Implementación de un algoritmo para la identificación de usuarios considerando problemas fisiológicos que afectan el habla

Implementation of an algorithm for users identification considering physiological problems affecting the speech

Diego Enrique Rey-Lancheros
Ingeniero Electrónico
Universidad Distrital Francisco José de Caldas
Bogotá, Colombia.
diegoerey@gmail.com

Hernán Julian Gavilán-Acosta
Ingeniero Electrónico
Universidad Distrital Francisco José de Caldas
Bogotá, Colombia.
hjga16@gmail.com

Helbert Eduardo Espitia-Cuchango
Ingeniero Electrónico
Universidad Distrital Francisco José de Caldas
Bogotá, Colombia.
heespitiac@udistrital.edu.co

Resumen– En este documento se presenta el diseño e implementación de un algoritmo para la identificación de usuarios por medio de voz, considerando problemas fisiológicos que afectan el habla de tal forma, que al presentarse este tipo de problemas en los usuarios se logre una baja tasa de falsos rechazos. Para el diseño del sistema se toma como base un algoritmo estándar que utiliza coeficientes cepstrales, al cual se le incorporan otras características que son definidas mediante un análisis acústico de la voz; de esta forma se puede manejar los problemas fisiológicos considerados. Con el fin de observar el desempeño del algoritmo, se lleva a cabo una prueba con varios archivos, tanto de personas sanas como también con afectaciones de la voz, observando que al incorporar las características establecidas del análisis de la voz, se logra un mejor resultado en relación con el caso donde solo se emplean coeficientes cepstrales.

Palabras clave– Identificación, problemas fisiológicos, voz.

Abstract- This paper shows the design and implementation of an algorithm for users voice identification, including considerations on physiologic issues affecting the speech so that when users manifest these problems, lower rates on fake rejections decrease. For purposes of managing the contemplated physiologic problems, algorithm design also takes a standard algorithm that uses cepstral coefficients which include additional characteristics determined by voice acoustic analysis. A test including several records from people with a healthy voice and those with voice affections is carried out in order to observe the performance of the algorithm; thus, observing that when applying those characteristics in voice analysis a better result is achieved regarding the case when cepstral coefficients are implemented.

Keywords– Identification, physiological problems, voice.

1. INTRODUCCIÓN

Los sistemas de seguridad actuales, utilizados tanto en ambientes empresariales así como en el hogar, tienden a basarse en el uso de contraseñas o reconocimiento de características físicas únicas como huella digital y reconocimiento de iris, entre otras. Para sistemas de mayor seguridad se aplican de manera combinada las técnicas de verificación como la huella digital y la de voz [1]. Particularmente el método de identificación de usuarios por voz presenta problemas bastante complejos y de una gran variedad [2].

Según [3] algunos de los factores más importantes por tener en cuenta para el análisis de la señal generada por la voz son:

Recibido: 06/12/2017 Aceptado: 07/04/2017 ITECKNE Vol. 14 Número 2 • ISSN 1692-1798 • ISSN Digital 2339-3483 • diciembre 2017 • 131 - 139
http://dx.doi.org/10.15332/iteckne.v14i2.1767

- El manejo de variaciones en la señal entre diferentes usuarios, ya sea con alta o baja variabilidad y con voces más parecidas entre sí.
- Manejo del ruido inducido por el ambiente.
- Diferencias entre volúmenes en la voz de los usuarios del sistema.
- Problemas de la salud de los usuarios en procesos de identificación.
- Análisis de cada uno de los máximos y mínimos individuales en la señal.
- Identificación de impostores y simulación de patrones del habla.

1.1 Sistemas de reconocimiento de voz

De acuerdo con [4] la voz contiene grandes cantidades de información, incluyendo género, sentimientos, un mensaje, una identidad e incluso estado físico. Usualmente es sencillo para un humano escuchar estos factores independientemente; por ejemplo, una mujer de noventa años y un hombre cansado se pueden diferenciar sin importar que pronuncien el mismo mensaje [5]. Esta información adicional, que no es necesaria para la transmisión del mensaje, puede ser utilizada para estimar la credibilidad, relevancia y otras propiedades del mensaje.

Desde una perspectiva computacional dos campos de investigación han recibido gran atención en los últimos años; estos son reconocimiento de lo que se habla (Speech recognition) y reconocimiento de quien está hablando (Speaker recognition) [6].

