

Traducción automática del lenguaje dactilológico de sordos y sordomudos mediante sistemas adaptativos

Daniel Betancur Betancur¹, Mateo Vélez Gómez¹, Alejandro Peña Palacio^{1,✉}

¹ Escuela de Ingeniería de Antioquia (EIA). Envigado, Colombia

Recibido 15 de noviembre de 2012. Aprobado 21 de marzo de 2013. En discusión hasta el 1 de junio de 2013

AUTOMATIC TRANSLATION OF THE DACTILOLOGIC LANGUAGE OF HEARING IMPAIRED BY ADAPTIVE SYSTEMS

Resumen — Una de las principales limitaciones que presentan las personas con discapacidad auditiva está directamente relacionada con su dificultad para interactuar con otras personas, ya sea de forma verbal o a través de sistemas auxiliares basados en la voz y el audio. En este artículo se presenta el desarrollo de un sistema integrado de *hardware* y *software*, para el reconocimiento automático del lenguaje dactilológico de señas utilizado por personas con este tipo de discapacidad. El *hardware* está compuesto por un sistema inalámbrico adherido a un guante, el cual posee un conjunto de sensores que capturan una serie de señales generadas por los movimientos gestuales de la mano, y un modelo por adaptación basado en los principios de la computación neuronal, el cual permite su reconocimiento en términos de un lenguaje dactilológico en particular. Los resultados arrojados por el sistema integrado mostraron gran efectividad en el reconocimiento de las vocales que conforman el lenguaje dactilológico en español, esto gracias a la capacidad que posee el modelo de asociar un conjunto de señales de entrada, con un movimiento dactilológico en particular.

Palabras clave — Computación neuronal híbrida, lenguaje dactilológico, Protocolo *ZigBee*, discapacidad auditiva, sistemas de reconocimiento.

Abstract — One of the main limitations of the people with hearing impairment is directly related to their difficulty interacting with others, either verbally or through auxiliary systems based on voice and audio. This paper presents the development of an integrated system of hardware and software for automatic fingerspelling sign language used by people with this type of disability. The hardware system comprises a glove which has a set of wireless sensors that capture a series of signals generated by the hand gestures, and a adaptive model based on the principles of neural computation, that allows recognition of a particular dactilologic language. Results from the integrated system showed great effectiveness in recognizing vowels from the dactilologic Spanish language. This recognition was influenced by the dimensionality reduction made by the neural model of the input signals representing movements, and the sensitivity factor that sets the limit between recognition and learning.

Keywords — Dactilologic language, Hearing impairment, Hybrid neural computing, Zigbee protocol.

I. INTRODUCCIÓN

La comunicación es el medio por el cual los seres humanos comparten sus ideas, pensamientos, sentimientos u opiniones, constituyéndose en un elemento primordial para el desarrollo de una persona dentro de una

✉ Dirección para correspondencia: pfjapena@gmail.com

sociedad. Como seres humanos, los individuos utilizan diferentes formas de comunicación, siendo el habla la más común, seguida de la escritura y finalmente de cualquier otro tipo de lenguaje de señas creado para tal fin [1,2]. Como individuos, las personas sordas y sordomudas poseen una capacidad auditiva nula que no les permite

aprender a hablar, por lo que se ven obligadas a emitir o a captar mensajes por medios diferentes al sonido, como son la escritura o el lenguaje dactilológico, para lo cual sólo requieren del sentido de la vista [3].

El lenguaje dactilológico es un lenguaje de señas generado por los movimientos gestuales de la mano, en donde se busca interpretar cada letra del alfabeto a través de diferentes figuras o formas. Este lenguaje varía de comunidad en comunidad, donde cada una maneja un gentilicio para su idioma gestual, por ejemplo las personas con discapacidad auditiva en comunidades de habla francesa utilizan el *langue des signes française* conocido como LSF [4], en los Estados Unidos de América (USA) este lenguaje se conoce como *american sign language* (ASL) [5], y en el idioma español en donde se enmarca el lenguaje de signos colombiano (Fig. 1), este se conoce como *lengua de signos española* (LSE) [6,7].

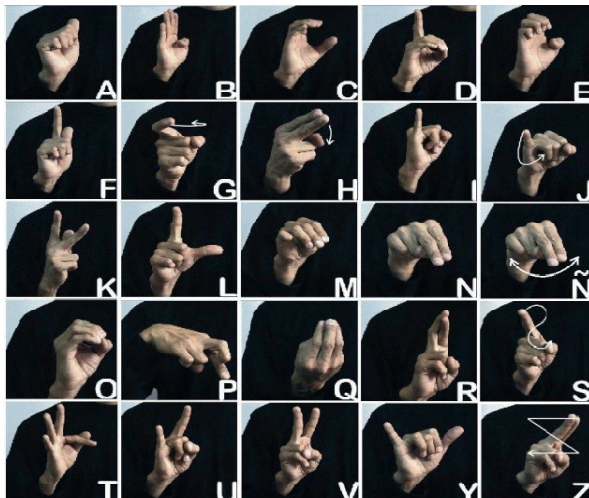


Fig.1. Movimientos gestuales que definen el lenguaje dactilológico en español utilizado en Colombia.

A pesar de la gran variedad de lenguajes dactilológicos existentes, y a pesar de la gran cantidad de signos que estos poseen, la comunicación verbal sigue siendo una limitación importante en la formación de personas con discapacidad auditiva, hecho que se evidencia aún más en su formación a través de ambientes virtuales en donde predomina la comunicación verbal, lo que supone un impacto directo en su inserción en la sociedad y de manera general en el mundo laboral [8].

La comunicación no verbal en personas con discapacidad auditiva involucra una serie de aspectos relacionados con los movimientos gestuales de la mano como son las posturas, en donde se encuentran la forma y la orientación de la mano, o los gestos temporales, los cuales se relacionan con su movimiento y posición [9,10], y de manera general con los estados de transición y las

posturas del cuerpo humano [11]. Para el reconocimiento de un movimiento gestual desde las posturas y los gestos temporales, se han desarrollado una serie de trabajos basados en el procesamiento digital de imágenes sobre video, los cuales reconocen un movimiento gestual en tiempo real, mediante la utilización de técnicas como la transformada wavelet, o las redes neuronales por aprendizaje no supervisado [11]. Para el caso del reconocimiento de movimientos gestuales a partir de imágenes que describen trayectorias sobre puntos específicos ubicados en la mano, se han utilizado sistemas como el *Kinect Motion de Microsoft* (www.xbox.com/es-ES/kinect), o modelos que utilizan regresiones no lineales múltiples para la descripción de dichas trayectorias en el espacio [12]. De manera general, estos sistemas de reconocimiento dependen en gran medida de la postura y la geometría de la mano, lo que genera un problema en el reconocimiento de dichos movimientos desde el punto de vista de gestos temporales, ya que estos requieren de la utilización de herramientas sofisticadas propias del procesamiento digital de imágenes, las cuales aumentan en complejidad cuando el número de símbolos o palabras aumenta de forma considerable [13].

