

“Desarrollo e Implementación de una Aplicación que Traduzca el Abecedario y los Números del uno al diez del Lenguaje de Señas a Texto para Ayuda de Discapitados Auditivos Mediante Dispositivos Móviles Android.”

María Gabriela Vintimilla Sarmiento
e-mail: gabriela_8724@hotmail.com

Darwin Alulema
e-mail: doalulema@espe.edu.ec

Fabian Saenz
e-mail: fgsaenz@espe.edu.ec

*Departamento de Eléctrica y Electrónica carrera de Ingeniería Electrónica en Redes y
Comunicación de Datos
Sangolquí, Ecuador*

RESUMEN

Este artículo se basa en el diseño e implementación de una aplicación para dispositivos móviles con sistema operativo Android, que permita la interrelación de las personas con discapacidad auditiva. La aplicación es capaz de aprender y reconocer una letra o número del lenguaje de señas que no tengan movimiento mediante la aplicación de redes neuronales artificiales, la aplicación indicará si la imagen captada pertenece a la letra que se desea reconocer, en caso de que la imagen no pertenezca a la letra correspondiente, la aplicación muestra un mensaje de error.

I. INTRODUCCION.

Las tecnologías móviles se consideran el sector de mayor auge dentro del ámbito tecnológico actual. Es por ello que un gran

número de novedades tecnológicas se centran en la telefonía móvil.

El móvil, hoy en día, ya no se considera únicamente un dispositivo para realizar llamadas, sino que posee múltiples usos cotidianos como puede ser el envío de mensajes, conexión a internet, reproducción de música, etc.

En cierto modo, el móvil se ha convertido en un ordenador personal. En la actualidad, se desarrollan nuevas tecnologías para la integración de las personas con discapacidad para así alcanzar una sociedad "plural" y "sin brechas digitales"; con el tiempo esta integración ha ido mejorando y se ha conseguido una mayor interacción en el uso del dispositivo móvil y las personas discapacitadas.

II. LENGUAJE DE SEÑAS.

El lenguaje de las señas es un lenguaje natural de expresión gesto espacial y de

percepción visual. Gracias a la lengua de señas las personas con deficiencias auditivas pueden establecer un canal de comunicación con su entorno social, el cual puede estar conformado por otras personas con las mismas deficiencias o por cualquier persona que conozca la lengua de señas empleada.

En casi todo el mundo se usan señas para representar las letras del alfabeto con el que se escribe la lengua oral del país. Esto se denomina alfabeto manual o alfabeto dactilológico. En el caso de los países de habla hispana donde se usa el alfabeto latino, las personas sordas usan un mismo alfabeto manual, común para todos los países con algunas variaciones de índole menor.

La representación manual de las letras del abecedario y de los números del uno al diez utilizados en la educación de personas sordas se muestra en la figura 1 y 2.

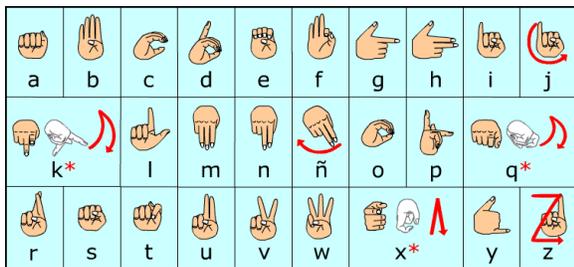


Figura 1. Abecedario.



Figura 2. Números del uno al diez.

III. REDES NEURONALES.

Las Redes Neuronales Artificiales están inspiradas en las redes neuronales biológicas del cerebro humano.

Están constituidas por elementos que se comportan de forma similar a la neurona biológica en sus funciones más comunes y presentan una serie de características propias del cerebro humano. Por ejemplo aprenden de la experiencia, generalizan de ejemplos previos a ejemplos nuevos y pueden ofrecer, dentro de un margen, respuestas correctas a entradas que presentan pequeñas variaciones debido a los efectos de ruido o distorsión.

Una de las principales características de las redes neuronales es su capacidad de aprendizaje a partir de un entrenamiento previo.

El objetivo del entrenamiento es conseguir que la aplicación, para un conjunto de entradas produzca el conjunto de salidas deseadas o mínimamente consistentes.

El algoritmo utilizado para el entrenamiento es el algoritmo Back Propagation, que consiste en el emparejamiento de cada vector de entrada con su correspondiente vector de salida, es decir se presenta un vector de entrada a la red, se calcula la salida de la red, se compara con la salida deseada, y el error o diferencia resultante se utiliza para realimentar la red y cambiar los pesos de acuerdo con el algoritmo que tiende a minimizar el error.

El algoritmo Back propagation es un método de entrenamiento de redes multicapa. Su potencia reside en su capacidad de entrenar capas ocultas.

En la aplicación del algoritmo back-propagation, se distinguen dos pasadas distintas de cómputo. La primera pasada se

refiere como la pasada hacia delante, y la segunda como la pasada hacia atrás.