Según [7] para el reconocimiento del usuario se suelen emplear diferentes mediciones de Jitter y Shimmer. Por su parte en [8], se señala que existe un crecimiento en los enfoques híbridos para una caracterización exitosa de la voz, particularmente en aplicaciones forenses, utilizando las características de la grabación de la señal como también de las obtenidas a través de filtros.

1.2 Reconocimiento de voz e identificación de usuarios

El reconocimiento de voz se basa en identificar las palabras pronunciadas por una persona. Esta identificación debe ser tolerante a diferentes factores como dialectos, malestar, edad, entre otros. La información importante es el mensaje que está siendo entregado [6]. Por su parte, la identificación de usuarios busca establecer la identidad de la persona que está hablando. El significado de lo que se está diciendo, para este caso, no es importante y es descartado. En este campo de investigación los dialectos son un gran apoyo, puesto que ayudan a distinguir entre las personas. Sin embargo, tal como en el caso de reconocimiento de voz, esta identificación debe ser tolerante a factores como malestar, edad, ruido ambiental, etc., los cuales le pueden dificultar al sistema hacer la identificación [6].

En relación con el reconocimiento de usuarios, se presenta una división en dos áreas de estudio generales, las cuales son:

Verificación de usuario: donde la entrada es un segmento de audio y un locutor. El problema al que se busca dar solución es verificar si el locutor pronunció o no dicho segmento. Esto es llamado una verificación 1:1, puesto que la comparación se hace solamente contra uno de los registros de voz almacenados.

Identificación del usuario: donde la entrada es un segmento de audio y una cantidad ρ de usuarios registrados. El problema al que se le busca solución es determinar a cuál de los usuarios registrados pertenece dicho segmento. Esto es llamado una identificación 1:ρ, puesto que la comparación se hace contra cada uno de los ρ registros de voz almacenados [9].

Sobre alternativas en el desarrollo de trabajos de identificación de usuarios, en [10] se describe la forma para construir un sistema independiente del texto, empleando coeficientes cepstrales y una máquina de soporte vectorial. Adicionalmente en [11] se tiene un enfoque donde se trata de establecer el género de una persona a partir de su voz, lo cual facilitaría la autenticación e identificación de usuarios en sistemas de alta seguridad. En este artículo se consideran tres características, las cuales son: autocorrelación, energía de señal y coeficientes cepstrales. También se utilizó una

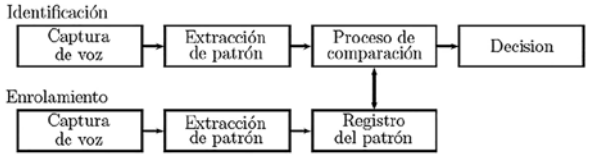
máquina de soporte vectorial lineal para la clasificación de características extraídas de la señal de voz. Un enfoque alternativo consiste en el reconocimiento automático de fenómenos paralinguísticos acústicos, los cuales se pueden utilizar para la identificación de usuarios mediante el habla. En [12] el método propuesto reconoce automáticamente fenómenos paralingüísticos como gritos, hiperarticulación y vacilación durante la interacción entre el usuario y el sistema, de esta forma la caracterización se realiza en función de la ocurrencia de estos fenómenos.

Como es de apreciar, existen diferentes enfoques y alternativas que pueden incrementar la complejidad del sistema de identificación. En consecuencia, este trabajo muestra el diseño de un sistema para la identificación de usuarios buscando implementar un algoritmo compacto que incluya los elementos más relevantes de la identificación de usuarios con patologías que afectan la voz. Para esto se toma como referente un algoritmo general de identificación de usuario basado en coeficientes cepstrales y se incorporan características obtenidas del análisis acústico de la voz como son los picos de la señal, el Shimmer y la frecuencia fundamental. Estas características se establecen teniendo presente algunos problemas fisiológicos que pueden afectar el habla, buscando que el algoritmo logre reconocer el usuario cuando se presente variación en su voz. Este algoritmo en futuros desarrollos se puede implementar en sistemas de seguridad como la identificación de usuarios en dispositivos móviles.