Para la solución del problema del reconocimiento de gestos temporales muchas empresas e investigadores han desarrollado una serie de dispositivos o guantes, los cuales en una etapa muy temprana tuvieron una fuerte tendencia hacia los videojuegos, como lo demuestra el guante *Syreglove* desarrollado en el año de 1977 por la empresa *Electronic Visualization Laboratory*, con el fin de controlar las acciones de un personaje dentro de un juego [14]. Posteriormente, en el año de 1989, se crearon otros guantes como el *powerglove*, desarrollado por la empresa *Nintendo Entertainment System* con ayuda de la empresa *Mattel* [15], y el *cyberglove*, desarrollado en el año de 1990 por la empresa *CyberGlove Systems* [16]. A pesar de los avances tecnológicos logrados durante la época, estos guantes fueron utilizados para fines recreativos y de diversión, no obstante en el año 2000 surgieron dos dispositivos capaces de asociar las letras del abecedario con los movimientos de las manos, el *American Sign Language Translator* creado por Rían Patterson en el año 2006 [17], y el *AcceleGlove*, creado en el año 2007 por los estudiantes José Hernández Rebollar y Nicholas Kyriakopoulos en George Washington University en Estados Unidos [18]. Actualmente, en el ámbito investigativo se han venido desarrollado otro tipo de guantes traductores del lenguaje de señas, los cuales centran su funcionamiento en una serie de dispositivos que permiten capturar los movimientos temporales de la mano, en términos de una serie de señales espacio temporales, eliminando de plano los sistemas que utilizan el procesamiento digital

de imágenes sobre video, tal y como ocurre en los sistemas de reconocimiento centrado en las posturas y la geometría de la mano de un individuo [13,19].

En este artículo se describe el desarrollo de un sistema integrado *hardware-software* para la traducción automática del lenguaje dactilológico utilizado por sordos y sordomudos. El *hardware* del sistema está compuesto por un guante que captura los movimientos gestuales de la mano de un individuo, en términos de una serie de señales espacio-temporales, y el *software* integra un modelo computacional por adaptación y aprendizaje, el cual permite el reconocimiento automático de dichos movimientos en términos de un lenguaje dactilológico en particular.

Para evaluar la efectividad del modelo frente al reconocimiento, se utilizaron como referencia los movimientos gestuales que representan las cinco (5) vocales del lenguaje dactilológico español, y realizados por un individuo sano. Para tal efecto, el modelo computacional fue evaluado teniendo en cuenta para ello dos submodelos que lo componen, un primer submodelo que permite la reducción de dimensionalidad en las señales de entrada, y un segundo submodelo o de reconocimiento, el cual permite la identificación de los movimientos gestuales de la mano en términos de un lenguaje dactilológico en particular. Los resultados arrojados por el sistema de reconocimiento propuesto muestran la potencialidad de dicho sistema para ser extendido no solo al reconocimiento de una mayor cantidad de símbolos, sino para el reconocimiento de otro tipo de lenguajes dactilológicos.

II. METODOLOGÍA

El lenguaje dactilológico está compuesto por una serie de movimientos gestuales que configuran la forma más usual de comunicación en personas con discapacidad auditiva. Para el reconocimiento automático de dicho lenguaje, se desarrolló un sistema integrado de *hardware* y *software*, en donde el *hardware* permite la adquisición de los movimientos de la mano en términos de un conjunto de señales, y el *software* permite el reconocimiento automático de dichos movimientos, en términos de las señas que componen un lenguaje dactilológico en particular. Para el reconocimiento automático de un lenguaje dactilológico cualquiera, el sistema efectúa la secuencia de procesos descrita en la Fig. 2.

2.1. Desarrollo del Hardware

Para la captura de los movimientos gestuales que conforman un lenguaje dactilológico en particular, se procedió con la construcción de un guante o *hardware*, el cual está provisto de seis (6) acelerómetros de tres ejes (MMA7361L), cinco (5) de los cuales están ubicados sobre la punta de cada uno de los dedos, y un acelerómetro adicional ubicado sobre el dorso de la muñeca. Para el procesamiento de las señales provenientes de cada acelerómetro, el *hardware* incorpora una tarjeta de procesamiento de datos del tipo Seeeduino Mega (Microcontrolador ATmega 1280), y para la comunicación inalámbrica, el *hardware* utiliza dos módulos de comunicación del tipo Zigbee (XBee Pro Serie 2), configurados para una velocidad de transmisión de 115200 bps (Fig.3) [20].



Fig. 2 Proceso general para el reconocimiento automático del lenguaje dactilológico a partir de los movimientos gestuales de la mano.

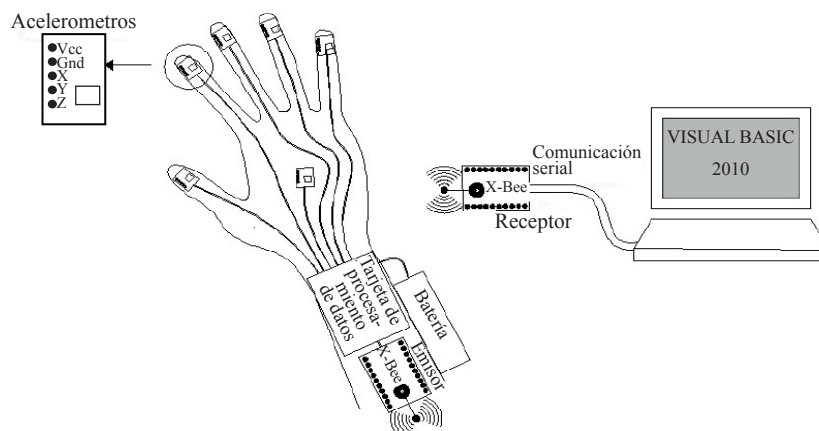


Fig.3. Estructura del guante que define el *hardware* utilizado para la captura de los movimientos gestuales de la mano.

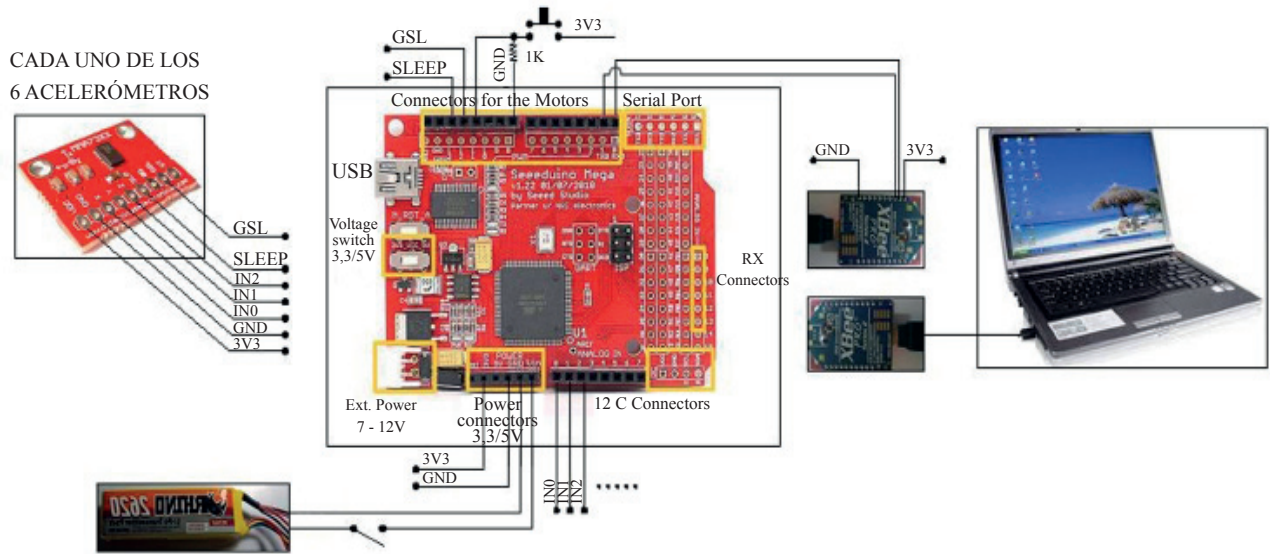


Fig.4. Esquema de conexiones entre los componentes del *hardware* del sistema integrado y el computador.

