sistemas analizados en dicho estudio tardan casi 10 segundos en procesar el vídeo del signante y traducirlo a texto Elakkiya (2020).

Los primeros estudios sobre el reconocimiento de la lengua de signos se basaban principalmente en el análisis de los equipos de uso que requieren que el signante porte un accesorio como un guante o una pulsera que permita transmitir los movimientos hacia el sistema de procesamiento que los interpretará y dará la salida correspondiente. Sin embargo, los rastreadores de movimiento y los guantes de datos no solamente son caros, sino que también tienen un valor práctico reducido. Con el rápido desarrollo del aprendizaje profundo y sus aplicaciones de visión por ordenador, la investigación del lenguaje de signos basada en imágenes RGB públicas se ha convertido en un punto de interés Huang et al. (2022).

Para reconocer las señales del lenguaje, se necesita una metodología de aprendizaje como K-nearest neighbors (KNN), redes neuronales artificiales (ANN) y Hidden Markov Models (HMM). Una red neuronal es un conjunto de neuronas artificiales interconectadas que utilizan modelos matemáticos para procesar información. Las múltiples conexiones entre las neuronas forman un sistema adaptativo cuyos pesos se actualizan mediante un determinado algoritmo de aprendizaje. Las redes neuronales se han utilizado en varios campos de aplicación con diferentes algoritmos de aprendizaje Espejel-Cabrera et al. (2021).

En los últimos años, el desarrollo de sistemas de reconocimiento del lenguaje de señas basado en sensores RGB-D se ha generalizado. Lo anterior debido a que la información 3D proporcionada por los sensores de profundidad mejora la interactividad, así como la comodidad para el usuario. También se debe tener en cuenta que los precios son cada vez más asequibles para este tipo de dispositivos.

Un sensor RGB-D, como su nombre lo dice nos permite capturar la información fotométrica en el esquema RGB. Además de poder tener los datos correspondientes a la profundidad con la que se encuentran los objetos en una escena. Un ejemplo es el sensor Kinect, un dispositivo que pertenece a este tipo de sensores RGB-D. El sensor Kinect fue creado por Microsoft principalmente para ser usado para la consola de video juegos Xbox. Sin embargo, la comunidad científica encontró en este una herramienta de bajo coste para ser usada en diversos proyectos y aplicaciones Trujillo Romero and García Bautista (2021).

No obstante, los dispositivos mencionados implican costos adicionales y requieren espacio fí-

sico para su instalación, lo que puede resultar inconveniente o incluso incómodo para algunos usuarios. Sería ideal llevar a cabo este proceso de manera óptima y económica, dado que dichos dispositivos podrían representar un gasto significativo para muchas personas en México.

Los sensores de los guantes sensoriales se sustituyen por cámaras, con lo que se reduce el coste de construcción del sistema. Hoy en día, los ordenadores portátiles y los teléfonos móviles llevan incorporadas cámaras de alta resolución. A pesar de todo ello, existen varios inconvenientes en el enfoque basado en la visión, como la necesidad de múltiples cámaras para producir resultados sólidos, los elevados costes computacionales, los problemas de profundidad y oclusión al utilizar webcam, etc.

Afortunadamente, existen investigaciones que hacen uso exclusivo de una cámara convencional, como es el caso del modelo basado en una Convolutional Neural Network (CNN) para el reconocimiento de gestos de la mano con los dedos. Este enfoque aborda la detección imprecisa de modelos de posturas similares mediante una detección de gestos estáticos basada en la postura de la mano, logrando una precisión del 98.6 % utilizando imágenes de esqueletos. La ubicación de la marca de la mano 2D en las poses de la mano se estimó utilizando Mediapipe y una cámara convencional Robert and Duraisamy (2023).

Con el objetivo de complementar los antecedentes presentados y considerar avances recientes en el área, se incorporaron estudios publicados en los últimos años relacionados con el reconocimiento automático de lengua de señas, la traducción asistida y el uso de inteligencia artificial con fines educativos. Estos trabajos permiten identificar tendencias actuales, enfoques metodológicos predominantes y áreas de oportunidad para el desarrollo de herramientas tecnológicas orientadas al aprendizaje de lenguas de señas.

En años recientes, la investigación en reconocimiento y aprendizaje de lenguas de señas ha mostrado un crecimiento significativo, impulsado principalmente por los avances en inteligencia artificial, visión por computadora y aprendizaje profundo. Diversos estudios coinciden en que estas tecnologías han permitido mejorar la detección, el modelado y la clasificación de gestos manuales a partir de secuencias de video, lo que ha favorecido el desarrollo de sistemas más precisos y robustos para el reconocimiento automático de señas Abdullah et al. (2024a).

Asimismo, se ha identificado una tendencia hacia la consolidación de enfoques basados en aprendizaje automático y aprendizaje profundo como herramientas clave para atender las necesidades de la comunidad sorda, especialmente en contextos educativos. Revisiones recientes destacan que estos paradigmas permiten capturar de manera más efectiva las variaciones espaciales y temporales presentes en las señas, en comparación con métodos tradicionales, lo que resulta particularmente relevante para aplicaciones de enseñanza y entrenamiento asistido Ab-

dullah et al. (2024b); Singh et al. (2020).

Por otro lado, investigaciones enfocadas en la traducción y el apoyo del aprendizaje de lenguas de señas mediante inteligencia artificial subrayan la importancia de desarrollar sistemas interactivos que no solamente reconozcan las señas, sino que también proporcionen retroalimentación al usuario durante el proceso de aprendizaje. En este sentido, los estudios recientes enfatizan la necesidad de integrar componentes pedagógicos y de usabilidad en los sistemas propuestos, con el fin de mejorar la experiencia del usuario y facilitar la adquisición progresiva de la lengua de señas Ortiz-Leon et al. (2023).

De manera complementaria, algunas propuestas han explorado el uso de tecnologías alternativas, como entornos virtuales inmersivos, avatares y dispositivos robóticos, con el objetivo de apoyar el aprendizaje y la práctica de lenguas de señas. Si bien estos enfoques presentan resultados prometedores en términos de interacción y motivación, diversos autores señalan que aún existen limitaciones en la representación precisa de los movimientos finos de las manos, lo que puede afectar la fidelidad de las señas mostradas al usuario Lara-Ortiz et al. (2023).

Finalmente, estudios recientes sobre reconocimiento de gestos manuales en video en tiempo real destacan la relevancia de emplear cámaras convencionales y técnicas de visión por computadora como una alternativa accesible y de bajo costo. Estos trabajos resaltan que la extracción de puntos clave de manos y rostro, junto con métodos de comparación temporal, constituye una base sólida para el desarrollo de sistemas de reconocimiento y entrenamiento de señas aplicables en contextos educativos reales Patel and Shah (2023); Williams and Carter (2022).

Tabla 1.1*Resumen de los trabajos consultados sobre reconocimiento de Lengua de Señas*

Autor(es)	Año	Título	Técnica / Enfoque	Resultados / Aportaciones
Cheok et al.	2019	A review of hand gesture and sign language recognition techniques	Procesamiento de imágenes y visión por computadora.	Identificación de desafíos técnicos como variabilidad de gestos y oclusión.
Elakkiya	2020	Machine learning based sign language recognition: a review and its research frontier	Reconocimiento basado en aprendizaje automático.	Evidencia tiempos de procesamiento de casi 10 segundos en traducción de video a texto en sistemas previos.
Espejel-Cabrera et al.	2021	Mexican sign language segmentation using color based neuronal networks...	Redes Neuronales Artificiales (ANN).	Uso de modelos matemáticos y pesos actualizables mediante algoritmos de aprendizaje.
Trujillo Romero & García Bautista	2021	Reconocimiento de palabras de la Lengua de Señas Mexicana utilizando información RGB-D	Sensores RGB-D (Kinect).	Captura de información fotométrica y de profundidad logrando una alternativa funcional de bajo coste.
Huang et al.	2022	Dynamic Sign Language Recognition Based on CBAM with Autoencoder...	Aprendizaje profundo e imágenes RGB.	Destaca la utilidad de bases de datos públicas RGB como punto de interés para la investigación.
Robert & Duraisamy	2023	A review on computational methods based automated sign language recognition...	Convolutional Neural Network (CNN) y Mediapipe.	Precisión del 98.6% en detección de gestos estáticos utilizando imágenes de esqueletos 2D con cámara convencional.
Lara-Ortiz et al.	2023	Assisted Sign Language Tutoring System Using a Bioinspired Virtual Controlled...	Avatares y dispositivos robóticos en entornos virtuales.	Resultados prometedores en interacción y motivación, con limitaciones en la representación de movimientos finos.
Patel & Shah	2023	Recent Trends and Advancements in Hand Gesture Recognition from Real Time Video	Cámaras convencionales y técnicas de comparación temporal.	Extracción de puntos clave de manos y rostro aplicable a la creación de sistemas en contextos educativos reales.

1.2. Justificación

En México, la discapacidad representa un reto relevante para la inclusión social y el acceso equitativo a la información y la comunicación. De acuerdo con el Instituto Nacional de Estadística y Geografía (INEGI), en el Censo de Población y Vivienda 2020 se estima que aproximadamente 6.1 millones de personas presentan algún tipo de discapacidad en el país. Entre las principales causas asociadas se encuentran las enfermedades, la edad avanzada y las condiciones presentes desde el nacimiento, factores que inciden en diversas limitaciones para realizar actividades de la vida cotidiana Instituto Nacional de Estadística y Geografía (2021a,b).

En particular, la discapacidad auditiva constituye un aspecto crítico dentro de este contexto. De los 6.1 millones de personas con discapacidad, el 21.9 % (aproximadamente 1.35 millones de personas) presenta dificultad severa para oír, aun usando aparato auditivo. Si a esto se suma la población que reporta alguna limitación auditiva leve, la cifra asciende a más de 5.1 millones de individuos.

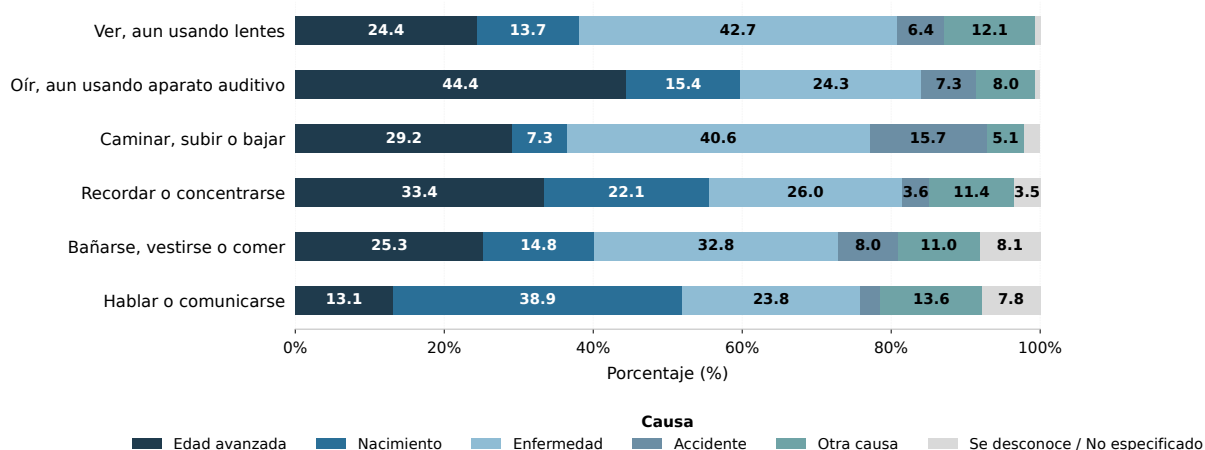
Los datos derivados del mismo censo indican que esta dificultad para oír se relaciona principalmente con la edad avanzada (44.4 %), seguida por enfermedades (24.3 %) y causas presentes desde el nacimiento (15.4 %). La Figura 1.1 muestra la distribución de la población con discapacidad según la causa y el tipo de actividad cotidiana afectada. Al observar esta comparativa, destaca que, a diferencia de limitaciones como el hablar o comunicarse —donde las condiciones de nacimiento son la causa predominante (38.9 %)—, la pérdida auditiva es mayoritariamente adquirida en etapas avanzadas de la vida o por enfermedad.

Estos resultados evidencian la diversidad de factores asociados a la discapacidad auditiva y la magnitud demográfica del problema. Que la mayoría de los casos sean adquiridos subraya que gran parte de esta población no tiene a la Lengua de Señas Mexicana como lengua materna, reforzando la importancia de desarrollar herramientas tecnológicas, como entrenadores virtuales basados en Inteligencia Artificial, orientadas a facilitar y mejorar sus procesos de comunicación y aprendizaje.

La Lengua de Señas constituye un medio de comunicación fundamental para las personas sordas, al permitirles expresar ideas, necesidades y emociones mediante el uso coordinado de movimientos de manos, brazos, cuerpo y expresiones faciales. No obstante, su conocimiento y uso siguen siendo limitados dentro de la comunidad oyente, lo que genera una barrera significativa de comunicación entre ambos grupos y restringe el acceso equitativo a diversos servicios y oportunidades.

Figura 1.1

Distribución porcentual de la población con discapacidad según causa, por tipo de actividad cotidiana



Nota. La categoría 'Se desconoce / No especificado' fue calculada y añadida para completar el 100 % de la distribución; los tabulados originales no grafican explícitamente esta categoría. Las etiquetas de datos menores al 3 % fueron omitidas por claridad visual. Fuente: Elaboración propia con datos del Censo de Población y Vivienda 2020 e informes oficiales del INEGI Instituto Nacional de Estadística y Geografía (2021a,b).

Esta barrera se manifiesta de manera crítica en ámbitos como la salud y el empleo. A partir de entrevistas realizadas a personas con amplia experiencia en el uso de la Lengua de Señas Mexicana, se identificaron problemáticas recurrentes asociadas a la ausencia de intérpretes en instituciones públicas y privadas. En el sector salud, esta carencia dificulta la atención médica adecuada y la expresión clara de síntomas o necesidades; mientras que, en el ámbito laboral, limita la comunicación efectiva en procesos administrativos, de capacitación y de desarrollo profesional. Estas situaciones evidencian una vulneración indirecta de derechos fundamentales y refuerzan la necesidad de promover acciones que reduzcan la brecha comunicativa.

En este contexto, fomentar el aprendizaje de la Lengua de Señas entre la población oyente resulta una estrategia clave para avanzar hacia una sociedad más inclusiva. El desarrollo de herramientas tecnológicas orientadas al apoyo del aprendizaje de la Lengua de Señas no solo contribuye a sensibilizar sobre esta problemática, sino que también ofrece alternativas accesibles y escalables para facilitar su adquisición.

Por lo anterior, el presente proyecto se justifica en la necesidad de diseñar y evaluar un entrenador virtual apoyado en técnicas de inteligencia artificial que facilite el aprendizaje de la Lengua de Señas Mexicana. Esta propuesta busca aportar una solución tecnológica que complemente

los métodos tradicionales de enseñanza, promueva la inclusión social y sienta las bases para el desarrollo de futuras aplicaciones orientadas al apoyo de personas con discapacidad auditiva.

1.3. Planteamiento del problema

La Lengua de Señas constituye el principal medio alternativo de comunicación para las personas con discapacidad auditiva o del habla. No obstante, su comprensión y uso siguen siendo limitados entre la población oyente, lo que dificulta la interacción cotidiana entre ambos grupos y perpetúa barreras de comunicación significativas Putera et al. (2018).