2. DESCRIPCIÓN DEL SISTEMA DE RECONOCIMIENTO

Un modelo genérico de reconocimiento por voz se divide en dos etapas, el enrolamiento inicial y la identificación [13]. El proceso de enrolamiento consiste en adquirir una muestra de voz de un usuario, luego dicha muestra es procesada con el fin de obtener las características que formarán su patrón, para finalmente ser almacenadas. La identificación consiste en obtener otra muestra de voz del mismo usuario para extraer el patrón de voz de la misma forma que en el enrolamiento, sin embargo, en lugar de almacenar el nuevo patrón, este es comparado contra los que ya habían sido almacenados previamente con el fin de buscar coincidencias y así poder establecer la identidad del usuario a quien corresponde la voz, este modelo se muestra en Fig.1.

Fig. 1. MODELO GENÉRICO PARA EL RECONOCIMIENTO POR VOZ



Fuente: Los autores.

Como es de apreciar en la Fig. 1, tanto para la identificación como para el enrolamiento se realiza la captura de la voz con el respectivo equipo de adquisición, luego se extrae el patrón o las características de la voz de un usuario, en el enrolamiento se registran las características para luego ser comparadas con una grabación de entrada, de tal forma que se pueda establecer si el usuario se encuentra registrado en la base de datos, lo cual corresponde al proceso final asociado a la decisión. Considerando los elementos involucrados en el sistema de identificación, el respectivo pseudocódigo del algoritmo se puede apreciar en la Fig. 2.

Fig. 2. PSEUDOCÓDIGO DEL ALGORITMO DE IDENTIFIZACIÓN DE USUARIO POR VOZ

Algoritmo 1: Identificación de usuario por voz.

```
Previamente realizar el proceso de enrolamiento de ρ usuarios;
begin
    Adquirir el archivo de voz del usuario a identificar;
    Extraer las características del archivo de entrada;
    for los archivos almacenados ρ do
        Comparar con las características del registro de entrada;
    end
    Establecer si el archivo de entrada corresponde a un registro
    almacenado;
end
```

Fuente: Los autores.

Estos pasos son generales en un sistema de identificación de usuario siendo de particular interés en este artículo las características que se emplean para poder caracterizar los usuarios.

Un enfoque tradicional para la identificación de usuarios consiste en la caracterización mediante coeficientes cepstrales, por lo cual se toma como base para la implementación del algoritmo de reconocimiento; adicionalmente se busca emplear otras características considerando el análisis de la voz para que el algoritmo pueda identificar

usuarios que presenten alguna enfermedad respiratoria.

3. OBTENCIÓN DE CARACTERÍSTICAS MEDIANTE COEFICIENTES CEPSTRALES

La voz se puede representar mediante un espectro en frecuencia que proporciona información de los parámetros de producción de la misma. Como características principales de cada registro de voz se calculan los coeficientes cepstrales, estos parámetros tienen la ventaja de ser invariantes frente a distorsiones que puedan ser introducidas por elementos externos como el micrófono o sistemas de transmisión [13].

De acuerdo con el modelo propuesto para un algoritmo genérico de identificación por voz [13], el proceso de obtención de características se realiza mediante los siguientes pasos:

Eliminar silencios de la señal: se evalúan segmentos de 10ms de la señal y se calcula la energía del segmento, si es menor que el promedio de energía de la señal, entonces es descartado.

Aplicar filtro de énfasis: se aplica un filtro a la señal para enfatizar las frecuencias de los formantes con el fin de evitar la pérdida de información durante el proceso de segmentación. La ecuación 1 muestra el filtro utilizado en el algoritmo base.

$$H(z) = 1 - 0.95z^{-1} \qquad (1)$$

Segmentación: para analizar la señal, esta es dividida en secciones donde, para cada sección, se asume que es estacionaria. Se aplica un ventaneo de Hamming de 30ms cada 10ms.

Obtención de coeficientes cepstrales: el proceso de obtención de coeficientes se ilustra en Fig. 3, donde se aprecia la manipulación de la señal en el dominio de la frecuencia para luego efectuar el cálculo de los respectivos coeficientes. Para esto se emplea la transformada rápida de Fourier FFT y su inversa IFFT, también se realiza el cálculo de la magnitud y el logaritmo.

Fig. 3. MODELO DE OBTENCIÓN DE COEFICIENTES CEPSTRALES

Fuente: Los autores.

Obtención de características: se obtienen mediante la normalización de coeficientes. Este proceso se lleva a cabo para reducir variabilidades espectrales en largos periodos de tiempo. Los coeficientes se expanden utilizando una representación polinomial ortogonal en intervalos de 90ms cada 10ms. Según [13] solamente se utilizan los dos primeros coeficientes.