Las conexiones entre los componentes del *hardware* y los acelerómetros se hicieron de forma serial mediante la utilización de los pines comunes (GND, 3V3, SLEEP y GSL), y los pines de las señales en el microcontrolador se unieron consecutivamente desde la entrada analógica IN0 hasta la entrada analógica IN16 (Fig.4).

De manera general, la estructura del guante está compuesta por dos capas de tela unidas por costuras (capa interna – capa externa). Sobre las puntas de los dedos el guante posee una serie de agujeros en donde van ubicados cinco (5) de los seis (6) acelerómetros mencionados anteriormente, y los circuitos son fijados a la capa interna mediante la utilización de resortes y pequeños bolsillos. La tarjeta de procesamiento de datos se ubicó justo debajo de la muñeca, con el módulo de comunicación (Xbee) ubicado en la mitad del antebrazo. La conexión de los componentes del *hardware* se realizó mediante cables

flexibles de múltiples hilos de cobre con el fin de facilitar el movimiento de la mano y evitar su rompimiento por fatiga (Fig. 5).

Así, cada movimiento gestual fue representado por un total de 18 señales analógicas determinadas por las coordenadas *x*, *y* y *z* asociadas con cada uno de los acelerómetros. Para el reconocimiento, el sistema utilizó como referencia 16 señales de las 18 disponibles, eliminando de plano las señales asociadas con el eje *z* sobre los dedos anular y meñique debido a la poca información espacial que estas señales arrojan.

Para cada señal se hizo un proceso de adquisición análogo-digital donde las señales de salida fueron representadas entre 0 y 1023 (adquisición de 10 bits), con relación a un voltaje generado por cada señal entrada, en un rango entre 0 V y 3,3 V. Cada señal de salida fue enviada al computador por el sistema Xbee cada 150 ms a lo largo de 3 segundos que dura un movimiento gestual, obteniéndose un total de 20 instantes de tiempo por señal, configurando así una matriz 20 x 16 datos por movimiento gestual como se muestra en la Fig. 6 y 7. Es de anotar que cada señal de salida fue normalizada por el sistema en un intervalo definido entre 0 y 1.

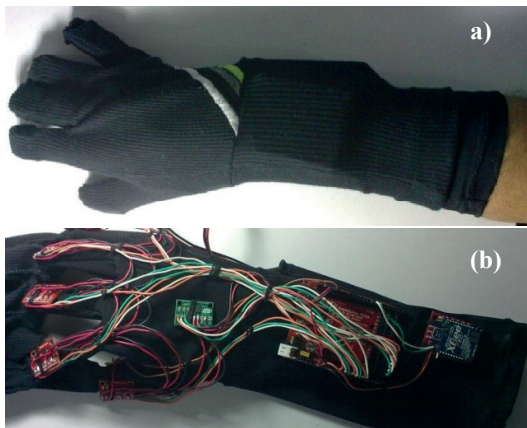


Fig.5. Apariencia general del guante o *hardware* utilizado para la captura de los movimientos gestuales de la mano (a) Capa Externa (b) Capa Interna.

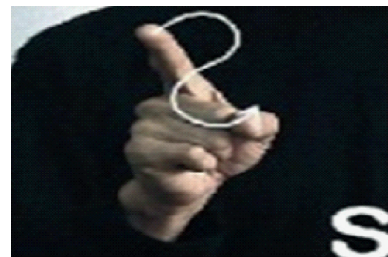


Fig.6. Movimiento gestual que representa una letra en un alfabeto dactilológico cualquiera.

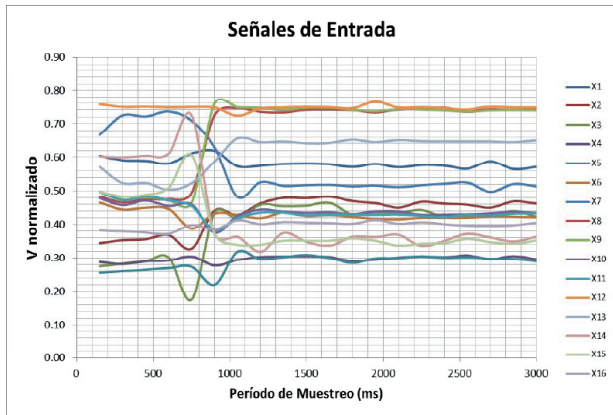


Fig.7. Conjunto de señales que representan un movimiento dactilológico de acuerdo con el hardware del sistema.

2.2. Desarrollo del Modelo Computacional.

Para el reconocimiento de los movimientos gestuales que conforman un lenguaje dactilológico, el sistema incorpora un modelo computacional por adaptación y aprendizaje compuesto por dos submodelos: el primero está basado en los principios de la inteligencia computacional, el cual permite la reducción de dimensionalidad en la matriz (20 x 16) que representa las señales de entrada en términos de una señal de referencia (20 x 1) definida por una serie de números aleatorios generados a partir de una distribución normal $N(0,1)$ (submodelo de reducción). Cada serie de números aleatorios permitió la representación de un movimiento gestual en particular, eliminando de plano la separación ciega de señales, ya que cada una de estas fue obtenida de forma independiente [21]. De manera general, este modelo se denota y define [21]:

$$yr_k = f(x_{i,k}, a_m) \quad (1)$$

Donde:

i : Indica el número de señales que representan un movimiento gestual ($i=1,2,3,\dots,16$).

k : Representa el número de instantes de tiempo que componen una señal para un movimiento gestual ($k=1,2,3,\dots,20$).

$x_{i,k}$: Representa el vector de entrada al modelo en el instante k y para la señal de entrada i (20 x 16).

a_m : Representa los parámetros y variables que establecen una relación funcional entre las señales de entrada y la señal de referencia.

$f()$: Representa la relación funcional entre la matriz de entrada y el vector de salida compuesto por las señales de referencia. Esta función puede considerarse como una red neuronal con funciones de activación de tipo lineal o sigmoideal a la salida [22], mientras que para los modelos neuronales borrosos o neuronales radiales esta relación agrupa las reglas de inferencia que relacionan las variables lingüísticas tanto de entrada como de salida [23-25].

Para establecer la relación funcional entre los datos de entrada y la señal de referencia el modelo incorpora un mecanismo por adaptación y aprendizaje que minimiza el error cuadrático medio:

$$\min \left\{ \sum_{k=1}^{ND} (yd_k - yr_k)^2 \right\} \quad (2)$$

Donde:

ND : Representa el número de instantes de tiempo que componen cada una de las señales que representan un movimiento gestual tanto a la entrada como a la salida.

yr_k : Representa la señal de salida arrojada por la relación funcional luego de la reducción de dimensionalidad (20 x 1).

yd_k : Representa la señal de números aleatorios de referencia que permiten la reducción de dimensionalidad (20 x 1).

Para el caso de modelos basados en los principios de la inteligencia computacional, esta minimización de hace en términos de la regla delta generalizada, la cual opera sobre los parámetros y variables que conforman una relación funcional. La regla delta generalizada se denota y define para un conjunto de parámetros a_m de la siguiente manera [22]:

$$a_m = a_m - \alpha \frac{\partial e_k^2}{\partial a_m} \quad (3)$$

Donde:

a_m : Número de parámetros y variables adaptables en un modelo por adaptación y aprendizaje.

e_k^2 : Representa el error cuadrático medio para un instante k de acuerdo con la señal de referencia.

α : Representa la tasa de aprendizaje para lograr el mínimo en el error cuadrático medio. Este valor está definido en el intervalo $[0,1]$.