En los últimos años, se han desarrollado diversas tecnologías orientadas a apoyar a la comunidad sorda, principalmente con el objetivo de facilitar la inclusión social y reducir las barreras comunicativas en distintos entornos. Entre estas tecnologías, el reconocimiento automático del lenguaje de señas ha sido ampliamente estudiado mediante enfoques basados en procesamiento de imágenes, visión por computadora y aprendizaje automático. Sin embargo, esta línea de investigación aún enfrenta importantes desafíos técnicos relacionados con la variabilidad de los gestos, la segmentación adecuada de las manos y la complejidad de los movimientos Cheok et al. (2019); Elakkiya (2020); Singh et al. (2020).

Entre los principales obstáculos identificados en la literatura se encuentran la oclusión de las manos, las variaciones en las condiciones de iluminación, las transformaciones afines de la mano durante el movimiento, la limitada disponibilidad de bases de datos suficientemente amplias y la alta demanda computacional necesaria para construir sistemas de reconocimiento precisos y robustos Cheok et al. (2019); Robert and Duraisamy (2023).

De manera paralela, el uso de dispositivos y tecnologías de captura, como sensores de profundidad y cámaras convencionales, ha permitido obtener información gestual con mayor facilidad. Estos dispositivos facilitan la adquisición de datos relacionados con la posición, movimiento y configuración de las manos, lo cual resulta fundamental para el reconocimiento automático de señas. Métodos basados en segmentación por color de piel, extracción de características geométricas y análisis de movimiento han sido ampliamente utilizados en diferentes investigaciones, mientras que enfoques más recientes incorporan modelos de aprendizaje automático y aprendizaje profundo para mejorar el rendimiento en la clasificación de gestos y señas Kumar et al. (2019); Huang et al. (2022); Mejía Pérez et al. (2022).

En el contexto específico de la Lengua de Señas Mexicana (LSM), diversos trabajos han abordado el reconocimiento automático utilizando información visual proveniente de cámaras RGB

o sensores de profundidad. Estas investigaciones pueden clasificarse, de manera general, en tres enfoques principales: reconocimiento del alfabeto y números, reconocimiento de palabras y reconocimiento de oraciones. No obstante, la mayoría de estas propuestas se centra principalmente en la identificación o traducción automática de señas, sin abordar de manera directa el proceso de aprendizaje de la lengua Trujillo Romero and García Bautista (2021); Carmona Arroyo et al. (2021).

Esta limitación evidencia una carencia de herramientas tecnológicas orientadas al aprendizaje de la Lengua de Señas Mexicana, lo cual dificulta que personas oyentes adquieran conocimientos básicos de este lenguaje y limita las posibilidades de interacción con la comunidad sorda. En consecuencia, surge la necesidad de desarrollar soluciones que no solo reconozcan señas, sino que también apoyen activamente el proceso de aprendizaje mediante mecanismos de retroalimentación y herramientas interactivas que faciliten la práctica y evaluación del usuario.

1.4. Preguntas de investigación

1. ¿Es posible desarrollar una herramienta tecnológica para el aprendizaje de la Lengua de Señas Mexicana que funcione como un entrenador virtual apoyado en técnicas de inteligencia artificial?
2. ¿Es posible proporcionar retroalimentación automática al usuario durante el proceso de aprendizaje de la Lengua de Señas Mexicana a partir del análisis de las señas realizadas por el propio usuario, captadas mediante una cámara convencional?

1.5. Hipótesis

El desarrollo de un entrenador virtual basado en el uso de una cámara convencional y técnicas de inteligencia artificial facilita el aprendizaje de un conjunto de palabras de la Lengua de Señas Mexicana, al funcionar como una herramienta tecnológica de apoyo para las personas interesadas en adquirir conocimientos en esta lengua, en comparación con los métodos tradicionales de enseñanza.

1.6. Objetivos

1.6.1. Objetivo General

Desarrollar un modelo de entrenador virtual, utilizando técnicas de inteligencia artificial y una cámara convencional, que funcione como herramienta de apoyo para el aprendizaje de un conjunto de palabras de la Lengua de Señas Mexicana por parte de los usuarios.

1.6.2. Objetivos Específicos

1. Diseñar e implementar un entrenador virtual basado en el uso de videos de intérpretes como modelo de referencia, con el fin de proporcionar retroalimentación efectiva a los usuarios interesados en aprender la Lengua de Señas Mexicana.
2. Obtener y sincronizar los puntos corporales del usuario y de los modelos de referencia, a partir de videos de intérpretes de Lengua de Señas Mexicana, con el propósito de verificar la correcta ejecución de una seña.
3. Desarrollar un algoritmo de comparación destinado a contrastar las señas ejecutadas por el usuario con aquellas obtenidas de los modelos de referencia.
4. Realizar pruebas del entrenador virtual con el fin de validar su funcionalidad y rendimiento.

Capítulo 2

Marco Teórico

El desarrollo de sistemas para el aprendizaje de la Lengua de Señas Mexicana requiere la integración de múltiples áreas de conocimiento, que van desde la visión por computadora hasta los métodos de evaluación del aprendizaje. En este capítulo se presentan los fundamentos teóricos que sustentan la propuesta de esta investigación, abordando los principales enfoques para el reconocimiento de señas, las técnicas de representación del movimiento humano mediante puntos clave, los métodos de comparación de secuencias temporales, así como los elementos relacionados con la retroalimentación y la evaluación de usabilidad en sistemas educativos. Estos conceptos permiten contextualizar y justificar las decisiones metodológicas adoptadas en el desarrollo del sistema propuesto.

2.1. Reconocimiento de lengua de señas mediante visión por computadora

El reconocimiento automático de lengua de señas representa un reto complejo dentro del campo de la visión por computadora, debido a la naturaleza multimodal de las señas, las cuales combinan movimientos de manos, expresiones faciales y postura corporal. En años recientes, los avances en aprendizaje profundo han permitido mejorar significativamente el desempeño de estos sistemas, particularmente mediante el uso de arquitecturas que modelan tanto la información espacial como temporal Camgoz et al. (2020).

Diversos enfoques han empleado redes neuronales convolucionales para extraer características visuales a partir de imágenes, combinadas con modelos secuenciales para capturar la dinámica del movimiento. Sin embargo, en aplicaciones donde se busca eficiencia computacional e interacción en tiempo real, se han explorado alternativas basadas en representaciones estructuradas del cuerpo humano, como los modelos de pose Zhang et al. (2020).

2.2. Representación mediante puntos clave y uso de MediaPipe

La representación del movimiento humano a través de puntos clave (keypoints) permite abstraer la información visual en estructuras numéricas que describen la posición de articulaciones relevantes. Este enfoque reduce la dependencia de condiciones de iluminación, fondo o calidad de imagen, facilitando el procesamiento de las señas.

Frameworks como MediaPipe han popularizado este tipo de aproximación al proporcionar soluciones eficientes para la detección de manos, rostro y cuerpo en tiempo real Lugaresi et al. (2019). En el contexto de esta investigación, el uso de puntos clave permite comparar ejecuciones de señas entre distintos usuarios, manteniendo consistencia en la representación de los movimientos.

2.3. Comparación de secuencias mediante Dynamic Time Warping

Una de las principales dificultades en el análisis de señas es la variabilidad en la velocidad y ejecución entre diferentes personas. Para abordar este problema, el algoritmo Dynamic Time Warping (DTW) ha sido ampliamente utilizado como una técnica para medir similitud entre secuencias temporales.

DTW permite alinear dos secuencias que pueden diferir en duración o ritmo, encontrando una correspondencia óptima entre sus elementos Müller (2007). Este enfoque ha demostrado ser efectivo en tareas de reconocimiento de gestos cuando se trabaja con datos derivados de esqueletos o puntos clave Zhao et al. (2021). En este trabajo, DTW se emplea para comparar las ejecuciones del usuario con modelos de referencia, permitiendo cuantificar el grado de similitud.

2.4. Retroalimentación en sistemas de aprendizaje asistido

La retroalimentación inmediata es un componente fundamental en entornos de aprendizaje interactivo. En sistemas basados en reconocimiento de gestos, proporcionar información sobre la

precisión de la ejecución permite al usuario identificar errores y mejorar progresivamente su desempeño.

Estudios en el área de interacción humano-computadora han demostrado que la retroalimentación basada en métricas cuantitativas mejora significativamente la experiencia del usuario y favorece el aprendizaje autónomo Alvarez et al. (2021). En este sentido, la integración de métricas de similitud como DTW dentro de una interfaz visual contribuye a generar un proceso de aprendizaje más intuitivo.

2.5. Evaluación de usabilidad en aplicaciones educativas

La evaluación de usabilidad permite determinar si un sistema cumple con criterios de facilidad de uso, eficiencia y satisfacción del usuario. En aplicaciones educativas, estos factores son especialmente relevantes, ya que influyen directamente en la adopción de la tecnología.

Instrumentos como la escala SUS (System Usability Scale) han sido ampliamente utilizados para medir la percepción del usuario en sistemas interactivos Brooke (2019). Asimismo, métricas complementarias como cuestionarios tipo Likert y análisis estadísticos permiten validar la consistencia y confiabilidad de los resultados obtenidos Lewis (2018).

En esta investigación, la evaluación de usabilidad se integra como un componente clave para validar no solo el desempeño técnico del sistema, sino también su aceptación por parte de los usuarios finales.

Capítulo 3

Metodología

En este capítulo se describe el diseño experimental y la arquitectura técnica del entrenador virtual para el aprendizaje de la lengua de señas mexicana (LSM). La metodología se fundamenta en un enfoque cuantitativo y se centra en la transformación de señales de video en trayectorias matemáticas, permitiendo una comparación objetiva entre un modelo de referencia y la ejecución del usuario mediante visión por computadora.

El flujo de trabajo propuesto se divide en cuatro etapas técnico-científicas: 1) Adquisición y curaduría del corpus de datos; 2) Extracción de características anatómicas; 3) Normalización y alineación temporal mediante *Dynamic Time Warping* (DTW); y 4) Evaluación de robustez y usabilidad.

3.1. Constitución y segmentación del corpus de datos

La construcción del corpus de datos representó una de las etapas más críticas de la investigación, debido a la necesidad de contar con modelos de alta fidelidad lingüística y un conjunto de pruebas que permitiera evaluar la robustez del sistema. Para este propósito, se generó un repositorio total de 670 videos, cuya distribución detallada según su función en el proyecto se presenta en la Tabla 3.1.

Tabla 3.1

Distribución del corpus de datos según su función en la investigación

Origen	Función en el proyecto	No. de videos	Lecciones
Intérprete	Modelos de referencia: Integrados en la aplicación para el aprendizaje del usuario.	335	12
Usuario	Conjunto de validación: Utilizados en pruebas para evaluar la robustez del algoritmo.	335	12
Total		670	

La selección de un vocabulario compuesto por exactamente 335 señas distintas responde a un doble criterio de diseño metodológico. Desde una perspectiva pedagógica, esta cantidad permite una cobertura léxica integral y equilibrada de un nivel introductorio y funcional de la Lengua de Señas Mexicana (LSM), distribuido de manera modular en las 12 lecciones estructuradas (abarcando el abecedario dactilológico, el sistema numérico, unidades temporales como días y meses, y vocabulario conversacional básico).

Desde un enfoque computacional y de visión por computadora, un catálogo de 335 señas proporciona una diversidad y variabilidad cinemática idónea para validar el desempeño del sistema. Este volumen de clases garantiza que el algoritmo de alineación temporal (DTW) y la extracción de puntos clave con MediaPipe sean expuestos a un amplio espectro de geometrías anatómicas: desde configuraciones manuales estáticas y bimanuales, hasta trayectorias dinámicas complejas con oclusiones y desplazamientos espaciales significativos. De este modo, se mitiga el riesgo de sobreajuste (*overfitting*) a un léxico reducido y se dota de validez estadística y científica al análisis de robustez del entrenador virtual.

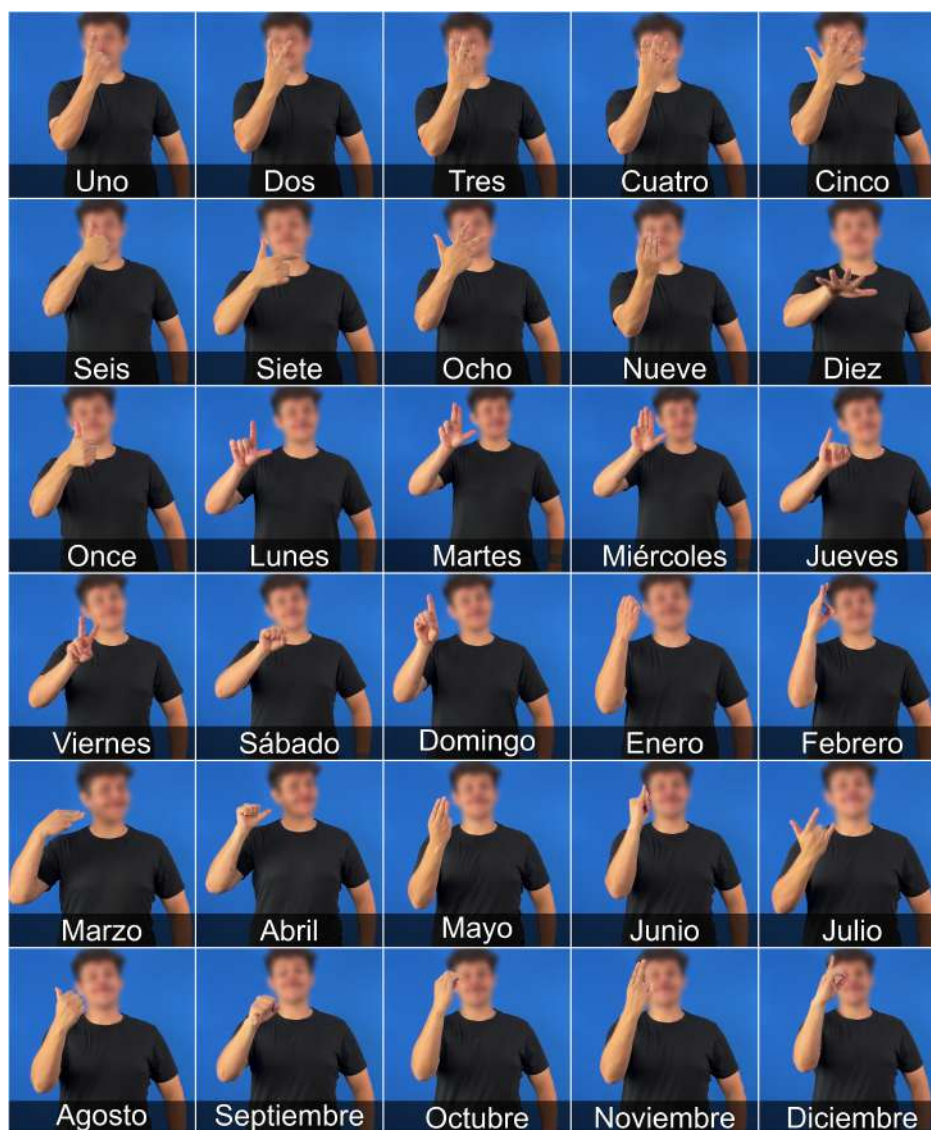
3.1.1. Modelos de referencia (Intérprete)

Para los modelos de ejecución técnica ideal que el sistema utiliza como estándar, se contó con la participación de un intérprete certificado en lengua de señas mexicana. El proceso de captura se realizó bajo condiciones controladas, empleando un fondo uniforme de color azul y vestimenta oscura ajustada para optimizar el contraste visual y facilitar el seguimiento de puntos clave. Estos videos constituyen la base de datos pedagógica que reside permanentemente en la aplicación.

En la Figura 3.1 se presenta una muestra de los modelos de referencia, abarcando categorías como el abecedario, los días de la semana y los meses del año.