4. PATOLOGÍAS CONSIDERADAS

En la valoración de la calidad de voz para un contexto clínico, las mediciones acústicas son aplicadas como apoyo del diagnóstico [14]. El análisis acústico de la voz, es decir, el estudio de las propiedades del sonido considerado como una vibración que se propaga y a un mayor detalle, la señal que genera dicho sonido y sus características como frecuencia, amplitud, picos, periodo, entre otras, ofrece diferentes ventajas, como bajo costo, fácil ejecución, además de ser una técnica no invasiva. El aspecto de mayor importancia de las mediciones acústicas para una voz sana y una enferma, es la interpretación en términos fisiológicos (de laringe) o funcionales (de la glotis).

Existen dos grupos de parámetros fundamentales para el análisis de voces con problemas patológicos: de ruido aditivo (turbulencia), parámetros que describen ruido en la modulación de la frecuencia (Jitter) y ruido en la modulación de la amplitud (Shimmer) [15], que corresponden a mediciones cuantificables de las señales. Dichas medidas corresponden a diferentes formas en las que el tracto vocal y en general el aparato fonador se ven afectados en mayor o menor medida, y por ende generan afectación en el proceso de generación de la voz.

De acuerdo con cifras de la Organización Mundial de la Salud, anualmente entre el 5 y el 15% de la población mundial es afectada con algún tipo de infección, en las vías respiratorias, debido al contagio de gripe o alguna otra infección produciendo amigdalitis o laringitis. Teniendo en cuenta que dichas enfermedades presentan en común algu-

nos síntomas que pueden generar afectación en la voz (congestión nasal, dolor de garganta, rinitis), entonces, principalmente se consideran estas afecciones para el diseño y también la implementación del algoritmo. Considerando lo reportado en [16] y [17] de forma general para el sistema de reconocimiento se consideran grabaciones de:

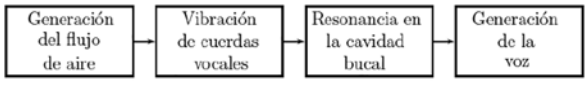
- Voz normal: Tomada como referencia para cada usuario.
- Voz afónica: Asociada a laringitis, infecciones virales, resfriado y alergia al polen.
- Voz nasal: Asociada a sinusitis, rinitis y alergias.

En las siguientes secciones se busca establecer las características más importantes que permiten manejar las afectaciones respiratorias que puede sufrir una persona.

5. CARACTERÍSTICAS BASADAS EN EL ANÁLISIS DE LA VOZ

En el proceso de generación de la voz (fonación) interviene el flujo de aire originado por los pulmones en su proceso de exhalación. Inicialmente el flujo de aire se estrella con las cuerdas vocales, las cuales generan una serie de vibraciones debido a los cambios en la presión del flujo del aire. De esta forma, el aire en movimiento pasa por la cavidad bucal, la cual, gracias a la acción de dientes y lengua, aporta diferentes niveles de resonancia para finalmente completar el proceso de fonación [18]. Este proceso es ilustrado en Fig.4.

Fig. 4. PROCESO PARA LA GENERACIÓN DE LA VOZ



Fuente: Los autores.

El análisis aerodinámico permite realizar la valoración objetiva de flujos de aire y presiones mediante la medición de diferentes parámetros, como son:

- Tiempo máximo de fonación.
- Índice fonorespiratorio.
- Capacidad vital.
- Cociente de fonación.

El análisis acústico brinda información sobre la calidad de la voz, considerando sus principales parámetros acústicos [18]. Dichos parámetros son:

- Frecuencia fundamental: es la onda de frecuencia más baja entre las que forman un sonido complejo.
- Intensidad: es la potencia acústica transferida por un sonido.
- Jitter: es la perturbación involuntaria de la frecuencia fundamental.
- Shimmer: es la perturbación de la amplitud de la frecuencia fundamental.
- Ruido glótico: es el ruido causado por la turbulencia del aire pasando por el aparato respiratorio.

Para el algoritmo propuesto se realiza el análisis acústico de la voz, empleando las mediciones de Jitter y Shimmer, ya que de acuerdo con [19], estas proveen información importante que varía dependiendo del tipo de voz. Jitter y Shimmer son mediciones de las variaciones de la frecuencia y amplitud respectivamente [19].