Para el reconocimiento, el modelo integra un segundo submodelo inspirado en la estructura de un modelo neuronal por aprendizaje no supervisado del tipo ARTx (submodelo de reconocimiento), el cual establece una

serie de categorías de acuerdo con un conjunto de patrones de reconocimiento establecidos para un lenguaje dactilológico en particular y que matemáticamente se representa por la ecuación (4) [22].

$$z_j = \sum_{k=1}^{20} y r_k \omega_{kj} \quad (4)$$

Donde:

z_j : Representa la salida asociada con cada uno de los patrones de referencia utilizados para el reconocimiento ($j=1,2,\dots, NP$).

NP : Representa el número de neuronas de salida o patrones de referencia.

Para el caso de la red neuronal ARTx, la selección de la neurona ganadora se denota y define:

$$z_j = \max\{z_j\} \quad (5)$$

$y r_k$: Representa la señal de salida arrojada por la relación funcional luego de la reducción de dimensionalidad (20×1).

ω_{kj} : Representan las conexiones hacia adelante del modelo neuronal (20×5). La actualización de los pesos hacia adelante se denota y define:

$$\omega_{kj} = \frac{T_{j,k}}{\sum T_{j,k} + \gamma} \quad (6)$$

Donde:

$T_{j,k}$: Representa la memoria a corto plazo del submodelo de reconocimiento propuesto. La actualización de estos pesos se denota y define:

$$T_{j,k} = T_{j,k} \cdot \gamma r_k \quad (7)$$

γ : Representa el factor de sensibilidad para la actualización de los pesos del submodelo de reconocimiento (FS).

A pesar de que las máquinas de vector soporte son consideradas clasificadores binarios, estas pueden ser extendidas para la clasificación de un conjunto mayor de patrones con un consecuente aumento del tiempo computacional ya que la clasificación se hace por lo general sobre espacios infinitos, lo que las diferencia claramente del modelo propuesto, el cual establece valores de similitud entre los patrones a clasificar sobre un solo espacio finito, mejorando ostensiblemente el tiempo de respuesta en la clasificación de patrones en tiempo real [27, 28].

El modelo propuesto fue implementado mediante la utilización de la programación orientada por objetos y tecnologías .NET, lo que permite la adquisición y procesamiento de señales, además de la interoperabilidad de operaciones con diferentes tipos de *software* y su posibilidad de ser migrado a sistemas operativos libres mediante la plataforma mono.NET [26]. Igualmente, el modelo posee un sistema de visualización basado en una librería *Open Source* llamada *ZedGraph* [29], el cual muestra en un gráfico de barras el comportamiento de los sensores a lo largo del tiempo que dura un movimiento gestual (Fig. 8). El sistema posee además un cuadro de texto que permite la visualización del puerto por el cual se da la adquisición de las señales, así como la letra obtenida como resultado del proceso de reconocimiento a partir de un movimiento gestual. Es de anotar que el proceso de reconocimiento se inicia con la presión de un interruptor ubicado sobre el pulgar en la estructura del *hardware* o *guante*.

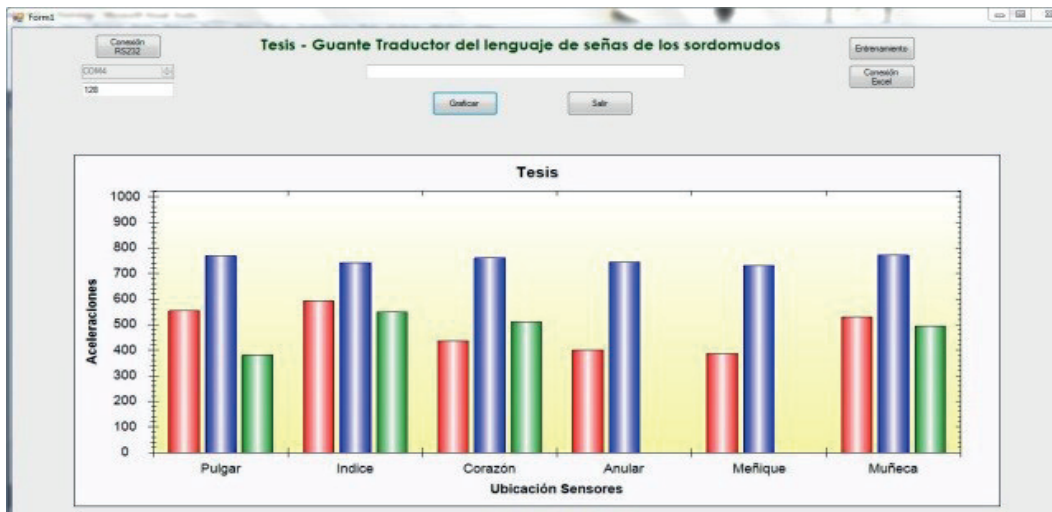


Fig.8. Gráfico de Barras que permite visualizar el comportamiento temporal de los sensores (16) ubicados en el *hardware* del sistema (guante dactilológico).

2.3. MATERIALES Y MÉTODOS

Para el análisis y la validación del modelo se llevaron a cabo una serie de pruebas de reconocimiento automático del lenguaje dactilológico en español sobre un individuo sano. Para tal efecto se tuvieron en cuenta tres etapas: una primera etapa en la cual se hizo la selección y la validación del submodelo de reducción de dimensionalidad, una segunda en la cual se llevó a cabo el proceso de validación del submodelo de reconocimiento a partir de las señales arrojadas por el submodelo de reducción; y en una etapa final, en la que se dio el proceso de validación general del modelo por reconocimiento *on-line*.

2.3.1. Validación del submodelo de reducción.

Para la validación del submodelo de reducción, se construyó una base de datos compuesta por cinco (5) movimientos gestuales asociados con cada una de las vocales del alfabeto dactilológico en español. Cada uno de estos movimientos se describió en términos de una matriz de datos de tamaño 20 x 16 (matriz de entrada), así como por una serie de números aleatorios de tamaño 20 x 1 (señal de referencia) obtenidos a partir de una distribución normal $N(0,1)$. De acuerdo con el número de patrones de referencia (vocales), tanto la matriz de entrada como la señal de referencia fueron extendidas hasta un tamaño de 100 x 16 (20 x 5 x 16) y 100 x 1 (20 x 5 x 1) respectivamente.

Debido a la dificultad que genera la representación de un movimiento gestual a partir de la matriz de entrada, se procedió con la evaluación y selección de un modelo por adaptación y aprendizaje que permita establecer la relación funcional entre dicha matriz y el vector de números aleatorios de referencia. Para tal efecto, se evaluaron cinco modelos basados en los principios de la inteligencia computacional, entre los que se cuentan las redes neuronales MADALINE y Base Radial [22], los modelos borrosos adaptativos de tipo ANFIS [23] y Takagi Sugeno [24,25], así como el modelo por evolución EPR (Evolutionary Polynomial Regression) [30]. Para establecer esta relación funcional, cada modelo fue configurado para un total de 16 valores de entrada y una sola salida, lo que arrojó como resultado la reducción de dimensionalidad en un factor de 16 a 1, para los 100 instantes de tiempo de los cuales se compone tanto la matriz de entrada como la señal de referencia.

2.3.2. Validación del submodelo de reconocimiento.

Para el análisis y la validación del submodelo de reconocimiento, el modelo incorpora un submodelo neuronal por aprendizaje no supervisado, inspirado en la estructura de una red neuronal del tipo ARTx

[31]. De acuerdo con lo anterior, este modelo posee 20 neuronas de entrada correspondientes con cada uno de los valores arrojados por el submodelo de reducción para un movimiento gestual en particular, mientras que la salida está compuesta por cinco (5) neuronas o categorías asociadas con cada uno de los patrones de reconocimiento establecidos.