Figura 3.1

Muestra del corpus de referencia: estandarización visual de las lecciones de números (1-11), días de la semana y meses del año ejecutadas por el intérprete profesional



3.1.2. Conjunto de validación y análisis de robustez (Usuario)

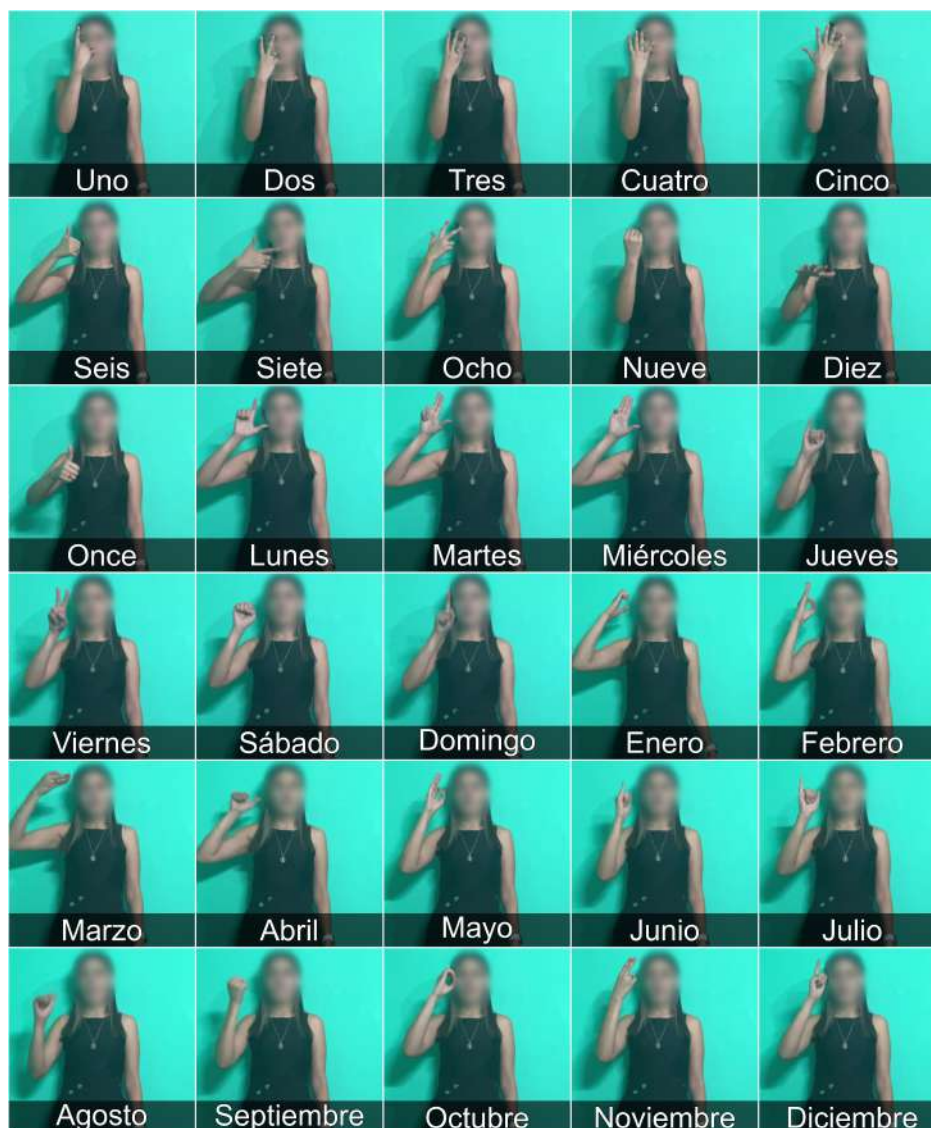
De manera homóloga, se registraron las ejecuciones de un usuario no experto realizando las mismas 335 señas. Estos datos no forman parte de la aplicación; su uso es estrictamente experimental, orientado a evaluar el algoritmo DTW frente a variaciones en velocidad, amplitud y precisión.

El conjunto fue sometido a pruebas de degradación técnica para analizar la sensibilidad del

sistema ante cambios en iluminación y resolución espacial (véase la Figura 3.2).

Figura 3.2

Muestra del conjunto de validación: réplica de las lecciones del corpus por un usuario no experto, utilizada para la evaluación de la robustez algorítmica y el cálculo de distancias DTW



3.1.3. Curaduría manual y optimización de archivos

Para garantizar la precisión de la comparación temporal, se realizó una **curaduría fotograma a fotograma** en los 670 videos. Se identificaron manualmente los puntos de inicio (*onset*) y finalización (*offset*) de cada seña para eliminar pausas irrelevantes. Asimismo, se aplicó un recorte espacial (*cropping*) y se eliminó la pista de audio de los archivos, lo que redujo drásticamente el

peso de los videos sin comprometer la visibilidad de los articuladores manuales y faciales.

3.2. Generación de animaciones a partir de video

Como parte del proceso de desarrollo del sistema, se generaron animaciones a partir de los videos del intérprete con el objetivo de evaluar su posible uso como recurso visual dentro de la aplicación.

Para ello, los videos previamente editados fueron cargados en la plataforma DeepMotion, donde se procesaron para obtener animaciones correspondientes a cada seña. En la Figura 3.3 se muestra la interfaz de la aplicación web utilizada, en la cual se observa un total de 382 videos cargados. Esta cantidad incluye tanto los videos definitivos como aquellos empleados durante pruebas, ajustes y repeticiones del proceso.

Figura 3.3

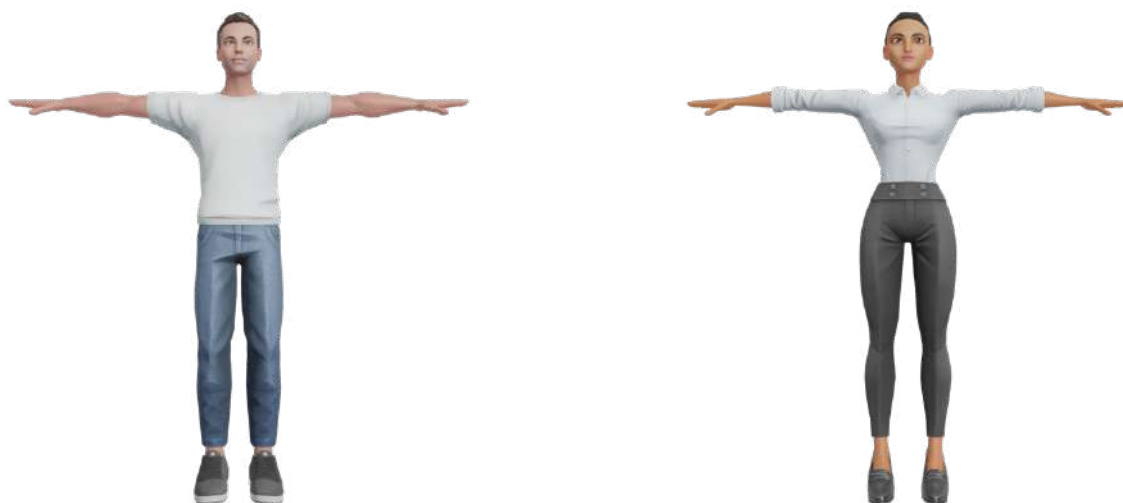
Interfaz de la plataforma DeepMotion mostrando el conjunto de videos cargados



Las animaciones generadas fueron descargadas en formato *.fbx* y almacenadas como parte del repositorio del proyecto. En la Figura 3.4 se presentan ejemplos de los modelos de avatar obtenidos a partir de este proceso.

Figura 3.4

Ejemplos de modelos de avatar generados a partir de los videos del intérprete



Modelo masculino

Modelo femenino

La evaluación de la calidad de estas animaciones y su viabilidad para su uso dentro del sistema se presenta en el capítulo de Resultados.

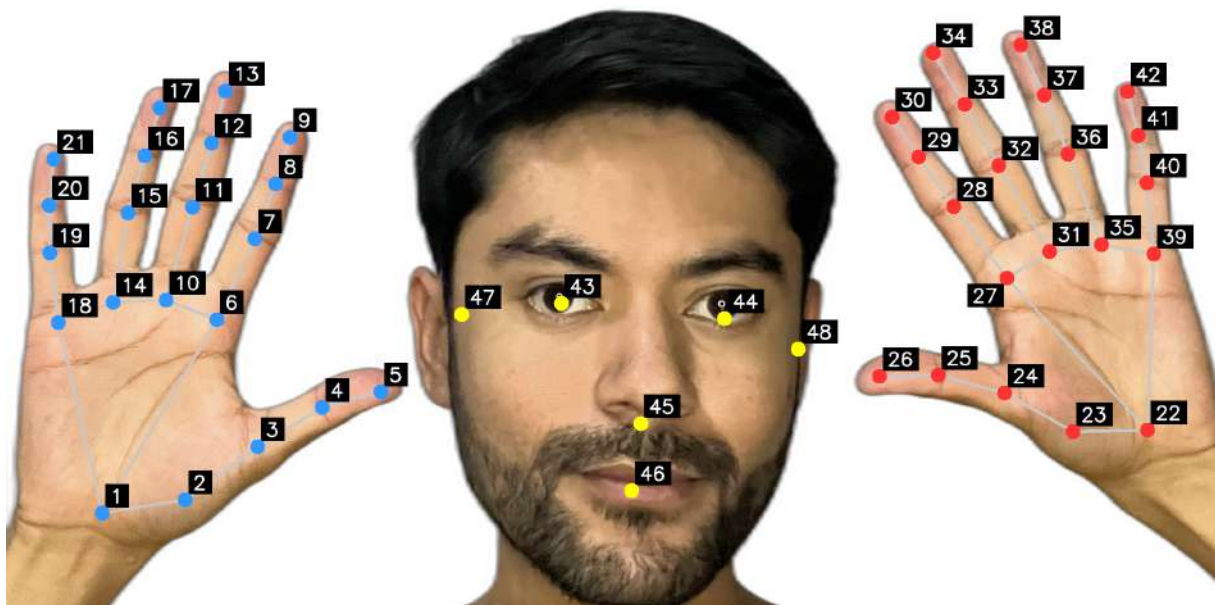
3.3. Extracción de puntos de las manos y la cara

Para la obtención de las características necesarias para el análisis, se utilizó la tecnología MediaPipe Lugaresi et al. (2019), mediante la cual se realizó la extracción de puntos clave en los ejes x e y correspondientes a las manos y la cara en cada uno de los videos.

En total, se extrajeron 21 puntos de la mano derecha, 21 puntos de la mano izquierda y 6 puntos del rostro. Estos puntos fueron obtenidos tanto de los videos del intérprete como de los videos del usuario, y sirven como referencia para la comparación de las señas, tal como se ilustra en la Figura 3.5.

Figura 3.5

Visualización de los puntos clave obtenidos mediante MediaPipe



Nota. Fuente: fotografía tomada por el autor.

Una vez realizada la extracción, los valores de cada punto fueron almacenados en archivos con formato CSV, lo que permitió una visualización y manipulación más eficiente de los datos.

3.3.1. Representación temporal de las señas

Con el propósito de ilustrar la estructura temporal de las señas empleadas en este trabajo, la Figura 3.6 presenta ejemplos de secuencias de imágenes correspondientes a distintas señas de la Lengua de Señas Mexicana. Cada seña se representa mediante cuatro frames consecutivos que permiten observar su ejecución a lo largo del tiempo.

Figura 3.6

Ejemplos ilustrativos de secuencias temporales correspondientes a cuatro señas de la Lengua de Señas Mexicana. Cada seña se muestra como una secuencia de cuatro frames consecutivos (de izquierda a derecha)



¿Qué?



¿Dónde?



Buenos días



¿Cómo estás?

3.4. Normalización y comparación de los puntos

Con el objetivo de hacer comparables las trayectorias de señas ejecutadas por el intérprete y el usuario, independientemente de variaciones debidas al encuadre de la cámara o a la distancia, se aplicó un proceso de normalización por eje basado en el método z-score. Este procedimiento permitió reducir los efectos de traslación y escala.

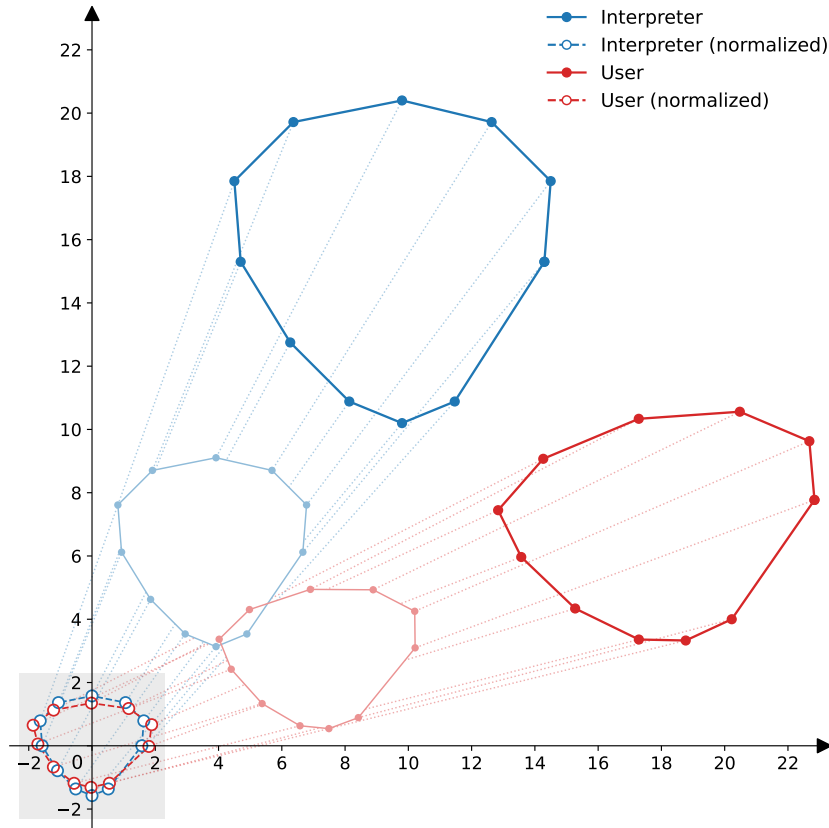
La normalización se realizó de manera independiente para cada punto y para cada eje espacial. Los valores originales en los ejes x e y se transformaron mediante una estandarización de acuerdo con las siguientes expresiones:

$$x' = \frac{x - \mu_x}{\sigma_x}, \quad y' = \frac{y - \mu_y}{\sigma_y} \quad (3.1)$$

Donde x y y representan los valores originales; x' e y' corresponden a los valores normalizados; μ son las medias; y σ representan las desviaciones estándar.

Figura 3.7

Esquema del proceso de normalización: las trayectorias originales se transforman a un espacio normalizado



3.4.1. Condiciones de iluminación y resolución para el análisis de robustez

Con el objetivo de evaluar la robustez del sistema ante variaciones en las condiciones de adquisición, se aplicaron modificaciones controladas en la iluminación y en la resolución espacial de las secuencias de entrada. La Figura 3.8 muestra un ejemplo ilustrativo de las condiciones consideradas durante este análisis.

Figura 3.8

Ejemplo ilustrativo de las condiciones de iluminación y resolución empleadas para el análisis de robustez, utilizando como referencia la ejecución de la seña correspondiente al número 5



Nota. De izquierda a derecha: (a) condición original, (b) reducción moderada de iluminación, (c) reducción severa de iluminación, (d) reducción moderada de resolución, (e) reducción severa de resolución, (f) combinación moderada de iluminación y resolución, y (g) combinación severa de ambas condiciones.