En la pasada hacia delante los pesos permanecen inalterados por toda la red, y las señales funcionales de la red se computan neurona por neurona.

En la pasada hacia atrás los cálculos de las modificaciones de todos los pesos de las conexiones empiezan por la capa de salida y continua hacia atrás a través de todas las capas de la red hasta la capa de entrada.

IV. DESARROLLO DE LA APLICACIÓN.

El objetivo del proyecto es llegar a reconocer las letras del lenguaje de señas, para lo cual existen dos fases en toda aplicación de las redes neuronales:

- Fase de entrenamiento: Se usa un conjunto de datos o patrones de entrenamiento para determinar los pesos que definen el modelo neuronal, estos datos son enviados a un servidor de base de datos que se encuentra alojado en la nube.
- Fase de prueba o funcionamiento directo: Una vez entrenado este modelo, se procesan los patrones de prueba que constituyen la entrada habitual de la red.

El diseño de la aplicación consiste en tres capas como se muestra en la figura 3:

- La capa del cliente, donde se encuentra la aplicación instalada en el móvil.
- La capa de base de datos MySQL, que se encuentra en la nube que está conformada por dos tablas, la primera para almacenar las imágenes capturas después de un tratamiento previo de la imagen, y la segunda para capturar los valores de los pesos generados en modo de aprendizaje.

- La capa de Servidor de inteligencia artificial, en donde se ejecuta el algoritmo de Back Propagation.

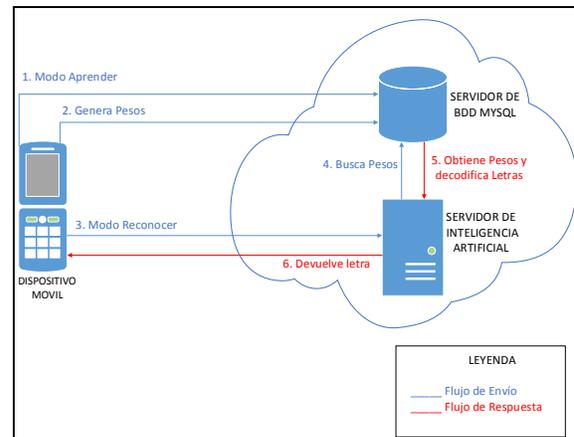


Figura 3. Diagrama de arquitectura Talk to Sign

5.1.Fase de entrenamiento:

Para la fase de entrenamiento se utilizan dos modos en la interfaz de la aplicación que se muestra en la figura 4:

- Modo aprender.
- Modo Genera pesos.

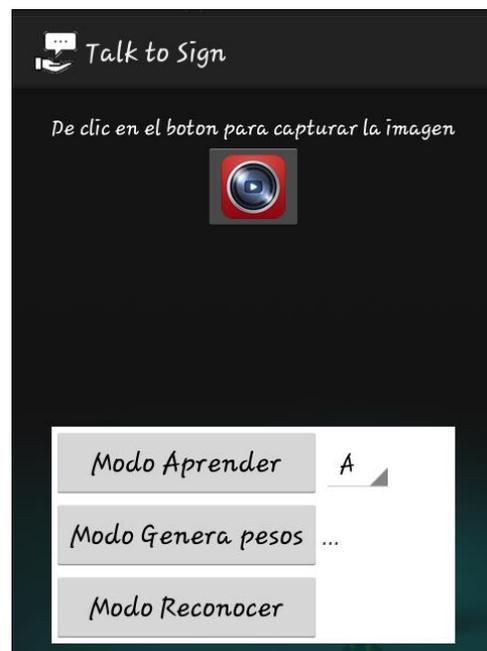


Figura 4. Interfaz gráfica de la aplicación.

El objetivo del entrenamiento es llegar a tener una matriz de bits de la imagen original, almacenarlas en la base de datos y generar los pesos respectivos de cada letra.

Para lograr este objetivo se realiza un tratamiento a la imagen, el tratamiento de la imagen consiste en:

- Obtención de escala de grises: se realiza la conversión de la imagen a escala de grises, con este proceso se obtiene una imagen en una escala comprendida en el intervalo [0,255], donde 0 es negro y 255 es blanco.

$$Y = 0.299 * R + 0.587 * G + 0.144 * B$$

Donde:

R= Rojo
G= Verde
B= Azul

- Binarización: proceso tras el cual se obtiene una imagen binaria (negro 0 y blanco 255), este proceso se lo realiza para descartar colores intermedios, y tener una imagen en dos rangos únicamente. Este proceso se lo realiza mediante un umbral con valor de 40, este valor fue obtenido mediante pruebas.

$$pixel[i] = \begin{cases} Max & \text{if } pixel[i] > threshold \\ 0 & \text{otherside} \end{cases}$$

Pixel[i] da el valor de gris del pixel número i,

siendo i ∈ [0, tamaño_imagen]

- Reducción: con el objetivo de optimizar la entrada para la red neuronal se reduce su tamaño a 40x40 píxeles, de esta manera no dependerá del tamaño de la imagen capturada, porque todas serán reducidas para así obtener una imagen tratable para la red neuronal, este valor se determinó mediante pruebas y se pudo observar que

era el valor mínimo para reducir una imagen, ya que en este tamaño se podía distinguir las letras similares.