Para tal efecto, se establecieron un total de 70 ciclos de aprendizaje *off-line*, los cuales comprendieron el ingreso de forma aleatoria de los cinco patrones de reconocimiento o vocales. Para cada ciclo, cada una de las neuronas de la capa de salida fue evaluada frente a un factor de sensibilidad (FS) o factor de reconocimiento, el cual determina el porcentaje de similitud de cada neurona con un movimiento gestual de entrada. Este factor de sensibilidad se denota y define [22]:

$$FS = C(X,V) = \frac{1 + \|X.V\|}{1 + \|X\|} < \rho \quad (8)$$

Donde:

X : Representa la señal arrojada por el modelo de reducción luego de un movimiento gestual. Este vector se denota y define:

$$X = [x_1, x_2, \dots, x_{20}] \quad (9)$$

V : Representa la matriz de conexiones entre la capa de entrada y la capa de salida de la red neuronal de reconocimiento. Esta matriz se denota y define $V_{j,i}$, donde j representa el número de patrones de salida considerados (5), mientras que i representa cada una de las componentes de la señal arrojada por el modelo reducción (20).

2.3.3. Validación general del modelo de reconocimiento.

Se procedió con la integración de los submodelos de reducción y reconocimiento configurando el modelo general de reconocimiento (MAD_ART₂). Este modelo fue validado *on-line*, mediante el ingreso de una serie de movimientos gestuales determinados por los patrones de referencia o vocales, hasta lograr un total de 300 aciertos en el reconocimiento. Esta validación incluyó la evaluación del comportamiento del modelo frente a otro tipo de modelos utilizados para el reconocimiento de patrones, entre los que se cuentan el modelo ART_MAP [32], el Fuzzy ART_MAP [33] y la Red Neuronal de Base Radial con funciones de tipo gaussiano (RNBR) [22]. De esta manera, la efectividad en el reconocimiento fue evaluada en términos del

número de intentos requeridos por cada uno de los modelos para lograr 300 aciertos en el reconocimiento.

Finalmente, y debido a que el modelo propuesto incorpora un mecanismo por adaptación y aprendizaje de tipo no supervisado, se procedió al análisis del efecto que el FS tiene sobre la creación de nuevas categorías, y de manera general sobre el porcentaje de reconocimiento *on-line*. Para la variación del FS se tomaron 6 valores, los cuales van desde 0,5 hasta 1,0, con variaciones de 0,1. Para cada FS se obtuvo el porcentaje de reconocimiento (PR), el cual mide el número de aciertos obtenidos por el modelo propuesto, frente a un total de 100 movimientos gestuales aleatorios ingresados correctamente al sistema.

III. ANÁLISIS DE RESULTADOS

3.1. Validación del submodelo de reducción.

Para la validación del modelo de reducción (*relación funcional*), se tomaron los cinco modelos por adaptación y aprendizaje mencionados anteriormente. Cada modelo fue evaluado frente a la señal aleatoria de referencia (100 x 1) (ND) de acuerdo con la matriz de entrada que representa los (5) patrones de referencia (100 x 16). Para ello se utilizaron ocho métricas estadísticas que miden el desempeño de cada modelo frente a los datos: *Fractional Bias* (FB), *Normalized Mean Square Error* (NMSE), *Geometric Bias Mean* (MG), *Geometric Bias Variance* (VG), *Within a Factor of Two* (FAC2), *Index of Agreement* (IOA), *Unpaired Accuracy of Peak* (UAPC) and *Mean Relative Error* (MRE). Cada una de las métricas anteriores es descrita cualitativamente en términos de los valores alcanzados

por cada una de ellas: *Good* (G), *OverFair* (OF), *Fair* (F), *UnderFair* (UF), *Poor* (P). Para obtener el valor general de desempeño de un modelo (ID), Park *et al* (2007) propone una serie de valores que son aditivos de acuerdo con las cualidades tomadas por cada métrica: *Good* 7-10 (promedio 8,5), *Fair* 4-7 (promedio 5,5), *OverFair* (promedio 6), *UnderFair* (promedio 5) and *Poor* 1-4 (promedio 2,5). De esta manera, el máximo valor que puede alcanzar un modelo es de 68 puntos, el cual puede ser expresado porcentualmente tomando como referencia este valor máximo [25]:

$$ID = (NG * 8,5 + NF * 5,5 + NOF * 6 + NUF * 5 + NP * 2,5) / 68 \quad (10)$$

Donde:

ID: Índice de desempeño de un modelo frente a los datos de referencia.

NG: Número de métricas que lograron la cualidad *Good*.

NF: Número de métricas que lograron la cualidad *Fair*

NOF: Número de métricas que lograron la cualidad *OverFair*

NUF: Número de métricas que lograron la cualidad *UnderFair*

NP: Número de métricas que lograron la cualidad *Poor*.

En la Tabla 1 y Fig. 9 se muestran los resultados arrojadas por cada uno de los modelos frente a la reducción de dimensionalidad.

Tabla 1. Resultados de la evaluación de los modelos por adaptación y aprendizaje utilizados para la reducción de dimensionalidad.

Index	MADALINE		TKS		ANFIS		RNBR		EPR	
ND	100		100		100		100		100	
FB	0,050	G	0,100	G	0,136	G	0,111	G	0,114	G
NMSE	0,0262	G	0,032	G	0,063	G	0,056	G	0,067	G
MG	1,038	G	1,094	G	1,117	G	1,084	G	1,098	G
VG	1,027	G	1,030	G	1,044	G	1,041	G	1,057	G
FAC2	1	G	1	G	0,998	G	1	G	0,992	G
IOA	0,880	G	0,919	G	0,862	G	0,848	G	0,769	G
UAPC2	0,096	G	0,266	G	0,345	G	0,246	G	0,169	G
MRE	0,024	G	0,076	G	0,091	G	0,062	G	0,068	G
ID	100		100		100		100		100	

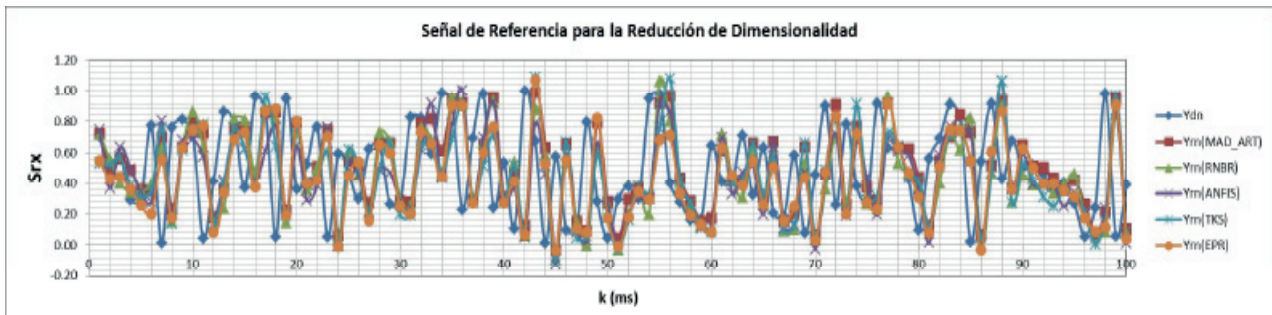


Fig.9. Respuesta de los modelos por adaptación y aprendizaje utilizados para la representación de la relación funcional establecida entre la matriz de entrada y la señal aleatoria de referencia (Reducción de Dimensionalidad).