3.5. Comparación cuadro a cuadro mediante distancia euclidiana

Una vez normalizadas las trayectorias, se procedió a calcular la similitud instantánea entre pares de cuadros (frames). Para ello, se empleó la distancia euclidiana.

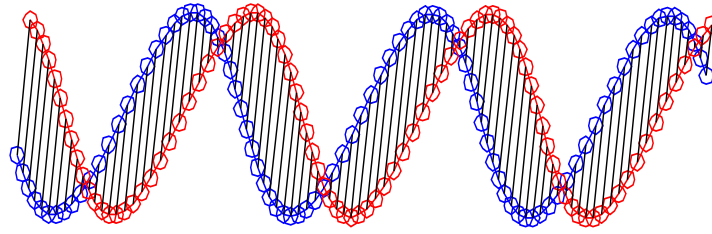
Sea p'_i el vector de puntos normalizados correspondiente al cuadro i de la seña de referencia, y q'_j el vector de puntos normalizados del cuadro j de la seña realizada por el usuario. La distancia euclidiana entre ambos cuadros se define como se indica en la Ecuación 3.2:

$$d(i, j) = \|p'_i - q'_j\|_2 = \sqrt{\sum_{k=1}^{96} (p'_{i,k} - q'_{j,k})^2} \quad (3.2)$$

Donde k recorre el conjunto de 96 coordenadas normalizadas correspondientes a los puntos clave. Esta métrica proporciona una medida de disimilitud instantánea.

Figura 3.9

Cálculo de la distancia euclidiana entre dos puntos homólogos en el espacio normalizado



3.6. Alineación temporal mediante Dynamic Time Warping (DTW)

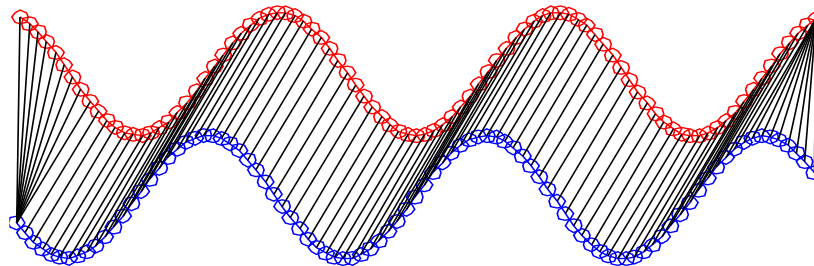
Dado que las señas pueden presentar variaciones en velocidad, se utilizó el algoritmo *Dynamic Time Warping* (DTW), el cual permite alinear temporalmente dos secuencias minimizando el costo acumulado.

A partir de la distancia euclidiana $d(i, j)$ calculada entre cuadros, el costo acumulado de DTW se define de forma recursiva como se muestra en la Ecuación 3.3:

$$DTW[i, j] = d(i, j) + \min \begin{cases} DTW[i - 1, j] \\ DTW[i, j - 1] \\ DTW[i - 1, j - 1] \end{cases} \quad (3.3)$$

Figura 3.10

Ejemplo del alineamiento de dos secuencias de señas utilizando el algoritmo Dynamic Time Warping (DTW)



3.7. Definición del umbral de aceptación basado en DTW

Para establecer un criterio práctico que permitiera determinar si una seña fue ejecutada correctamente, se analizó la distribución de las distancias DTW obtenidas en todas las lecciones del sistema ($N = 335$). Se identificó que las muestras del abecedario presentaban valores más altos debido al cambio de mano del intérprete, por lo que fueron excluidas.

Con base en la distribución empírica, se definió un umbral global de aceptación de $DTW \leq 600$. Adicionalmente, se realizó un análisis de sensibilidad con umbrales de 500, 600 y 700 unidades DTW, confirmando que 600 constituye un criterio robusto.

3.8. Evaluación de robustez ante ruido

Con el objetivo de evaluar la sensibilidad del sistema, se generaron versiones perturbadas de las secuencias originales mediante la adición de ruido gaussiano a las coordenadas (x,y) , con valores de desviación estándar $\sigma \in \{0,5, 1, 2, 3\}$. Los efectos de estas perturbaciones sobre las distancias DTW fueron analizados y los resultados se presentan en la sección de Resultados.

3.9. Comparación de los puntos normalizados

Se llevó a cabo un análisis comparativo entre los puntos previamente normalizados empleando DTW. La técnica de DTW se basa en encontrar una correspondencia óptima entre dos secuencias temporales, minimizando la distancia total entre ellas.

La implementación de DTW en este contexto permitió:

1. Identificar y cuantificar las discrepancias entre los movimientos realizados por el usuario y los del intérprete.
2. Compensar diferencias temporales y espaciales.
3. Generar métricas de similitud para ofrecer retroalimentación.

Capítulo 4

Resultados

En este capítulo se presentan los resultados obtenidos a partir de la comparación entre las señas realizadas por un intérprete profesional de lengua de señas mexicana y las ejecutadas por los usuarios mediante la aplicación desarrollada. Para la evaluación de la similitud entre ambas ejecuciones se empleó el algoritmo Dynamic Time Warping (DTW), el cual permite comparar secuencias temporales considerando variaciones en la duración y velocidad de ejecución.

La distancia DTW proporciona una medida cuantitativa de la similitud entre los movimientos de las señas evaluadas. Valores cercanos a cero indican una alta similitud entre las trayectorias comparadas, mientras que valores elevados reflejan diferencias significativas en la ejecución. Esta métrica permite evaluar de manera objetiva el desempeño del usuario, identificar discrepancias en el gesto y analizar la consistencia de la ejecución bajo distintas condiciones.

A partir de esta métrica, los resultados se analizaron mediante visualizaciones gráficas, comparaciones por vocabulario y lecciones, así como pruebas de robustez ante condiciones adversas y validación con usuarios reales, las cuales se describen en las secciones siguientes.

4.1. Evaluación de animaciones generadas para las señas

Se generaron animaciones a partir de los videos del intérprete utilizando la herramienta DeepMotion, obteniendo archivos en formato *.fbx* para cada una de las señas del corpus, así como algunos casos adicionales destinados a pruebas.

Durante el proceso de generación y revisión, se realizó una comparación visual directa entre las animaciones obtenidas y los videos originales del intérprete. Este análisis permitió identificar diferencias significativas en la ejecución de las señas.

En particular, los movimientos de las manos en las animaciones no correspondían con precisión a los del video original, presentando trayectorias poco similares. Asimismo, se observaron problemas en la fluidez del movimiento, donde en varios casos las manos parecían detenerse o

comportarse de manera poco natural durante la ejecución. De igual forma, la configuración de los dedos no coincidía con la del intérprete, mostrando posiciones incorrectas o poco definidas.

Dado que en la Lengua de Señas Mexicana la precisión en la forma y movimiento de las manos es fundamental, estas discrepancias afectan directamente la claridad y correcta interpretación de las señas.

Con base en estos resultados, se determinó que las animaciones generadas no alcanzan el nivel de fidelidad necesario para su uso dentro del sistema. Por esta razón, se optó por utilizar directamente los videos del intérprete como referencia principal para la enseñanza de las señas.

Finalmente, los archivos generados en formato *.fbx* fueron conservados como parte del repositorio del proyecto, ya que representan un recurso potencial para futuras líneas de investigación orientadas a la mejora o corrección de este tipo de animaciones.

4.2. Comparación de trayectorias mediante DTW y distancia euclidiana

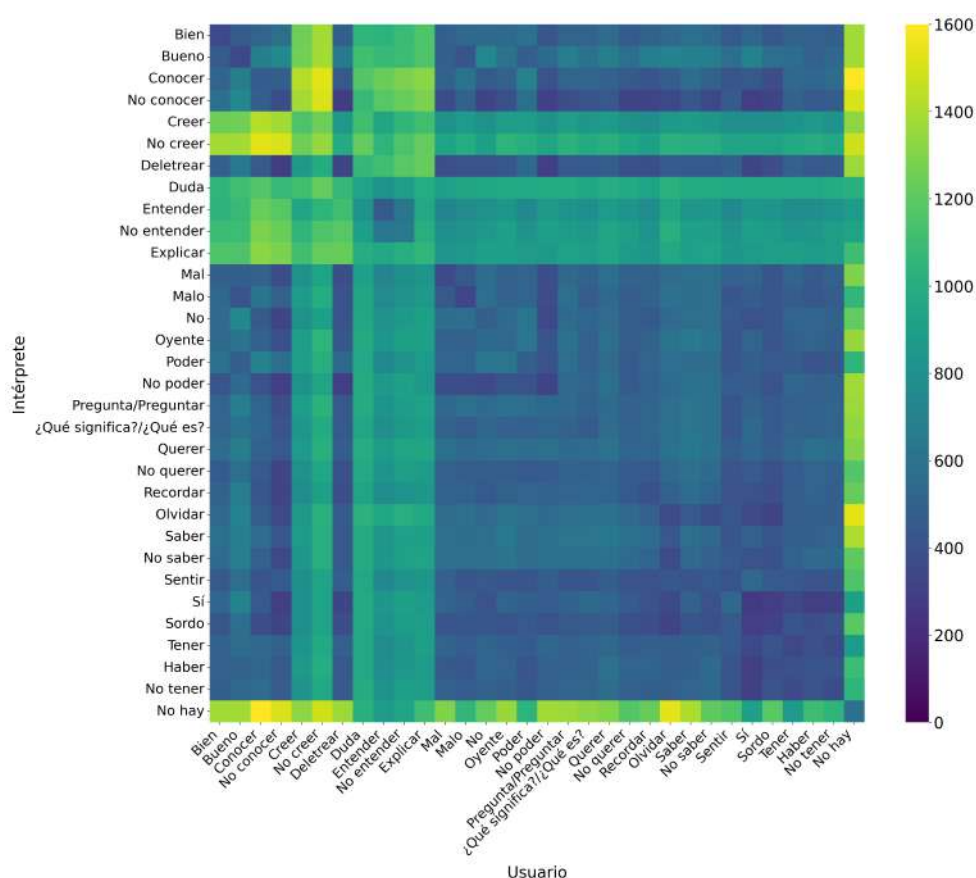
4.2.1. Análisis mediante mapas de calor con DTW

Con el objetivo de visualizar de manera detallada las similitudes y diferencias entre las trayectorias del intérprete y del usuario, se generaron mapas de calor a partir de las distancias DTW calculadas entre los puntos clave de ambas manos y del rostro.

En la Figura 4.1 se presenta un mapa de calor que muestra la distancia DTW obtenida al comparar directamente todas las secuencias de puntos del intérprete y del usuario. Esta representación permite identificar regiones temporales donde se concentran mayores discrepancias, así como zonas donde la ejecución del usuario presenta mayor similitud con la referencia profesional.

Figura 4.1

Mapa de calor de la distancia DTW entre los puntos de las manos y el rostro del intérprete y el usuario



Durante las pruebas se observó que los usuarios, al no ser intérpretes profesionales, suelen tardar más tiempo en ejecutar una misma seña en comparación con el intérprete de referencia, quien realiza los gestos de manera más fluida y rápida. Como consecuencia, los videos del usuario contienen un mayor número de fotogramas que los del intérprete.

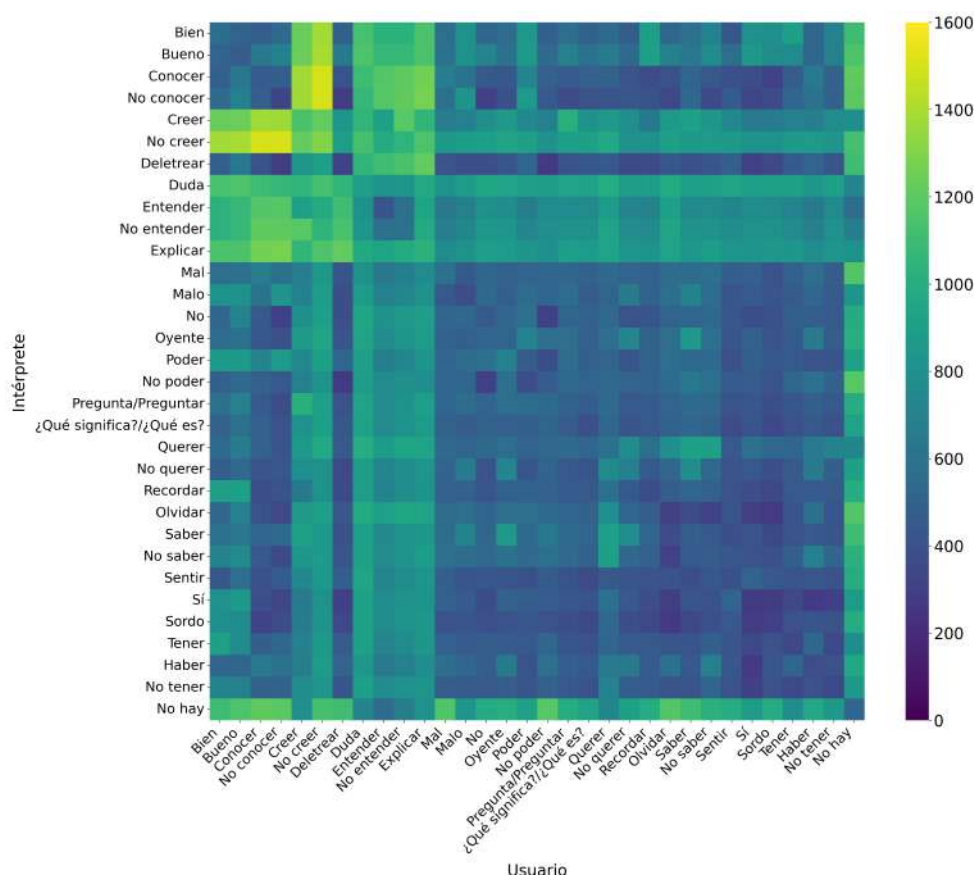
Con el objetivo de realizar una comparación visual más equilibrada en los mapas de calor, se llevó a cabo un ajuste temporal seleccionando únicamente los fotogramas centrales de ambas secuencias. En este proceso, la cantidad de frames considerados en el video del usuario se redujo para igualarla a la del intérprete, preservando el núcleo del movimiento de la seña y evitando incluir pausas iniciales o finales.

En la Figura 4.2 se muestra el mapa de calor obtenido tras esta normalización temporal, la cual se utilizó exclusivamente con fines de visualización y comparación gráfica. Cabe destacar que el cálculo de la distancia DTW se mantiene como la métrica principal del sistema, ya que este

algoritmo permite manejar de forma natural diferencias en la duración de las secuencias sin requerir recortes estrictos.