- Obtención de características: obtención de un array de 0 y 1 que contiene los píxeles de la imagen tratada. El array está conformado por 1600 bits obtenidos de la imagen de 40x40 ya binarizada. Este array servirá como entrada a la red neuronal.

La reducción del tamaño de la imagen, se realiza para conseguir un array tratable para la red neuronal para que la aplicación sea rápida y no tenga un retardo muy grande.

La generación de pesos se realiza en función al número de muestras almacenadas en la base de datos. Mientras mayor sea el número de muestras se tendrá una mayor presión en el reconocimiento, este peso se almacenará en otra tabla de la base de datos.

5.2.Fase de prueba o funcionamiento directo:

Una vez terminada la fase de entrenamiento se procede a la fase de pruebas.

Internamente la red neuronal está formada por la capa de entrada que son el producto de la imagen binarizada, a continuación los datos ingresan a las tres capas ocultas de forma secuencial en la red neuronal, y por último la capa de salida, que contará con 8 unidades devolviendo un valor en cada una de ellas entre 0 y 1. Este vector de 8 componentes hace referencia al valor binario de la letra o número que se puede observar en la tabla número 1.

Tabla 1. Letras y números del alfabeto en binario.

LETRAS		NUMEROS	
A	01000001	1	00000001
B	01000010	2	00000010
C	01000011	3	00000011
D	01000100	4	00000100
E	01000101	5	00000101
F	01000110	6	00000110
G	01000111	7	00000111
H	01001000	8	00001000
I	01001001	9	00001001
L	01001100	10	00001010
M	01001101		
N	01001110		
O	01001111		
P	01010000		
R	01010010		
S	01010011		
T	01010100		
U	01010101		
V	01010110		
W	01010111		
X	01011000		
Y	01011001		

V. RESULTADOS

Se realizaron pruebas para la letra A, para lo que se capturó letras A correctas y letras incorrectas. Se debe tener en cuenta que el escenario para las pruebas fue un fondo negro y sin brillo debido a que puede provocar inconvenientes al realizar el tratamiento de la imagen, generando unos y veros falsos en la binarización de la imagen.

La tabla numero 2 resume el resultado de las pruebas obtenidas.

Tabla 2. Tabla de resultados de pruebas.

Letra Seleccionada	Letra Capturada	Respuesta
A	A	LETRA A
A	O	TRAMA NO CONOCIDA
A	R	TRAMA NO CONOCIDA
A	MUESTRA CUALQUIERA	LETRA A
A	5	TRAMA NO CONOCIDA
A	MUESTRA CUALQUIERA	TRAMA NO CONOCIDA

VI. CONCLUSIONES

- La aplicación es capaz de reconocer símbolos a partir de imágenes previamente aprendidas, en caso de no reconocer la imagen tomada, la aplicación no devolverá ninguna letra, por lo que podría ser la base para una aplicación que reconozca palabras completas.
- No todo el conjunto de símbolos del lenguaje de signos internacional ha sido tratado, debido a que los símbolos de las letras como como la j, k, ñ, q, x, z necesitan de movimiento por lo que se debía realizar tomas en video y adaptar el sistema para que sea capaz de identificar movimiento.
- La variación en la iluminación y el color de fondo pueden provocar dificultad para reconocer las imágenes porque el entrenamiento no aporta suficiente información a la red. Esto se debe a que al momento de realizar el tratamiento de la imagen el brillo o un color claro de fondo se van a mostrar como "negro" y la aplicación responde como error.
- La aplicación está formada por tres niveles, un nivel constituye la aplicación en el dispositivo móvil, los otros dos

niveles se encuentran en la nube, el de inteligencia artificial y el de Base de datos, de esta manera se si estos se implementaban en el dispositivo móvil su tiempo de procesamiento aumentaría, y por ende su tiempo de respuesta.

VII. REFERENCIAS

- [1] Sordos Ecuador. (s.f.). *Lenguaje de señas*. Recuperado el 23 de enero de 2014 <http://www.sordosecuador.com/lengua-de-senas/>
- [2] Fiszelew, A. (2002). *Generación automática de redes neuronales con ajuste de parámetros basado en algoritmos genéticos*: Universidad de Buenos Aires.
- [3] Hilera, J.R. (2000). *Nuevas técnicas de modelización y predicción de fenómenos complejos*. Recuperado el 10 de marzo de 2014 de *Redes Neuronales Artificiales y Algoritmos Genéticos*. Alcalá de Henares, España.
- [4] Andes. (2012). *Vicepresidencia del Ecuador impulsa el primer Diccionario Oficial de Lengua de Señas Ecuatoriana*. (2012). Recuperado el 10 de marzo de 2014 de <http://www.andes.info.ec/es/sociedad/7820.html>