De acuerdo con la Tabla 1 y la Fig. 9, podemos observar que el modelo que mostró el mejor comportamiento frente al índice de correlación (IOA) fue el modelo borroso de Takagi Sugeno, seguido del modelo neuronal Madaline. Si se observa este modelo, este mostró los mejores resultados frente a índices como el *Fractional Bias* (FB) y el UAPC2, los cuales miden la subestimación y sobrestimación de los datos por parte de los modelos. Si se observa el comportamiento de los modelos borrosos de tipo ANFIS y *Takagi Sugeno*, estos tienden a subestimar la serie aleatoria de referencia, debido, principalmente, a la forma de los conjuntos borrosos de tipo gaussiano que definen las señales de entrada al sistema.

Igualmente los índices VG, MG, MRE y NMSE, mostraron valores aceptables en cuanto a la reconstrucción de la serie. Por su parte, el modelo EPR no arrojó buenos resultados en la reducción ya que este depende fundamentalmente de los retardos en las señales, efecto que para este modelo no fue tenido en cuenta, ya que el reconocimiento no se considera como un proceso

dinámico en el tiempo. De acuerdo con el índice establecido por Park and Seok [34], el modelo neuronal MADALINE mostró una mejor flexibilidad frente a la reconstrucción de la señal aleatoria de referencia, a partir de las señales obtenidas del *hardware* como resultado de un movimiento gestual.

3.2. Análisis del submodelo de reconocimiento

Los resultados arrojados por el submodelo de reconocimiento frente al FS asociado con cada una de las neuronas de salida luego de 70 ciclos aleatorios de aprendizaje (Fig.10), mostró que este factor fue en aumento a medida que cada neurona de salida se especializaba en una letra o movimiento gestual en particular. Aunque el FS obtuvo valores diferenciales al final del aprendizaje, esto se debió al orden en que los patrones de aprendizaje fueron ingresados al sistema (A-E-I-O-U) en un instante inicial. No obstante, cada una de las neuronas de salida obtuvo FS cercanos al 80%, lo que muestra el buen comportamiento del modelo frente al reconocimiento de los patrones de referencia.

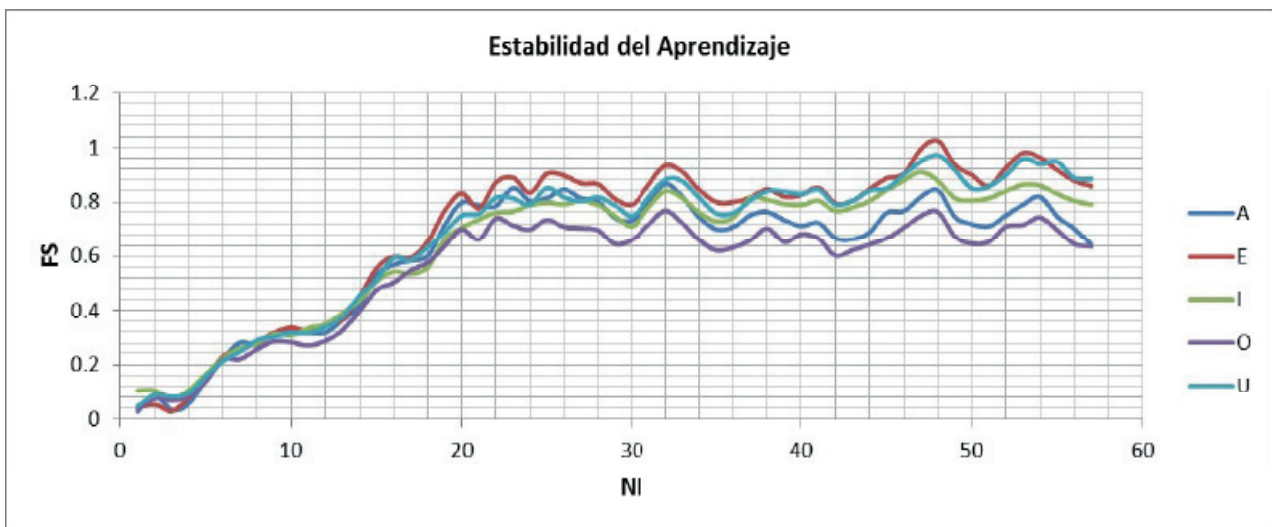


Fig.10. Comportamiento del factor de sensibilidad (FS) frente a cada una de las neuronas de salida (que representa cada uno de los movimientos gestuales o patrones de referencia vinculados a las vocales del alfabeto dactilológico en español).

3.3. Validación del modelo general.

Los resultados arrojados por el modelo general en cuanto a la efectividad en el reconocimiento se muestran en la Tabla 2.

Tabla 2. Matriz de confusión que muestra la efectividad del modelo propuesto frente a la identificación de 300 aciertos en el proceso de reconocimiento *on-line*

Modelo	Positivo (NI)	Negativo	Indiferente	Total	% Efectividad
ART_MAP	300	92	273	665	45,11%
F_ART_MAP	300	11	271	582	51,55%
RNBR	300	181	277	758	39,58%
MAD_ART2	300	7	77	384	78,13%

La matriz de confusión consolidada para la identificación de cada uno de los 300 aciertos logrados por el modelo propuesto para cada patrón de aprendizaje de acuerdo con el número total de aciertos se muestra en la Tabla 3.

Tabla 3. Matriz de confusión que muestra los resultados del reconocimiento para el modelo MAD_ARTx luego del ingreso de 300 vocales aleatorias.

Patrón	Patrón Reconocido						Total	PR	
	A	E	I	O	U	N/A			
Patrón Ingresado	A	50	3	2	2	2	3	62	80,65%
	E	55	5	3	4	3	5	75	73,33%
	I	79	3	3	4	4	4	97	81,44%
	O	65	3	3	4	3	4	82	79,27%
	U	51	3	3	2	5	4	68	75,00%
Total		300	17	14	16	17	20	384	

De acuerdo con la Tabla 2, podemos observar que el modelo MAD_ART2 logró los mejores índices de reconocimiento, ya que este modelo requirió de un menor número de intentos para lograr un total de 300 aciertos en el reconocimiento *on-line*. De esta misma manera, podemos observar que el modelo logró índices de desempeño cercanos al 75% en el reconocimiento, sobre cada uno de los movimientos gestuales que representan cada una de las vocales del lenguaje dactilológico en español. Para cada acierto el proceso de reconocimiento fue evaluado en términos del PR (Porcentaje de Reconocimiento), hasta llegar a la estabilización en el aprendizaje (Fig. 11).

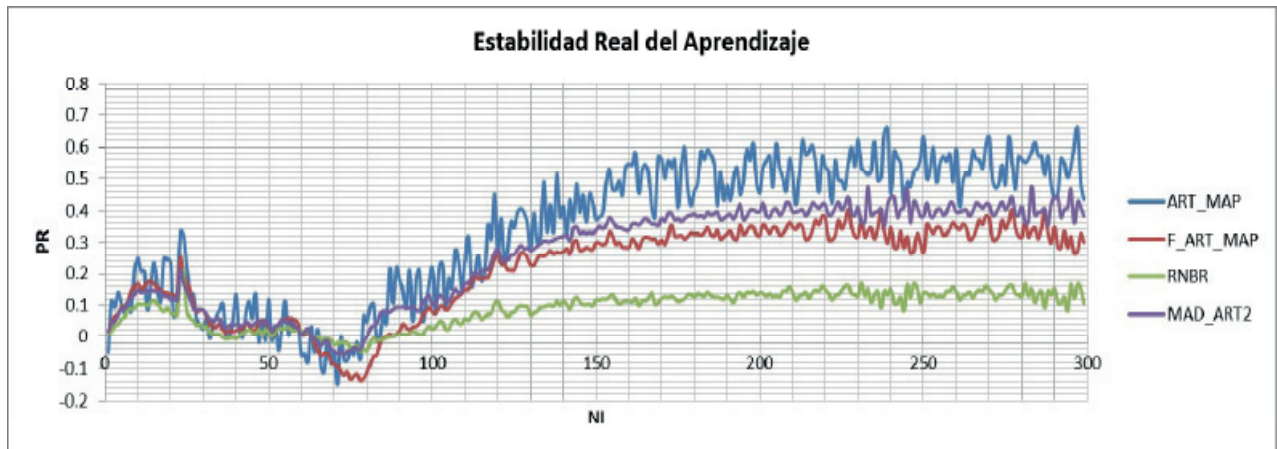


Fig.11. Desempeño del modelo propuesto (MAD_ART) frente a la estabilidad para el proceso de reconocimiento *on-line*.