Figura 4.2

Mapa de calor de la distancia DTW considerando únicamente los fotogramas centrales de las secuencias del intérprete y del usuario, con el fin de igualar la cantidad de frames para la comparación visual



4.2.2. Comparación con distancia euclidiana

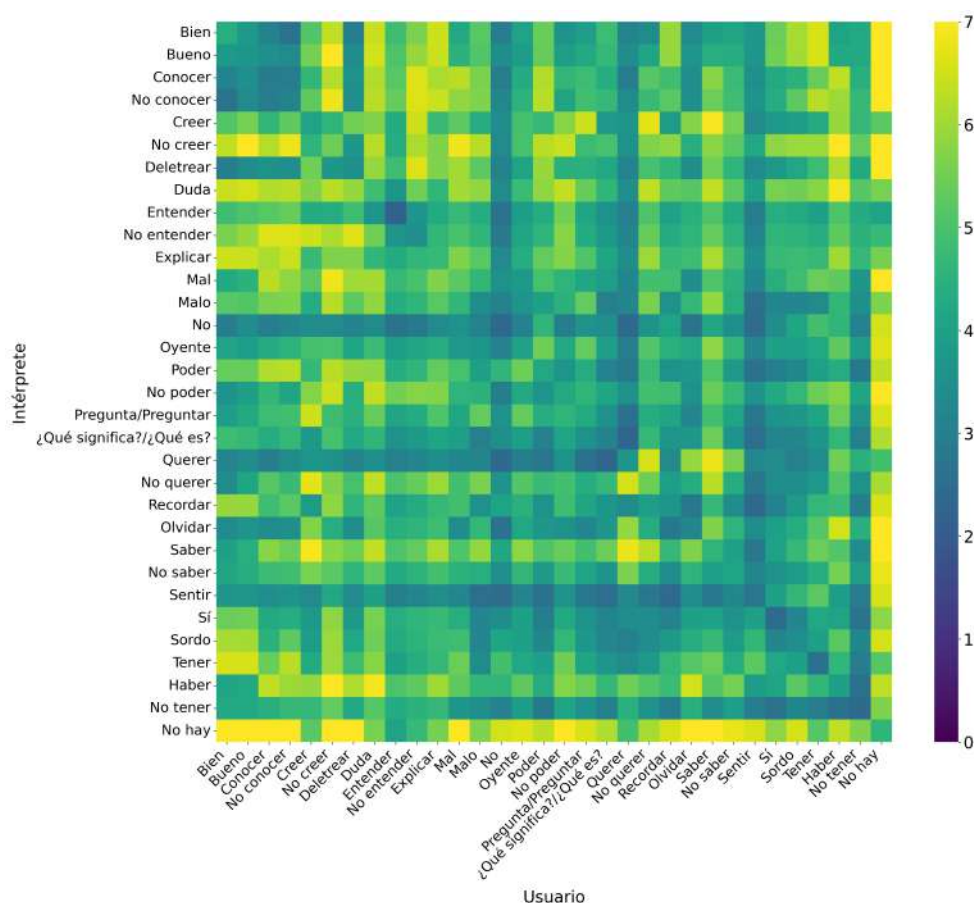
Con el propósito de contrastar el desempeño del algoritmo DTW con una métrica más simple, se realizó una comparación adicional utilizando la distancia euclidiana. Al igual que en el caso anterior, se omitieron los fotogramas iniciales y finales para igualar la longitud de las secuencias comparadas.

En la Figura 4.3 se presenta el mapa de calor correspondiente a la distancia euclidiana. A diferencia del DTW, esta métrica compara directamente los puntos en cada fotograma sin considerar desalineamientos temporales, lo que limita su capacidad para evaluar señas ejecutadas a distintas

velocidades. Como consecuencia de su formulación matemática, los valores obtenidos mediante distancia euclidiana se encuentran en un rango numérico considerablemente menor que los del DTW, lo cual explica que la escala del mapa de calor difiera respecto a las visualizaciones basadas en DTW. Este resultado refuerza la pertinencia del uso de DTW como métrica principal en el sistema propuesto.

Figura 4.3

Mapa de calor de la distancia euclidiana entre los puntos de las manos y el rostro del intérprete y el usuario, mostrando una escala de valores menor en comparación con DTW debido a la naturaleza punto a punto de esta métrica



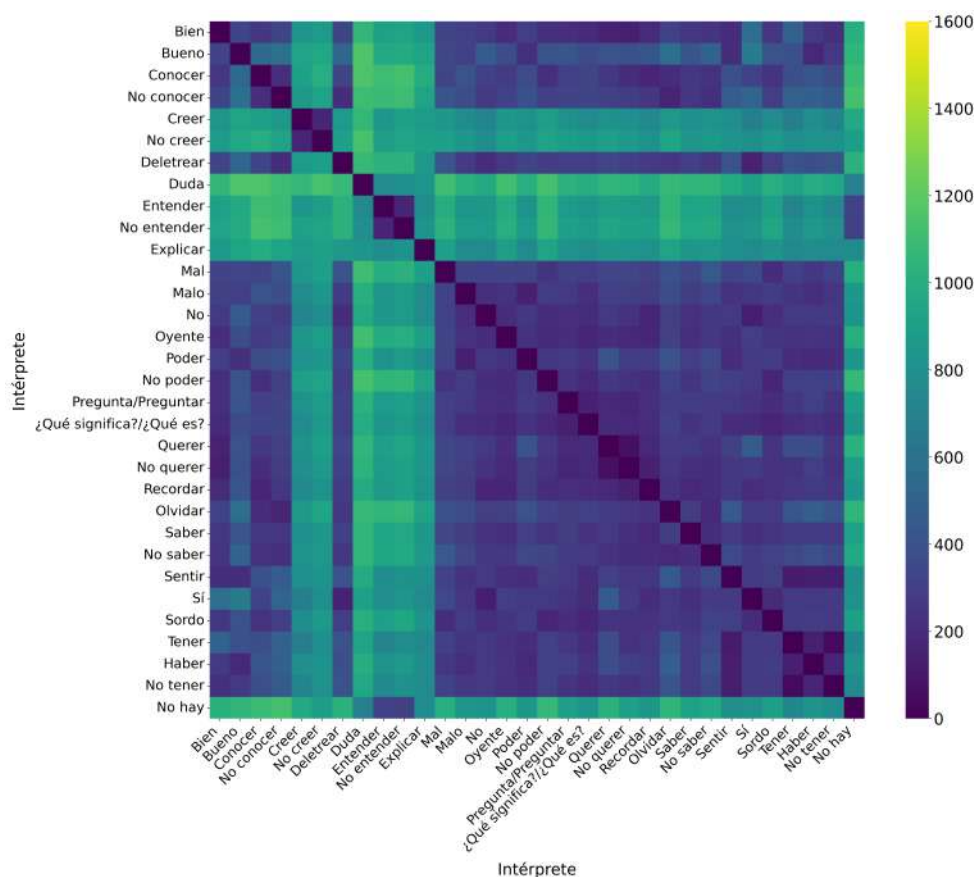
Con el fin de verificar el correcto funcionamiento del algoritmo DTW implementado, se realizó una comparación entre las secuencias de video del intérprete contra sí mismas, considerando cada seña dentro del vocabulario y las distintas lecciones definidas en el corpus. En este análisis, cada ejecución del intérprete fue comparada únicamente con su correspondiente secuencia original, lo que permite evaluar la consistencia interna del método.

En la Figura 4.4 se presenta el mapa de calor resultante de estas comparaciones. Como se obser-

va, la matriz muestra una diagonal principal claramente definida con valores cercanos a cero, lo cual indica que la distancia DTW entre una seña y ella misma es mínima. Este comportamiento es el esperado teóricamente y confirma que el algoritmo responde de manera adecuada cuando se comparan secuencias idénticas.

Figura 4.4

Mapa de calor de la distancia DTW obtenido a partir de la comparación de las secuencias del intérprete consigo mismas, donde la diagonal principal indica distancia cero al comparar cada seña con su correspondiente secuencia original



4.3. Desempeño del algoritmo DTW por lecciones

Los experimentos con Dynamic Time Warping (DTW) permitieron comparar las ejecuciones de los usuarios contra las secuencias de referencia realizadas por intérpretes profesionales. Una distancia DTW baja indica una alta similitud en la trayectoria y sincronización de los movimientos, mientras que valores elevados reflejan discrepancias en la forma de ejecución o en la alineación temporal.

En la lección del abecedario, las distancias DTW fueron consistentemente superiores a 1000. Este comportamiento se atribuye principalmente a la diferencia de mano dominante entre el intérprete (mano izquierda) y los usuarios (mano derecha). En contraste, en las lecciones de vocabulario básico, preguntas y frases cortas, la mayoría de las distancias DTW se concentraron en el rango de 300 a 600.

La Tabla 4.1 presenta la media y la desviación estándar de las distancias DTW por lección.

Tabla 4.1

Media y desviación estándar de las distancias DTW por lección, obtenidas a partir de la comparación entre las ejecuciones del intérprete y de los usuarios

Lección	DTW medio	Desv. estándar
Abecedario	1419.5	169.4
Vocabulario básico	518.3	255.6
Preguntas	517.4	148.1
Cortesía y saludos	610.4	278.0
Números	390.4	200.1
Adverbios de tiempo	510.6	188.8
Días de la semana	271.5	34.3
Meses	244.4	50.6
Parentescos y relaciones	517.6	217.6
Conectores / preposiciones	556.7	216.0
Emociones	483.9	169.7
Verbos comunes (parte 1)	635.3	238.1

4.3.1. Optimización del almacenamiento de los recursos multimedia

Como parte de la evaluación del sistema y con el objetivo de mejorar la eficiencia en su implementación, se llevó a cabo un proceso de optimización sobre los videos correspondientes al intérprete de Lengua de Señas Mexicana (LSM). Estos recursos constituyen un componente fundamental del sistema, ya que son utilizados tanto en la visualización de las señas como en el análisis de las trayectorias mediante el algoritmo DTW. En consecuencia, su tamaño impacta directamente en el consumo de memoria, los tiempos de carga y la viabilidad de despliegue del sistema.

Las capturas originales, almacenadas en formato .MOV, incluían pistas de audio y metadatos

inherentes al proceso de grabación que no aportan valor al análisis ni a la funcionalidad del sistema. Por ello, se diseñó un procedimiento de limpieza y estandarización orientado a reducir el volumen de almacenamiento sin comprometer la integridad de la información visual.

Para este propósito, se implementó un proceso automatizado mediante la herramienta FFmpeg. Utilizando el parámetro `-c:v copy`, se extrajo exclusivamente el flujo de video sin recodificación y se encapsuló en el formato .mp4, eliminando las pistas de audio y los metadatos. Este procedimiento garantiza la preservación íntegra de la información espacio-temporal de los movimientos del intérprete, asegurando que el algoritmo DTW opere sobre secuencias temporales fieles a la captura original.

En la Tabla 4.2 se presenta una comparación entre el tamaño de los archivos originales y los procesados, tanto a nivel individual como a nivel global del corpus.

Tabla 4.2

Comparación de métricas de almacenamiento del corpus de LSM ($N = 335$)

Métrica	Formato original (.MOV)	Formato procesado (.mp4)
Tamaño promedio por seña	23.93 MB	1.91 MB
Tamaño total del corpus	8017.95 MB	639.35 MB

A partir de los resultados obtenidos, se observa una reducción significativa en el volumen de almacenamiento del corpus. El tamaño total disminuyó de 8017.95 MB a 639.35 MB, lo que representa una optimización global del 92.03 %. De manera consistente, el tamaño promedio por seña se redujo de 23.93 MB a 1.91 MB. Esta optimización resulta crítica para la implementación del sistema, ya que reduce el consumo de memoria, mejora los tiempos de lectura de los archivos y favorece la eficiencia general del sistema, sin afectar la calidad visual ni la precisión en el análisis de las trayectorias.

4.4. Evaluación del desempeño computacional del método DTW

Con el objetivo de cuantificar los requerimientos computacionales del método propuesto basado en DTW, se realizó un experimento de desempeño utilizando una seña representativa correspondiente a la Lección 10 (Sign_0011). Las pruebas se ejecutaron en un equipo con sistema operativo Windows 11, procesador Intel Core i7 y 32 GB de memoria RAM, sin el uso de aceleración por GPU.

La Tabla 4.3 presenta el tiempo de ejecución y el consumo de memoria para cada una de las etapas del proceso.

Tabla 4.3

Evaluación del desempeño computacional (Tiempo y Memoria)

Stage	Description	Time (ms)	Δ RSS (MB)	Peak Total (MB)
Keypoint (interpreter)	96 features per frame using MediaPipe	8551.73	+60.04	291.25
Keypoint (user)	Same process applied to user video	4417.78	+0.70	232.88
DTW computation	Temporal-spatial trajectory comparison	285.76	+0.04	91.94

Las matrices extraídas contienen 96 coordenadas normalizadas por fotograma. Los resultados muestran que el principal cuello de botella computacional se encuentra en la etapa de extracción de puntos clave. En contraste, el cálculo de DTW presenta un costo computacional reducido.

4.5. Análisis de robustez del método DTW

Con el fin de evaluar la robustez del método DTW ante variaciones en las condiciones visuales, se realizaron experimentos controlados que simulan degradaciones comunes en escenarios reales de uso.

4.5.1. Efecto de variaciones en iluminación y resolución

Para este experimento se generaron variantes de los videos del usuario mediante la reducción de brillo (Light50, Light25) y resolución espacial (Res50, Res25). Como se observa en la Tabla 4.4, los resultados indican que degradaciones moderadas en iluminación o resolución producen variaciones menores al 5 % en la distancia DTW. En contraste, la condición más severa (Light25_Res25) genera un incremento notable de aproximadamente 37.5 %.

Tabla 4.4*Resultados del análisis de robustez: Impacto de las variaciones de iluminación y resolución*

Condición del Video del Usuario	Promedio DTW	Desv. Est. ($\pm$)	Variación vs Base
Base (Condición Original)	583.02	335.72	0.00 %
Reducción de luz (50 %)	563.30	341.49	-3.38 %
Reducción de luz (25 %)	583.53	334.91	+0.09 %
Reducción de resolución (50 %)	582.56	336.06	-0.08 %
Reducción de resolución (25 %)	606.85	332.78	+4.09 %
Luz 50 % + Res. 50 %	571.50	339.54	-1.98 %
Luz 25 % + Res. 25 %	802.13	339.52	+37.58 %

4.5.2. Efecto de oclusiones temporales (Pérdida de rastreo)

En un segundo experimento de robustez, se evaluó el impacto de oclusiones temporales. Para evaluar la robustez del cálculo de la distancia DTW ante fallos en la estimación de pose, se implementó un algoritmo que simula la pérdida temporal de puntos clave (*keypoints*) generados por MediaPipe. En lugar de aplicar oclusiones espaciales sobre los píxeles del video, se intervinieron directamente las matrices de coordenadas extraídas, forzando los valores a cero para emular la incapacidad del modelo de detectar la zona anatómica.

Las degradaciones se generaron combinando tres variables paramétricas:

1. **Región Anatómica Afectada:** Se anularon selectivamente los puntos correspondientes a la mano izquierda (*L*), mano dominante derecha (*R*), ambas manos (*B*), los puntos faciales (*Face*) y la totalidad de los puntos (*All*).
2. **Severidad de la pérdida (%):** Se estableció la duración de la pérdida de información correspondiente al **10 %**, **30 %** y **60 %** del total de los cuadros (*frames*) del video.
3. **Distribución Temporal:**
 - **Continua (c):** Los cuadros afectados se agruparon en un bloque único y continuo situado en la región central del video. Esto simula una oclusión física prolongada o un bloqueo temporal del campo de visión.
 - **Aleatoria (r - Random):** Los cuadros afectados se distribuyeron estocásticamente a lo largo de todo el video, simulando inestabilidad en el algoritmo de rastreo (*flicke-*

ring), caídas severas de *framerate* o pérdida de paquetes en una transmisión de baja calidad.

Bajo esta nomenclatura, una variable algorítmica como Comp_occR30c indica la pérdida total del rastreo de la mano derecha en un bloque continuo equivalente al 30 % de la duración del video, mientras que Comp_occFace10r indica una pérdida aleatoria de los puntos faciales distribuida en el 10 % de los cuadros.

Los resultados cuantitativos de este experimento se consolidan en la Tabla 4.5, mostrando un incremento progresivo de la distancia DTW conforme aumenta el nivel de oclusión.