En la Fig. 11 se puede observar que el modelo que logró un mayor valor para PR luego de la estabilización en el aprendizaje fue el modelo ART_MAP; no obstante este modelo tuvo fuertes variaciones frente a la estabilidad en el aprendizaje, hecho que se evidenció en el número de intentos requeridos para lograr un total de 300 aciertos (665 movimientos gestuales). Estas variaciones se debieron fundamentalmente a que un movimiento gestual no se hace de la misma forma, ni de la misma manera para una misma vocal, lo que generó una distorsión en el reconocimiento. Por su parte el

modelo MAD_ART2 logró un valor medio para PR algo menor al del modelo ART_MAP, pero con una reducción significativa de intentos, y con una mejor atenuación de las posibles variaciones en el reconocimiento de un movimiento gestual. Esta atenuación se hizo evidente en los modelos RNBR y Fuzzy_art_MAP, los cuales por su concepción, asumen mucho mejor las posibles variaciones en el ingreso de un movimiento gestual dado su carácter borroso. Sin embargo, estos modelos requieren de un mayor número de intentos para llegar a la estabilización en el reconocimiento (PR).

Es de anotar que el PR obtenido por el modelo Fuzzy ART_MAP luego de la estabilidad del aprendizaje estuvo muy cercano a los valores obtenidos por el modelo MAD_ART2 propuesto. Sin embargo, el modelo Fuzzy ART_MAP mostró una serie de desventajas relacionadas con el aumento de la dimensionalidad en las señales de entrada (20 x 32) lo que trajo consigo un aumento en la complejidad computacional asociada con el reconocimiento, y en general en el proceso por adaptación y aprendizaje. Por su parte el modelo que mostró los PR más bajos estuvo dado por el modelo neuronal de base radial (RNBR), debido principalmente a que el reconocimiento estuvo fuertemente influenciado por los pesos de salida de red, los cuales generaron grandes variaciones en la salida.

Debido a que el modelo propuesto incorpora un mecanismo por adaptación y aprendizaje de tipo no supervisado, se procedió al análisis del efecto que el FS genera sobre el PR, y en general sobre la creación de nuevas categorías para el reconocimiento (Fig. 12).

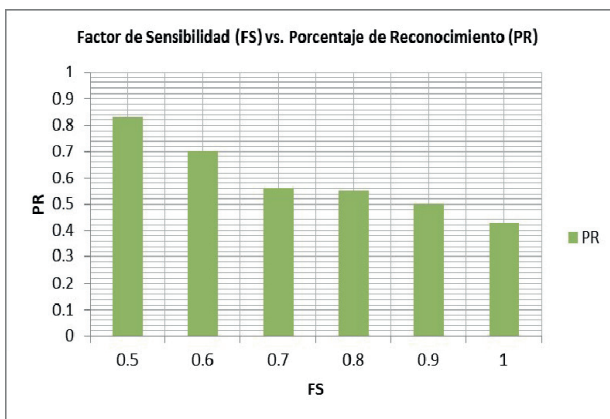


Fig. 12. Comportamiento del factor de sensibilidad (FS) frente al porcentaje de reconocimiento (PR) de acuerdo con el número de patrones ingresados de forma aleatoria al modelo.

En la Fig. 12, se puede observar que el PR logrado por el modelo general estuvo por encima del 80%, para un factor de sensibilidad de 0,50, lo que indica que para factores de sensibilidad muy bajos el modelo trata de asociar diferentes movimientos gestuales con una misma categoría, haciendo que la sensibilidad sea muy baja frente a la actualización de los pesos y frente a la creación de nuevas categorías. A medida que el FS fue en aumento el PR disminuyó hasta un 40%, lo que indica que a un mayor valor del FS, el modelo se ve presionado a la creación de una mayor cantidad de categorías, esto debido a que un movimiento gestual no se hace de la misma forma ni de la misma manera para una misma vocal.

IV. CONCLUSIÓN

El modelo general propuesto permitió el desarrollo de un sistema para el reconocimiento automático del lenguaje dactilológico en español, teniendo en cuenta para ello una integración entre un *hardware* de captura de los movimientos espacio-temporales de la mano y la utilización de modelos reconfigurables por adaptación y aprendizaje utilizados para el reconocimiento de patrones.

El buen funcionamiento del modelo propuesto estuvo determinado en gran medida por el submodelo neuronal de reducción de dimensionalidad, el cual utilizó una señal de referencia aleatoria obtenida a partir de una distribución normal para la integración de las señales de entrada que representan un movimiento gestual. Esto hizo que el modelo general (MAD_ART2) tuviera muy pocas distorsiones en el reconocimiento de un movimiento gestual sobre una única señal, lo que marcó una diferencia significativa con el modelo Fuzzy ART_MAP, en donde la señal de reducción tiende a aumentar de dimensionalidad por efecto de los conjuntos borrosos que definen cada una de las entradas del submodelo de reconocimiento.

Debido a la gran cantidad de movimientos gestuales utilizados para la validación del sistema de reconocimiento para un individuo sano, y al amplio rango de gestos temporales utilizados para la representación de una misma vocal, el sistema puede ser utilizado para cualquier individuo y para cualquier tipo de lenguaje dactilológico, esto dada su capacidad para asociar diferentes posiciones de la mano a un mismo movimiento gestual.

La eliminación del pulsador utilizado para la captura de un movimiento gestual se puede hacer mediante la implementación de comunicaciones mucho más rápidas que permitan identificar tiempos muertos entre movimientos gestuales. No obstante, un aumento en la velocidad de comunicación traerá consigo la utilización de modelos más robustos que permitan eliminar vibraciones (orden de 1,5 gravedades) generadas por la sensibilidad de los acelerómetros. Igualmente, esta sensibilidad puede ser manejada a través de sensores basados en unidades inerciales *IMU* [29].

Debido a que un movimiento gestual no se realiza de la misma forma ni de la misma manera, el logro de porcentajes de reconocimiento más allá del 90% para el modelo propuesto requerirá de la utilización de algoritmos de optimización que mejoren el proceso de aprendizaje en línea. Es por esto que los algoritmos de estimación de la distribución (EDA's) surgen como una alternativa para dar mayor robustez al modelo general de reconocimiento en tiempo real.