Tabla 4.5

Resultados del análisis de robustez: Impacto de oclusiones temporales (pérdida de rastreo) en la distancia DTW

Condición de Oclusión	Promedio DTW (c / r)	Desv. Est. ($\pm$) (c / r)	Incremento vs Base (c / r)
Base (Condición Original)	Promedio: 583.02 Desv. Est: 335.72 Incremento: 0.00 %		
Mano Izquierda (10 %)	569.21 / 623.43	311.23 / 335.78	-2.37 % / +6.93 %
Mano Izquierda (30 %)	619.66 / 688.16	290.41 / 318.91	+6.28 % / +18.03 %
Mano Izquierda (60 %)	685.20 / 728.43	295.42 / 316.32	+17.53 % / +24.94 %
Mano Derecha (10 %)	648.74 / 680.70	308.57 / 314.51	+11.27 % / +16.75 %
Mano Derecha (30 %)	827.87 / 880.05	248.33 / 276.01	+41.99 % / +50.94 %
Mano Derecha (60 %)	1012.70 / 1001.57	232.35 / 281.40	+73.69 % / +71.79 %
Ambas Manos (10 %)	651.38 / 685.14	295.95 / 321.02	+11.72 % / +17.51 %
Ambas Manos (30 %)	822.07 / 890.02	236.37 / 297.08	+41.00 % / +52.65 %
Ambas Manos (60 %)	1098.20 / 1172.14	230.41 / 294.67	+88.36 % / +101.04 %
Rostro (10 %)	628.82 / 666.52	322.45 / 327.14	+7.85 % / +14.32 %
Rostro (30 %)	739.29 / 832.29	294.41 / 314.37	+26.80 % / +42.75 %
Rostro (60 %)	953.87 / 1056.81	282.08 / 319.73	+63.60 % / +81.26 %
Área Completa (10 %)	672.70 / 722.23	290.82 / 319.42	+15.38 % / +23.87 %
Área Completa (30 %)	872.84 / 997.62	226.00 / 297.92	+49.71 % / +71.11 %
Área Completa (60 %)	1192.55 / 1375.00	249.30 / 324.97	+104.54 % / +135.84 %

Nota. Los valores separados por una diagonal (/) corresponden a las métricas obtenidas utilizando los patrones de pérdida continua *c* y aleatoria *r*, respectivamente. La métrica “Incremento vs Base” representa la dificultad adicional del algoritmo para emparejar la seña ante la falta de puntos clave.

Como se puede observar en los resultados de la Tabla 4.5, existe una relación directamente proporcional entre la severidad de la degradación de los datos y el incremento en la métrica DTW.

En el caso específico de las oclusiones (pérdida temporal de rastreo), el incremento en la distancia no representa una falla del software, sino el comportamiento matemático esperado del algoritmo. Al anular las coordenadas no detectadas (forzándolas a cero), el DTW calcula la diferencia espacial entre el movimiento real y fluido del intérprete contra los vacíos de información del usuario. Esta discrepancia genera una penalización acumulativa que eleva el valor final de la distancia.

A pesar de este aumento progresivo, los datos demuestran la robustez estructural del método propuesto. El sistema no colapsa ni interrumpe su ejecución ante la falta parcial o severa de fotogramas; por el contrario, el algoritmo DTW logra encontrar la ruta de alineación óptima (*warping path*) basándose en la información espacial y temporal residual.

Resulta particular destacar el comportamiento entre los distintos patrones de pérdida. Las interrupciones estocásticas (patrón *r*, que simulan *flickering* o pérdida intermitente de paquetes) tienden a incrementar la distancia DTW de forma más severa que las oclusiones en un bloque continuo (patrón *c*). Esto ocurre porque el ruido intermitente fragmenta constantemente la trayectoria de la seña, obligando al algoritmo a reajustar la alineación repetidas veces. En contraparte, un bloque continuo de pérdida permite que el DTW empareje de forma más estable los segmentos limpios que ocurren antes y después de la oclusión.

Análisis de impacto por mano dominante vs. mano pasiva

Al desglosar los resultados por región anatómica, se revela un comportamiento asimétrico que resulta fundamental en el contexto de la Lengua de Señas Mexicana (LSM). Los datos indican que el sistema es altamente sensible a la oclusión de la mano derecha (considerada la mano dominante para los usuarios de la muestra), alcanzando un incremento del 73.69 % en la distancia DTW bajo condiciones de pérdida severa (R60c).

En contraste, la oclusión de la mano izquierda (mano pasiva en muchas de las señas evaluadas) genera un impacto considerablemente menor, llegando apenas a un incremento máximo del 24.94 % en el peor de los escenarios. Resulta de especial interés el caso de la oclusión continua leve del 10 % en la mano izquierda (L10c), donde la distancia DTW no solo no aumentó, sino que experimentó una ligera mejoría respecto a la condición base (-2.37 %).

Este fenómeno demuestra empíricamente que, en las lecciones analizadas (principalmente de vocabulario básico y abecedario), la mano no dominante aporta una cantidad mínima de información dinámica para el reconocimiento de la seña. Por el contrario, su rastreo puede introducir un ligero “ruido” algorítmico cuando se encuentra en reposo. En términos prácticos para el en-

torno de aprendizaje, esto confirma que el sistema evalúa con mayor rigor la trayectoria de la extremidad que articula la seña principal, tolerando en mayor medida los fallos de cámara o de rastreo en la extremidad inactiva.

4.6. Evaluación del sistema en entornos reales

Con el objetivo de evaluar el desempeño del algoritmo DTW en condiciones reales de uso, se diseñó y desarrolló una aplicación interactiva en Python que permite a los usuarios practicar señas mediante el uso de una cámara web.

La integración de esta aplicación permitió que el modelo propuesto trascendiera una evaluación puramente teórica, validándose a través de la interacción directa con usuarios finales. No obstante, durante la fase de validación se identificaron factores que afectaron la precisión, como la calidad de las cámaras web y la iluminación no controlada.

Para poner a prueba el sistema en escenarios reales y heterogéneos, se adoptó una estrategia de validación mixta (instalación en equipos personales y pruebas presenciales supervisadas).

4.7. Funcionamiento de la aplicación desarrollada

Antes de detallar las interfaces individuales, es pertinente describir el flujo general de operación del sistema, el cual integra los componentes analizados en las secciones previas. El ciclo de interacción comienza cuando el usuario selecciona una lección y visualiza el modelo de referencia ejecutado por el intérprete. Posteriormente, el usuario realiza la seña frente a la cámara web, momento en el cual el sistema captura el video en tiempo real y utiliza el *framework* MediaPipe para extraer las coordenadas espaciales de los puntos clave de las manos y el rostro.

Una vez extraídas, estas trayectorias son normalizadas para compensar posibles diferencias de escala y traslación derivadas de la posición del usuario. A continuación, los datos normalizados son evaluados mediante el algoritmo DTW contra las secuencias temporales del modelo ideal, lo que permite calcular la distancia métrica entre ambas ejecuciones. Finalmente, el sistema traduce esta distancia en un porcentaje de similitud y presenta una retroalimentación visual inmediata, completando así el ciclo de aprendizaje asistido propuesto.

4.7.1. Menú principal de la aplicación

Al ejecutar la aplicación, el usuario es recibido por una interfaz gráfica intuitiva que presenta el menú principal, donde se listan las distintas lecciones disponibles (véase Figura 4.5).

Figura 4.5

Menú principal de la aplicación

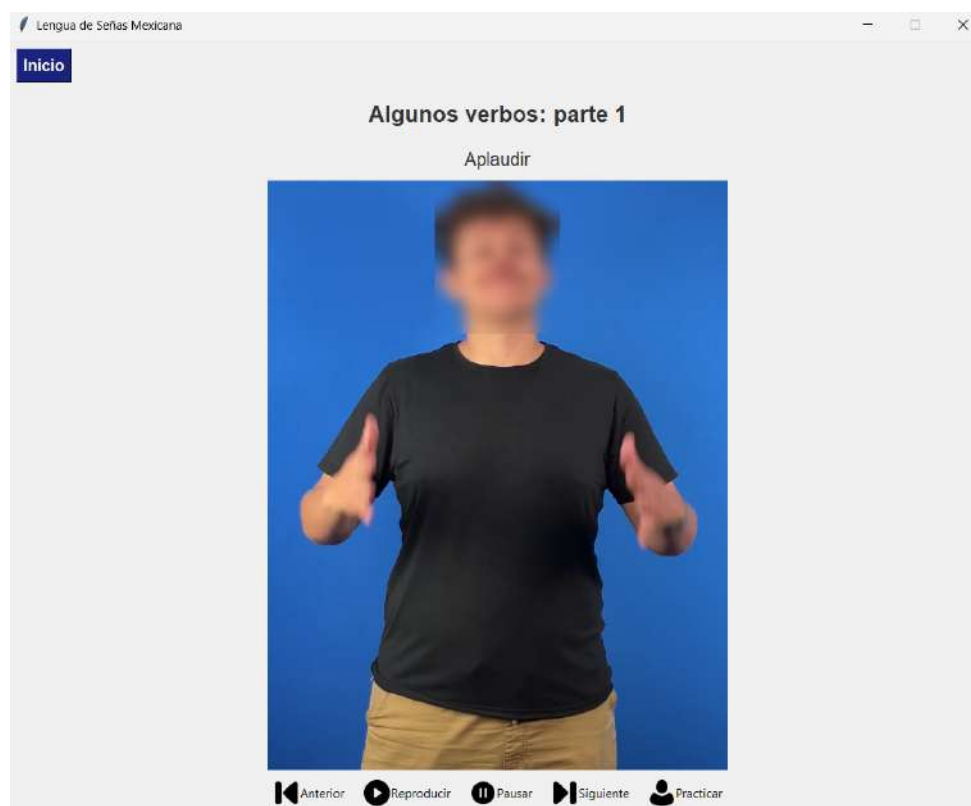


4.7.2. Selección de una lección con señas

Al seleccionar una lección, se abre una ventana que permite al usuario explorar y reproducir las señas asociadas. Esta disposición favorece un aprendizaje visual, gradual y autónomo (véase Figura 4.6).

Figura 4.6

Ventana de visualización de una lección con controles de navegación y práctica

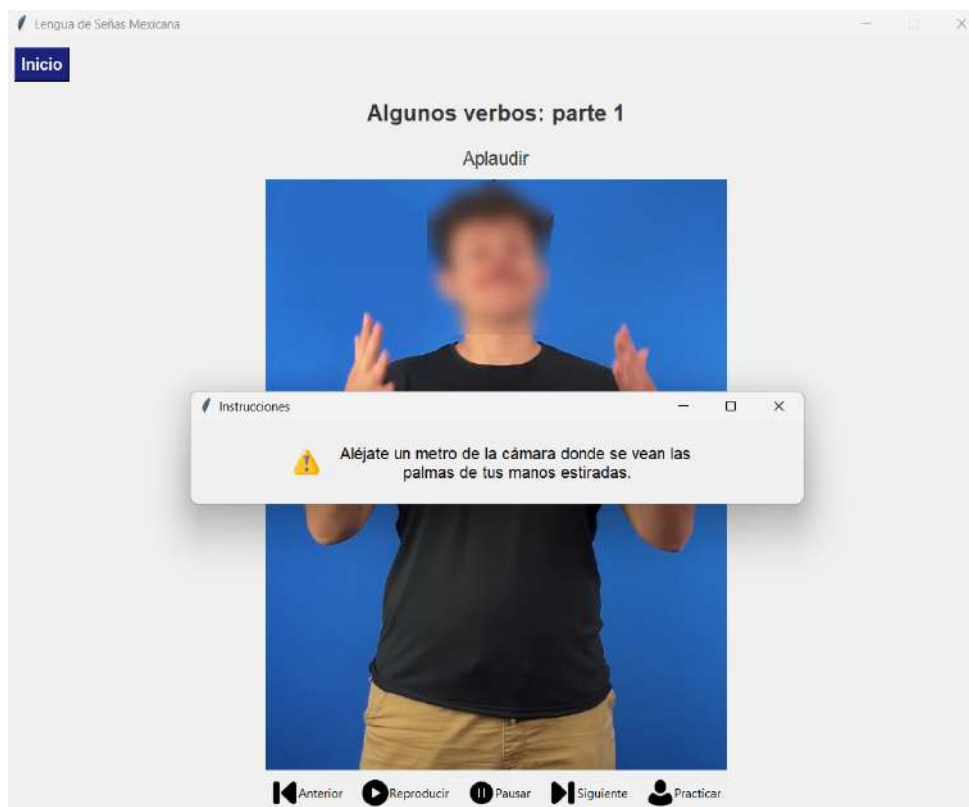


4.7.3. Captura de señas

Una vez que el usuario ha observado el video demostrativo, puede iniciar el proceso de práctica. Antes de comenzar la grabación, el sistema muestra un mensaje con instrucciones claras para facilitar una detección óptima de los puntos clave (véase Figura 4.7).

Figura 4.7

Mensaje de preparación mostrado antes de iniciar la práctica de una seña



4.7.4. Visualización de puntos clave

Una vez iniciada la práctica, el sistema comienza la captura en tiempo real de los puntos clave anatómicos. En la Figura 4.8 se observa la visualización de este proceso.

Figura 4.8

Visualización de los puntos clave durante la ejecución de una seña

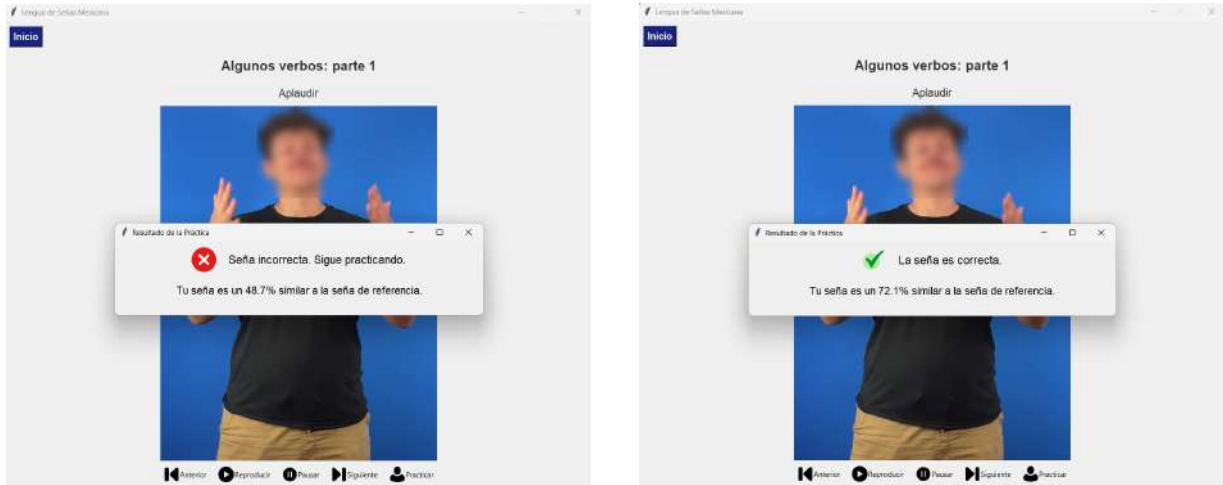


4.7.5. Retroalimentación

Cuando el valor de la distancia DTW supera un umbral establecido de 600, se muestra un mensaje indicando que la seña no fue reconocida correctamente (véase Figura 4.9a). En caso contrario ($DTW \leq 600$), se despliega un mensaje de retroalimentación positiva (véase Figura 4.9b).

Figura 4.9

Mensajes de retroalimentación del sistema basados en el umbral de distancia DTW



(a) Retroalimentación negativa ($DTW > 600$)

(b) Retroalimentación positiva ($DTW \leq 600$)

Nota. El umbral de 600 unidades fue determinado tras las pruebas de validación con usuarios.

4.8. Justificación del umbral de aceptación para la distancia DTW

Para determinar si una seña realizada por el usuario coincide con la versión de referencia, se analizó la distribución de las distancias DTW obtenidas en todas las lecciones ($N = 335$). Se observó que el valor $DTW \leq 600$ se ubica entre la mediana y el percentil 75, y se aproxima a la media global del conjunto de datos (≈ 583).

Con base en estos resultados, se adoptó un umbral global de aceptación de $DTW \leq 600$. Asimismo, se realizó un análisis de sensibilidad aplicando diferentes umbrales (500, 600 y 700), obteniéndose tasas de aceptación del 55.8 %, 69.5 % y 80.5 %, respectivamente.

4.8.1. Transformación de la distancia DTW a porcentaje de similitud

Si bien el algoritmo *Dynamic Time Warping* (DTW) proporciona una métrica matemática robusta para evaluar la diferencia temporal y espacial entre dos secuencias de movimiento, el valor crudo de esta distancia carece de interpretabilidad intuitiva para el usuario final. Para lograr un entorno de aprendizaje efectivo, es fundamental traducir esta distancia en una métrica de

retroalimentación comprensible, representada como un porcentaje de similitud.

Inicialmente, una transformación lineal simple a partir de un valor máximo teórico presenta una limitante pedagógica: el umbral óptimo de aceptación de la distancia ($DTW \leq 600$) no necesariamente se alinea con la percepción psicológica estándar de una “calificación aprobatoria” (por ejemplo, un 70 %). Para resolver esta discrepancia y mejorar la experiencia de usuario, se implementó una función de mapeo lineal por tramos (*piecewise linear mapping*).

Esta función garantiza que el límite de aceptación algorítmica ($DTW = 600$) equivalga exactamente a un porcentaje de similitud predefinido como aprobatorio ($P_a = 70\%$). Matemáticamente, el cálculo del porcentaje de similitud $S(d)$ en función de la distancia d devuelta por el algoritmo DTW se define como:

$$S(d) = \begin{cases} 100, & \text{si } d \leq 0 \\ 100 - \left(\frac{d}{600}\right) \times 30, & \text{si } 0 < d \leq 600 \\ 70 - \left(\frac{d-600}{1500}\right) \times 70, & \text{si } 600 < d < 2100 \\ 0, & \text{si } d \geq 2100 \end{cases} \quad (4.1)$$

Donde:

- El **primer tramo** ($0 < d \leq 600$) evalúa las ejecuciones correctas, asignando una similitud proporcional entre el 100 % y el 70 %. Esto asegura que cuando el sistema determina que la seña es correcta, el usuario visualice un porcentaje igual o mayor a 70 %.
- El **segundo tramo** ($600 < d < 2100$) evalúa las ejecuciones incorrectas, reduciendo gradualmente la similitud desde el 69,9 % hasta el 0 %, estableciendo a 2100 como la distancia máxima de tolerancia antes del fallo absoluto.

Finalmente, para blindar el sistema contra anomalías en la captura de datos (como pérdida temporal de seguimiento de los puntos clave anatómicos o movimientos atípicos que generen distancias exageradamente altas), la función incluye un acotamiento (*clamping*) en sus extremos. Cualquier distancia que exceda el límite máximo establecido ($d \geq 2100$) es truncada a una similitud de 0 %, evitando el cálculo de porcentajes negativos y manteniendo la consistencia de la retroalimentación visual mostrada en la interfaz.

4.9. Evaluación de usabilidad mediante encuestas

Una vez concluida la fase de desarrollo, se consideró necesario incorporar la perspectiva de los usuarios finales. Con este propósito, se diseñó y aplicó un cuestionario de usabilidad enfocado en población general interesada en aprender la Lengua de Señas Mexicana (LSM), dado que el objetivo primordial del sistema es fungir como una herramienta de apoyo educativo para personas sin dominio previo de la lengua. Por consideraciones éticas y de privacidad en la investigación, no se solicitaron datos médicos o audiológicos a los participantes, garantizando un muestreo por conveniencia basado estrictamente en el interés de uso.

El cuestionario fue distribuido junto con el ejecutable de la aplicación, obteniéndose $N = 33$ respuestas completas. El tamaño de esta muestra se justifica metodológica y estadísticamente por dos razones fundamentales. En primer lugar, en el ámbito de la Interacción Humano-Computadora (HCI), la literatura establece que para estudios cuantitativos de usabilidad mediante cuestionarios, una muestra de entre 20 y 30 usuarios es suficiente para estabilizar las métricas y obtener resultados fiables Lewis (2018).

En segundo lugar, desde la perspectiva estadística, alcanzar una muestra superior a 30 participantes ($N \geq 30$) permite satisfacer el Teorema del Límite Central, el cual establece que la distribución de las medias muestrales tiende a la normalidad independientemente de la distribución original de los datos. Esta condición es crucial, ya que proporciona la robustez matemática necesaria para validar el uso de pruebas paramétricas, como la prueba t de Student aplicada en el análisis inferencial de este estudio.

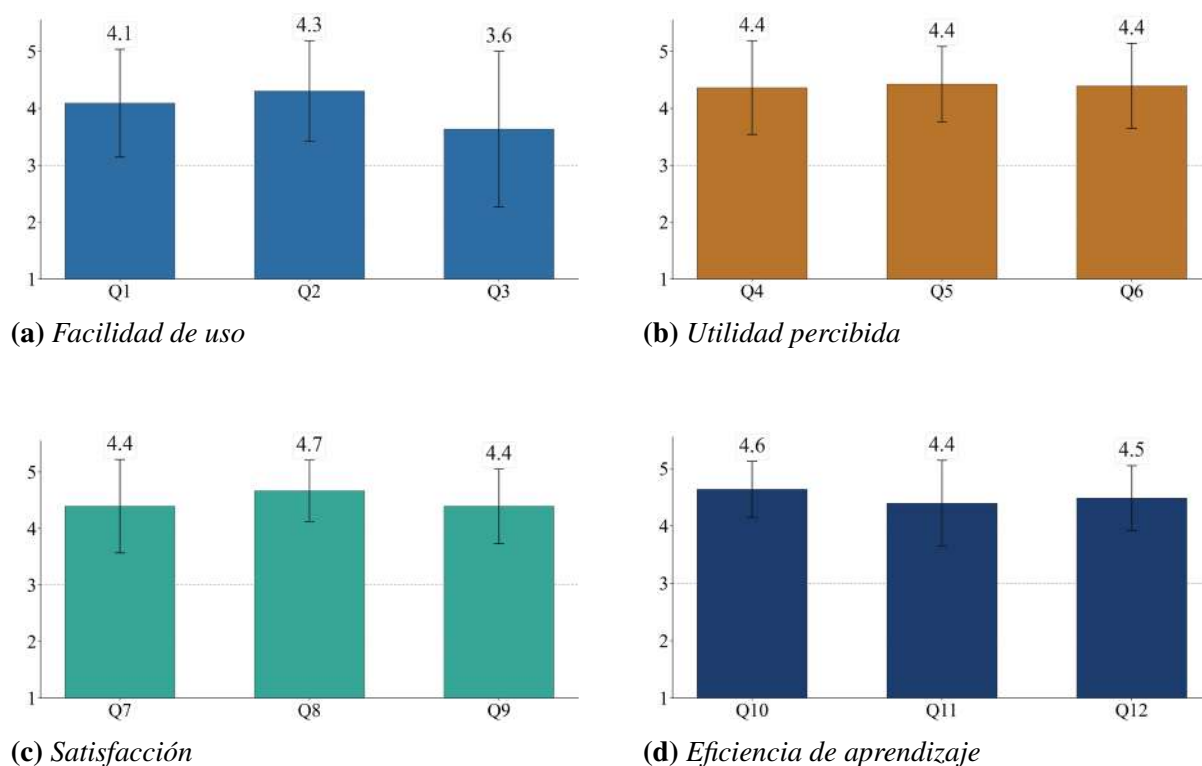
4.10. Resultados del cuestionario de evaluación con usuarios

El cuestionario se basó en una escala Likert de cinco puntos (1–5). Los ítems se agruparon en cuatro categorías: Facilidad de uso, Utilidad percibida, Satisfacción y Eficiencia del aprendizaje.

La Figura 4.10 presenta la distribución de las respuestas obtenidas. La línea punteada indica el valor neutral de referencia (3).

Figura 4.10

Distribución de las respuestas del cuestionario de usabilidad por categoría (N = 33)



Nota. Los resultados reflejan la percepción de los usuarios tras interactuar con el prototipo final de la aplicación.

En conjunto, los resultados cuantitativos indican una valoración global positiva de la aplicación en todas las categorías evaluadas.

4.10.1. Confiabilidad interna – Alfa de Cronbach

Se calculó el coeficiente Alfa de Cronbach para los ítems comprendidos entre las preguntas Q1 y Q12. El valor global obtenido fue $\alpha = 0,81$, lo cual indica una alta confiabilidad del instrumento.

4.11. Análisis inferencial mediante prueba t

Se realizaron pruebas t de una muestra con el objetivo de determinar si la media obtenida en cada ítem del cuestionario era significativamente mayor al valor neutro de referencia (3) en la

escala Likert utilizada. Este valor representa una percepción neutral, por lo que valores significativamente superiores indican una evaluación positiva por parte de los usuarios.

El nivel de significancia considerado fue $\alpha = 0,05$. En todos los casos donde $p < 0,05$, se rechaza la hipótesis nula y se concluye que la media del ítem es significativamente mayor que el valor de referencia.

Los resultados se presentan a continuación, organizados por categorías de evaluación.

4.11.1. Facilidad de uso

Tabla 4.6

Resultados de la prueba t para la categoría Facilidad de Uso

Ítem	Media	Desv. Est.	p -valor	Interpretación
Facilidad del menú	4.09	0.95	$1,86 \times 10^{-7}$	Signif. > 3
Instrucciones claras	4.30	0.88	$1,10 \times 10^{-9}$	Signif. > 3
Botones funcionales	3.64	1.37	0.0116	Signif. > 3

Los resultados muestran que todos los ítems de la categoría *Facilidad de uso* presentan medias superiores al valor neutro y diferencias estadísticamente significativas. Esto indica que los usuarios perciben la interfaz como intuitiva y funcional, destacando especialmente la claridad de las instrucciones y la facilidad de navegación dentro del sistema.

4.11.2. Utilidad percibida

Tabla 4.7

Resultados de la prueba t para la categoría Utilidad Percibida

Ítem	Media	Desv. Est.	p -valor	Interpretación
Claridad videos	4.36	0.82	$7,33 \times 10^{-11}$	Signif. > 3
Comparación DTW	4.42	0.66	$1,04 \times 10^{-13}$	Signif. > 3
Diseño app	4.39	0.75	$4,10 \times 10^{-12}$	Signif. > 3

En la categoría *Utilidad percibida*, todos los ítems presentan valores altamente significativos ($p \ll 0,05$), lo que evidencia que los usuarios consideran que el sistema cumple eficazmente su propósito. Destaca particularmente la funcionalidad de comparación mediante DTW, lo que sugiere que la retroalimentación proporcionada resulta clara y útil para el aprendizaje.

4.11.3. Satisfacción

Tabla 4.8

Resultados de la prueba t para la categoría Satisfacción

Ítem	Media	Desv. Est.	p -valor	Interpretación
Satisfacción gral.	4.39	0.83	$4,94 \times 10^{-11}$	Signif. > 3
Recomendación	4.67	0.54	$4,14 \times 10^{-18}$	Signif. > 3
Motivación	4.39	0.66	$1,55 \times 10^{-13}$	Signif. > 3

Los resultados de la categoría *Satisfacción* muestran niveles elevados en todos los ítems evaluados. En particular, la disposición de los usuarios a recomendar la aplicación presenta la media más alta, lo cual es un indicador sólido de aceptación y satisfacción general con el sistema desarrollado.

4.11.4. Eficiencia del aprendizaje

Tabla 4.9

Resultados de la prueba t para la categoría Eficiencia del Aprendizaje

Ítem	Media	Desv. Est.	p -valor	Interpretación
Organización	4.64	0.49	$3,74 \times 10^{-19}$	Signif. > 3
Retroalim. útil	4.39	0.75	$4,10 \times 10^{-12}$	Signif. > 3
Aprendí señas	4.48	0.57	$4,31 \times 10^{-16}$	Signif. > 3

En la categoría *Eficiencia del aprendizaje*, los resultados indican que los usuarios perciben una mejora significativa en su proceso de aprendizaje. Las medias elevadas sugieren que el sistema facilita la organización del contenido, proporciona retroalimentación útil y contribuye efectivamente a la adquisición de nuevas señas.

4.11.5. Interpretación general de resultados

En conjunto, los resultados obtenidos permiten afirmar que todas las categorías evaluadas presentan diferencias estadísticamente significativas respecto al valor neutro de referencia. Esto implica que la percepción de los usuarios hacia el sistema es consistentemente positiva en términos de usabilidad, utilidad, satisfacción y eficiencia de aprendizaje.

Por lo tanto, se rechaza la hipótesis nula en todos los casos analizados y se concluye que el sistema desarrollado cumple de manera efectiva con los objetivos planteados, ofreciendo una herramienta funcional y bien valorada para el aprendizaje de la Lengua de Señas Mexicana.

4.12. Conclusión del análisis cuantitativo

Los resultados obtenidos a través del cuestionario indican una valoración positiva por parte de los usuarios hacia la aplicación desarrollada. En todas las categorías evaluadas: facilidad de uso, utilidad percibida, satisfacción y eficiencia del aprendizaje, las medias de los ítems superaron de manera significativa el valor neutro de referencia (3). Estos resultados permiten concluir que la aplicación no solo fue bien recibida en términos funcionales, sino que también logró impactar positivamente en el proceso de aprendizaje de los participantes.

Capítulo 5

Conclusiones y Trabajo Futuro

5.1. Conclusiones

La Lengua de Señas Mexicana (LSM) constituye un medio de comunicación esencial para las personas sordas y con dificultades en el habla, ya que permite la expresión efectiva de ideas, emociones y necesidades, además de favorecer su inclusión social y educativa. En este contexto, el desarrollo de herramientas tecnológicas que apoyen el aprendizaje de la lengua de señas resulta fundamental para ampliar el acceso a procesos educativos más inclusivos, flexibles e innovadores.

En este trabajo se desarrolló una aplicación orientada al aprendizaje de LSM que integra técnicas de inteligencia artificial para la captura y análisis del movimiento de manos y rostro. A través del uso de MediaPipe, se extrajeron 42 puntos espaciales de las manos y 6 puntos del rostro, los cuales fueron normalizados y comparados mediante el algoritmo *Dynamic Time Warping* (DTW). Esta metodología permitió cuantificar de manera objetiva la similitud temporal y espacial entre las señas ejecutadas por un intérprete de referencia y aquellas realizadas por el usuario.

Los resultados experimentales demostraron que el uso de DTW es adecuado para la evaluación de la ejecución de señas, particularmente en escenarios donde existe variabilidad temporal entre las muestras. En las pruebas realizadas, se observó que las señas ejecutadas con manos distintas (intérprete con mano izquierda y usuario con mano derecha) generaron valores de distancia superiores a 1000, lo que refleja una baja similitud. En contraste, cuando las señas fueron ejecutadas utilizando la mano correcta, los valores de distancia se mantuvieron consistentemente por debajo de dicho umbral. Estos resultados permitieron establecer un criterio empírico para determinar si una seña fue realizada de manera adecuada, el cual fue integrado exitosamente en el sistema de retroalimentación de la aplicación.