REFERENCIAS

- [1]. INSOR - Observatorio Social, población sorda colombiana. Registro para la localización y caracterización de personas con discapacidad: Número de personas sordas que trabajan según sexo y actividad económica, 2009. Consultado el 13 de marzo de 2011: <http://www.observatorio.insor.gov.co/index.shtml?apc=1;;a;1;;;02&x=3414>
- [2]. Universidad de Alicante. Notas de clase: Lenguaje y Comunicación. Departamento de Psicología de la Salud, 2007. Consultado el 19 de noviembre de 2011 en <http://rua.ua.es/dspace/bitstream/10045/4298/6/TEMA%206.LENGUAJE%20Y%20COMUNICACION%20C3%93N.pdf>
- [3]. García I. S. Lenguaje de señas entre niños sordos de padres sordos y oyentes. Tesis de Grado, Licenciatura en Lingüística. Universidad Nacional Mayor de San Marcos, Lima, Perú, 2002.
- [4]. Cuxac C. Les langues des signes: une perspective sémiogénétique. Acquisition et interaction en langue Étrangère, 2005. Consultado el 22 de enero de 2013: <http://aile.revues.org/document495.html>
- [5]. Padden C.A. Folk Explanation in Language Survival in: Deaf World: A Historical Reader and Primary Sourcebook. New York: University Press, 2001. ISBN: 0-8147-9853-5.
- [6]. Pérez D.H. Organización Colombia Incluyente. Con "Julis" las personas sordas mejoran su comunicación, 2010. Consultado el 22 de enero de 2013: <http://colombianincluyente.org/vernot.php?id=21>
- [7]. XTEC - Xarxa Telemàtica Educativa de Catalunya. Las Lenguas de Signos. La expresión natural de las personas sordas, (s.f.) Consultado el 22 de enero de 2013: <http://www.xtec.cat/~cllomar/espanol/llenguasignes/llenguasignes.htm>
- [8]. OIT - Organización Internacional del Trabajo, Oficina de Actividades para los Empleadores y Departamento de Conocimientos Teóricos y Prácticos y Empleabilidad. Discapacidad en el lugar de trabajo. Prácticas de las Empresas. Ginebra, Suiza, 2010. ISBN:978-92-2-123871-3
- [9]. Wu Y., Huang T.S. Vision-Based Gesture Recognition: A Review. *Gesture-Based Communication in Human-Computer Interaction. Lecture Notes in Computer Science*. Springer. 1739, 103-115, 1999. ISBN 978-3-540-66935-7, doi:10.1007/3-540-46616-9
- [10]. Stokoe W.C. Sign language structure: An outline of the visual communication systems of the american deaf. *Journal of Deaf Stud. Deaf Educ.* 10(1), 3-37, 2005. ISSN:1081-4159.
- [11]. Kelly D., McDonald J., Markham C. A Person Independent System for Recognition of Hand Postures Used in Sign Language. *Pattern Recognition Letters*. 31, 1359-1368, 2010. doi:10.1016/j.patrec.2010.02.004
- [12]. Boulares M., Jemni M. 3D Motion Trajectory Analysis Approach to Improve Sign Language 3D-based Content Recognition. *Procedia Computer Science*. 13, 133-143, 2012. doi: 10.1016/j.porsc.2012.09.122.
- [13]. Kin Fun Li, Lothrop K., Gill E., Lau S. Web based sign language traslator using 3D video processing. *Proceedings of the 14th. International Conference on Network-Based Information Systems (NBIS)*, Tirana, Albania, septiembre 2011.
- [14]. Bartneck C., Rauterberg M. HCI Reality - An Unreal Tournament? *International Journal on Human-Computer Studies*. 65(8), 737-743, 2007.
- [15]. ABC Televisión. The Power Glove. The Good Game, 2008. Consultado el 22 de enero de 2013. <http://www.abc.net.au/tv/goodgame/stories/s2248843.htm>
- [16]. Wikipedia, The Free Encyclopedia. Wired Glove, 2012. Consultado el 22 de enero de 2013. http://en.wikipedia.org/wiki/Wired_glove
- [17]. LEMELSON-MIT. Inventor of the Week Archive: Ryan Pattersen, American Sign Language Translator, 2006. Consultado el 22 de enero de 2013. <http://web.mit.edu/invent/iow/patterson.html>
- [18]. The Lemelson Center for the Study of Invention & Innovation. AccedeGlove: Hernández-Rebollar J. Patente nº 11.836.136, 2007. Estados Unidos de América. Consultado el 22 de enero de 2013. http://invention.smithsonian.org/resources/fa_rebollar_index.aspx
- [19]. Vogler C., Metaxas D. Handshapes and Movements: Multiple-Channel American Sign Language Recognition. *Gesture-Based Communication in Human-Computer Interaction. Lecture Notes in Computer Science*. Springer, 2915, 247-258. 2004.
- [20]. Moreno J., Ruiz D. Informe Técnico: Protocolo Zigbee (IEEE 802.15.4). Universidad de Alicante, 2007. Consultado el 21 de enero de 2013. http://rua.ua.es/dspace/bitstream/10045/1109/7/Informe_ZigBee.pdf
- [21]. Oliva M.L., Alemán A.M., Olvera A.M. El Análisis de Componentes Independientes y sus Aplicaciones, 2013. Consultado el 21 de enero de 2013. <http://www.boletin.upiita.ipn.mx/boletin21/cyt/ica.htm>
- [22]. Isazi P., Galván I.M. Redes de Neuronas Artificiales. Un Enfoque Práctico. Madrid: Pearson Education, S.A., 2004. ISBN:84-205-4025-0.
- [23]. Zuñiga A., Jordán C. Pronóstico de Caudales Medios Mensuales Empleando Sistemas Neurofuzzy. *Revista Tecnológica ESPOL*. 18(1), 17-23, 2005. ISSN:0257-1749.
- [24]. Peña M., Alvarez H., Carelli R. Modelado e Identificación con un Modelo Borroso del Tipo Takagi Sugeno. *V Jornadas de Sistemas de Instrumentación y Control*. Caracas, Venezuela, 2000.
- [25]. Peña A., Hernández, R. Compression of Free Surface Based on the Evolutionary Optimization of A NURBS Takagi Sugeno. *Proceedings 23rd. ISPE International Conference on CAD/CAM, ROBOTICS & Factories of the Future*, Bogotá. 100-117, 2007. ISBN: 978-958-978-597-3.
- [26]. Xamarin. mono project, 2013. Consultado el 24 de junio de 2013. http://www.mono-project.com/Main_Page
- [27]. Solera R. Máquinas de Vectores Soporte para Reconocimiento Robusto del Habla. Tesis de Grado, Doctorado Interuniversitario en Multimedia y Comunicaciones. Madrid, España. Universidad Carlos III de Madrid. 2011.
- [28]. Collobert R., Bengio S. SVM Torch: Support Vector Machines for Large Scale Regression Problem. *Journal of Machine Learning Research* 1, 143-160, 2001.
- [29]. Champion J. A flexible charting library for .NET. The Code Project, 2007. Consultado el 22 de enero de 2013. <http://www.codeproject.com/KB/graphics/zedgraph.aspx>
- [30]. Giustolisi O., Doglioni A., Savic D., Webb B. A Multimodel Approach to Analysis of Environmental Phenomena. *Environmental Modelling & Software* 22(5), 674-682, 2007. doi: 10.1016/j.envsoft.2005.12.026.

- [31]. Basogain X. Redes Neuronales Artificiales y sus Aplicaciones. OpenCourseWare - Universidad del País Vasco, 2008. Consultado el 22 de enero de 2013. <http://ocw.ehu.es/enseanzas-tecnicas/redes-neuronales-artificiales-y-sus-aplicaciones/contenidos/pdf/libro-del-curso>
- [32]. Liu W., Seto K.C., Wu E.Y., Gopal S., Woodcock C.E. ART-MMAP: A Neural Network Approach to Subpixel Classification. *IEEE Transactions on Geoscience and Remote Sensing* 42(9), 1976-1983, 2004.
- [33]. Zhe L. Fuzzy ARTMAP Based Neurocomputational Spatial Uncertainty Measures. *Photogrammetric Engineering & Remote Sensing* 74(12), 1573-1584, 2008.
- [34]. Park O.H., Seok M.G. Selection of Appropriate Model to Predict Plume Dispersion in Coastal Areas. *Atmospheric Environment*. (41), 6095-6101, 2007. doi: 10.1016/j.atmosenv.2007.04.010.