Adicionalmente, se llevaron a cabo pruebas de robustez mediante la modificación controlada de los videos de entrada, considerando variaciones en la calidad del video, reducción de resolución, cambios en las condiciones de iluminación, aplicación de desenfoque y oclusiones parciales de

las manos. Los resultados mostraron que, aun bajo estas condiciones no ideales, las distancias DTW correspondientes a señas correctamente ejecutadas no presentaron variaciones significativas. Esto indica que el sistema mantiene un comportamiento estable y confiable ante escenarios reales de uso, donde el hardware y las condiciones ambientales pueden variar entre usuarios.

Desde la perspectiva del usuario final, la evaluación de usabilidad realizada mediante encuestas reflejó una percepción ampliamente positiva en términos de facilidad de uso, utilidad percibida, satisfacción general y apoyo al proceso de aprendizaje. Los participantes destacaron la claridad de los videos demostrativos, la utilidad de la retroalimentación automática basada en DTW y la organización progresiva de las lecciones. Asimismo, el análisis estadístico del cuestionario evidenció una alta consistencia interna, lo que respalda la confiabilidad de los resultados obtenidos.

En conjunto, los resultados técnicos y de usabilidad confirman que la aplicación desarrollada cumple con los objetivos planteados, al ofrecer una herramienta funcional, accesible y efectiva para el apoyo en el aprendizaje de la Lengua de Señas Mexicana. La integración de técnicas de inteligencia artificial, junto con el uso de métodos de comparación temporal como DTW, demostró ser una estrategia viable para proporcionar retroalimentación automática sobre la ejecución de señas, contribuyendo así al desarrollo de soluciones tecnológicas orientadas a la inclusión y accesibilidad educativa.

Un hallazgo relevante del presente trabajo fue la evaluación práctica de animaciones generadas a partir de video como posible recurso para la representación de señas dentro del sistema. A partir de los resultados obtenidos, se observó que dichas animaciones no reproducían con suficiente precisión los movimientos del intérprete, particularmente en la configuración de las manos, la articulación de los dedos y la fluidez general del gesto, presentando diferencias notables respecto a los videos de referencia.

En consecuencia, se optó por utilizar directamente los videos reales del intérprete como base para la visualización y evaluación de las señas, lo cual permitió garantizar una representación fiel del movimiento y una retroalimentación más confiable para el usuario. No obstante, las animaciones generadas fueron conservadas en formato *.fbx* como parte del repositorio del proyecto, reconociendo su posible utilidad en futuros trabajos, conforme evolucionen las herramientas de generación y mejora de animaciones.

A pesar de estos resultados favorables, es importante reconocer algunas limitaciones en la versión actual del sistema. En primer lugar, la aplicación opera con un corpus de señas restringido, lo que acota su uso en una diversidad más amplia de contextos comunicativos. En segundo lugar, aunque el sistema extrae puntos clave del rostro, la evaluación recae predominantemente en la trayectoria de las manos, dejando un área de oportunidad en el análisis profundo de carac-

terísticas corporales y expresiones no manuales, las cuales son fundamentales en la estructura de la LSM. Finalmente, la retroalimentación actual es de carácter evaluativo, ya que se limita a indicar el nivel de similitud de la seña, pero carece de un enfoque correctivo que brinde instrucciones pedagógicas detalladas sobre qué aspecto exacto de la postura o el movimiento debe ajustarse. Estas áreas de oportunidad constituyen el punto de partida directo para las mejoras que se abordan en el trabajo futuro.

Finalmente, con base en los resultados cuantitativos de la evaluación de usabilidad y el desempeño técnico del algoritmo de alineación temporal, es posible concluir de manera categórica que la hipótesis planteada al inicio de esta investigación ha sido confirmada. El desarrollo del entrenador virtual, fundamentado en visión por computadora e inteligencia artificial mediante el uso de una cámara convencional, efectivamente facilita el aprendizaje de la Lengua de Señas Mexicana. La herramienta demostró funcionar como un apoyo interactivo válido y robusto que complementa la enseñanza tradicional, brindando a las personas oyentes una plataforma accesible para adquirir e interiorizar conocimientos de esta lengua.

5.2. Trabajo Futuro

A partir de los resultados obtenidos, este trabajo sienta las bases para futuras investigaciones orientadas a la ampliación del corpus de señas, incorporando un vocabulario más extenso y diverso que permita cubrir un mayor número de contextos comunicativos. Asimismo, se plantea la incorporación de un mayor número de características faciales y corporales, con el fin de mejorar la precisión en la evaluación de señas que dependen en mayor medida de la expresión no manual.

En el ámbito pedagógico, se sugiere la implementación de nuevos módulos de práctica y gamificación. Por ejemplo, diversificar las dinámicas de estudio integrando ejercicios interactivos de opción múltiple, donde el sistema muestre distintas ejecuciones en video y el usuario deba identificar visualmente la seña correcta. Ampliar estas opciones, yendo más allá de la repetición guiada, enriquecería significativamente la experiencia de usuario y fortalecería la retención del conocimiento.

Por otro lado, dado que la aplicación actual fue desarrollada en *Python*, la lógica del sistema cuenta con una arquitectura altamente escalable. Por lo tanto, se proyecta como trabajo futuro la migración y adaptación del entorno a dispositivos móviles (*smartphones* y *tablets*). Esta transición tecnológica incrementaría drásticamente la accesibilidad de la herramienta, permitiendo a

las personas practicar desde cualquier lugar sin depender de un equipo de cómputo de escritorio.

Finalmente, se propone la mejora de los mecanismos de retroalimentación personalizada, de modo que el sistema no solo indique si una seña fue ejecutada correctamente, sino que también proporcione indicaciones específicas sobre cómo corregir la postura, el movimiento o la secuencia temporal de la seña. Estas líneas de trabajo permitirán fortalecer el alcance y el impacto de la aplicación, consolidándola como una herramienta de apoyo más completa para el aprendizaje de la Lengua de Señas Mexicana.

Bibliografía

- Abdullah, M. et al. (2024a). Advancements in sign language recognition: A comprehensive review and future prospects. *Artificial Intelligence Review*.
- Abdullah, M., Rahman, N., and Hassan, R. (2024b). Machine learning and deep learning paradigms in sign language recognition for the hearing impaired: A comprehensive review. *Journal of Intelligent Systems*.
- Alvarez, L. et al. (2021). The effect of immediate feedback on learning in interactive systems. *Computers & Education*, 173:104253.
- Brooke, J. (2019). Sus: A retrospective. *Journal of Usability Studies*, 8(2):29–40.
- Camgoz, N. C., Koller, O., Hadfield, S., and Bowden, R. (2020). Sign language transformers: Joint end-to-end sign language recognition and translation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6114–6129.
- Carmona Arroyo, G., Rios Figueroa, H. V., and Avendaño Garrido, M. L. (2021). Mexican sign-language static-alphabet recognition using 3d affine invariants. *Programming and Computer Software*, 47:171–192.
- Cheok, M. J., Omar, Z., and 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.
- Elakkiya, R. (2020). Machine learning based sign language recognition: a review and its research frontier. *Journal of Ambient Intelligence and Humanized Computing*, 12:1–28.
- Espejel-Cabrera, J., Cervantes, J., García-Lamont, F., Ruiz-Castilla, J. S., and Jalili, L. D. (2021). Mexican sign language segmentation using color based neuronal networks to detect the individual skin color. *Expert Systems with Applications*, 183:115295.
- Huang, Y., Huang, J., Wu, X., and Jia, Y. (2022). Dynamic sign language recognition based on cbam with autoencoder time series neural network. *Mobile Information Systems*, 2022.
- Instituto Nacional de Estadística y Geografía (2021a). Censo de población y vivienda 2020: Resultados. Technical report, INEGI, Aguascalientes, México.

- Instituto Nacional de Estadística y Geografía (2021b). Estadísticas a propósito del día internacional de las personas con discapacidad (3 de diciembre). Technical report, INEGI, Aguascalientes, México.
- Jauregi, A., Bernarás, E., and Jaureguizar, J. (2020). Derechos humanos desde la diversidad lingüística e identitaria de las personas sordas. *Deusto Journal of Human Rights*, 5(5):187–213.
- Kumar, D. A., Sastry, A. S. C. S., Kishore, P. V. V., Kumar, E. K., and Kumar, M. T. K. (2019). S3drgf: Spatial 3-d relational geometric features for 3-d sign language representation and recognition. *IEEE Signal Processing Letters*, 26(1):169–173.
- Lara-Ortiz, V., Fuentes-Aguilar, R., and Chairez, J. I. (2023). Assisted sign language tutoring system using a bioinspired virtual controlled robotic arm. *SSRN Electronic Journal*.
- Lewis, J. R. (2018). Measuring perceived usability: The csuq, sus, and umux. *International Journal of Human-Computer Interaction*, 34(12):1148–1156.
- Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M. G., Lee, J., Chang, W.-T., Hua, W., Georg, M., and Grundmann, M. (2019). Mediapipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172*.
- Mejía Pérez, K., Córdova Esparza, D. M., Terven, J., Herrera Navarro, A. M., García Ramírez, T., and Ramírez Pedraza, A. (2022). Automatic recognition of mexican sign language using a depth camera and recurrent neural networks. *Applied Sciences*, 12(11).
- Müller, M. (2007). Dynamic time warping. In *Information Retrieval for Music and Motion*, pages 69–84. Springer.
- Ortiz-Leon, C., Yupanqui-Allecca, F., and Meneses-Claudio, B. (2023). Using artificial intelligence for sign language translation: a systematic literature review. In *Salud, Ciencia y Tecnología - Serie de Conferencias*, volume 2, page 446.
- Patel, S. and Shah, M. (2023). Recent trends and advancements in hand gesture recognition from real time video. *Journal of Computer Vision and Applications*.
- Putera, Z. P., Anasanti, M. D., and Priambodo, B. (2018). Designing translation tool: Between sign language to spoken text on kinect time series data using dynamic time warping. *SINERGI*, 22(2):91.

- Robert, E. J. and Duraisamy, H. J. (2023). A review on computational methods based automated sign language recognition system for hearing and speech impaired community. *Concurrency and Computation: Practice and Experience*.
- Singh, A., Sharma, R., and Kaur, P. (2020). Bridging the communication gap: A review of automatic sign language recognition techniques. *International Journal of Computer Applications*.
- Trujillo Romero, U. and García Bautista, G. (2021). Reconocimiento de palabras de la lengua de señas mexicana utilizando información rgb-d. *Revista Facultad de Ingeniería*, 10(2):1–23.
- Williams, J. and Carter, L. (2022). Assistive technologies for american sign language users: A systematic mapping study. *Assistive Technology*.
- Zhang, Z., Wang, Q., et al. (2020). Deep learning-based human pose estimation: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1):1–21.
- Zhao, G. et al. (2021). Gesture recognition using skeleton-based dynamic time warping. *Pattern Recognition Letters*, 145:64–70.

Apéndice A

Cuestionario de Validación de la Aplicación de Lengua de Señas

A continuación se presenta el instrumento original utilizado para evaluar la utilidad, facilidad de uso y áreas de mejora de la aplicación desarrollada.

Este cuestionario busca evaluar la utilidad, facilidad de uso y áreas de mejora de una aplicación diseñada para apoyar el aprendizaje de la Lengua de Señas Mexicana. La información recolectada se usará con fines educativos y de investigación, como parte de un artículo académico. Este proyecto busca fomentar la inclusión de personas sordas mediante herramientas accesibles para toda persona interesada en aprender LSM. ¡Gracias por participar!

1. ¿Tienes alguna condición relacionada con la audición?
 - ☐ Sí, soy una persona sorda
 - ☐ Sí, tengo una pérdida auditiva parcial
 - ☐ No
 - ☐ Prefiero no responder

2. ¿Qué tan fácil fue navegar por el menú principal para encontrar las lecciones?
 - ☐ Muy difícil
 - ☐ Difícil
 - ☐ Ni fácil ni difícil
 - ☐ Fácil
 - ☐ Muy fácil

3. ¿Las instrucciones para usar la función de práctica (botón de “Practicar”) fueron claras y comprensibles?
 - ☐ Muy en desacuerdo

- ☐ En desacuerdo
 - ☐ Ni de acuerdo ni en desacuerdo
 - ☐ De acuerdo
 - ☐ Muy de acuerdo
4. ¿Los botones de “Anterior”, “Reproducir”, “Pausar” y “Siguiente” funcionaron correctamente y fueron fáciles de usar?
- ☐ Muy en desacuerdo
 - ☐ En desacuerdo
 - ☐ Ni de acuerdo ni en desacuerdo
 - ☐ De acuerdo
 - ☐ Muy de acuerdo
5. ¿Consideras que los videos son claros para entender las señas?
- ☐ Muy en desacuerdo
 - ☐ En desacuerdo
 - ☐ Ni de acuerdo ni en desacuerdo
 - ☐ De acuerdo
 - ☐ Muy de acuerdo
6. ¿El sistema de práctica (comparación de puntos con DTW) te ayudó a mejorar tus señas?
- ☐ Muy en desacuerdo
 - ☐ En desacuerdo
 - ☐ Ni de acuerdo ni en desacuerdo
 - ☐ De acuerdo
 - ☐ Muy de acuerdo
7. ¿El diseño de la aplicación contribuyó a tu aprendizaje de Lengua de Señas?
- ☐ Muy en desacuerdo
 - ☐ En desacuerdo
 - ☐ Ni de acuerdo ni en desacuerdo
 - ☐ De acuerdo
 - ☐ Muy de acuerdo

8. ¿Qué tan satisfecho estás con tu experiencia general usando la aplicación?

- ☐ Muy insatisfecho
- ☐ Insatisfecho
- ☐ Ni satisfecho ni insatisfecho
- ☐ Satisfecho
- ☐ Muy satisfecho

9. ¿Recomendarías esta aplicación a otras personas interesadas en aprender Lengua de Señas?

- ☐ Muy en desacuerdo
- ☐ En desacuerdo
- ☐ Ni de acuerdo ni en desacuerdo
- ☐ De acuerdo
- ☐ Muy de acuerdo

10. ¿Qué tan motivado te sentiste para continuar practicando las señas usando la aplicación?

- ☐ Nada motivado
- ☐ Poco motivado
- ☐ Neutral (ni motivado ni desmotivado)
- ☐ Motivado
- ☐ Muy motivado

11. ¿Crees que las lecciones están bien organizadas para aprender de manera progresiva?

- ☐ Muy en desacuerdo
- ☐ En desacuerdo
- ☐ Ni de acuerdo ni en desacuerdo
- ☐ De acuerdo
- ☐ Muy de acuerdo

12. ¿Consideras útil el sistema de retroalimentación (distancia DTW) para corregir tus señas?

- ☐ Muy en desacuerdo
- ☐ En desacuerdo

- ☐ Ni de acuerdo ni en desacuerdo
- ☐ De acuerdo
- ☐ Muy de acuerdo

13. ¿La aplicación te ayudó a aprender señas que antes no conocías?

- ☐ Muy en desacuerdo
- ☐ En desacuerdo
- ☐ Ni de acuerdo ni en desacuerdo
- ☐ De acuerdo
- ☐ Muy de acuerdo

14. ¿Qué mejorarías en la aplicación? _____

Nota Final: *Tus respuestas son confidenciales y anónimas, y serán utilizadas únicamente con fines educativos y de investigación. Tu participación es voluntaria y puedes salir del cuestionario en cualquier momento sin consecuencia alguna. Agradecemos sinceramente tu colaboración. Tu aporte ayudará a mejorar esta herramienta educativa y a promover investigaciones que impulsen la inclusión a través del aprendizaje de la Lengua de Señas Mexicana.*