Jitter: Es la perturbación involuntaria de la frecuencia fundamental entre cada ciclo vocal y el siguiente. Es una medición que permite establecer el porcentaje de estabilidad de la fonación. En condiciones normales, la frecuencia entre ciclo y ciclo de fonación no es exactamente igual, pero dicha variación tiene un grado de tolerancia, una voz con problemas tendrá una variación en la frecuencia mucho más alta [18].

Shimmer: Es la perturbación en la amplitud de la frecuencia fundamental entre cada ciclo vocal. Se mide en porcentaje y permite determinar el grado de disfonía de una voz, aunque no es posible relacionar esta medición a una patología determinada [18].

Existen diferentes aproximaciones para calcular el Jitter y el Shimmer, según [19] una adecuada for-

ma para realizar estos cálculos es mediante una aproximación relativa para el Jitter y un cálculo en decibeles para el Shimmer. Las ecuaciones 2 y 3 corresponden al Jitter (relativo) y Shimmer (dB), respectivamente.

$$Jitter(\%) = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|T_i - T_{i+1}|}{\frac{1}{N}\sum_{i=1}^{N}|T_i|} \quad (2)$$

$$Shimmer = \frac{1}{N}\sum_{i=1}^{N-1}\left|20Log_{10}\left(\frac{A_{i+1}}{A_i}\right)\right| \quad (3)$$

Para el Jitter medido en porcentaje T_i es el periodo de la frecuencia fundamental para una ventana i y N es el número total de ventanas [20]. Por su parte, para el Shimmer A_i son los datos de amplitud de pico a pico extraídos y N es el número de periodos de frecuencia fundamental extraídos [21].

6. ANÁLISIS ACÚSTICO DE LA VOZ

Con el fin de implementar el algoritmo se lleva a cabo un análisis acústico de la señal de voz, tanto para voces normales como para las que sufren alguno de los trastornos seleccionados. El análisis acústico de la voz proporciona información sobre la calidad de la misma y se realiza mediante el estudio de los diferentes parámetros que la componen. La mayor importancia del análisis acústico radica en la relevancia que tiene en la cuantificación de la disfonía y su grado de evolución [18]. Para el algoritmo se emplea el cálculo del Jitter, Shimmer y la medición de la frecuencia promedio de la muestra de voz.

Para este análisis las grabaciones utilizadas corresponden a secuencias de números del 0 al 9, similar a las utilizadas en base de datos XM2VTS la cual cuenta con grabaciones de video, imágenes y audio para ser empleadas en el reconocimiento de personas [22]. La primera muestra corresponde a una voz normal, la segunda a una voz afónica y la tercera a una voz nasal.

El primer paso para analizar las grabaciones es remover los silencios, esto se hace utilizando una función que cuantifique la energía de la voz en cada segmento. Si la medición en el segmento es menor a un umbral de energía mínimo para la voz humana, este segmento es ignorado [13]. A las grabaciones de voz normal, afónica y nasal se les remueve el silencio.

El siguiente paso consiste en realizar el análisis acústico de características de la voz que han sido seleccionadas, la Tabla I contiene la información recopilada para las grabaciones analizadas.

TABLA I
CARACTERÍSTICAS DE GRABACIONES

Tipo de voz	Frecuencia fundamental (Hz)	Jitter (%)	Shimmer (dB)
Normal	137,6	8	18
Afónica	156,3	36,5	16,05
Nasal	142,3	10,3	20,3

Fuente: Los autores.

Al analizar los resultados obtenidos es posible identificar que la mayor variación se tiene en el Jitter de una voz afónica. Por otro lado, el Shimmer es la medición que presenta menor variación en cada tipo de voz, esta medida tiende a ser estable pese a la afectación en la voz, por tanto se utiliza como característica adicional para la toma de decisiones en el proceso de identificación.

7. PROCESO DE IDENTIFICACIÓN

El proceso de identificación se efectúa comparando patrones de registros de voz para identificar coincidencias. Durante este proceso las características empleadas son: la medición de distancias mediante el cálculo de coeficientes cepstrales [13], comparación de formantes en la señal de voz [23], cálculo de Shimmer y comparación de frecuencia fundamental [23]. Como parte del proceso de enrolamiento de usuarios, se calculan los valores de las características mencionadas previamente. De esta forma, al realizar el proceso de identificación, dichas características son medidas de nuevo y así es posible la comparación entre registros.

El proceso de identificación se ejecuta sobre un conjunto de muestras de voz. Adicionalmente, se

crean registros de control para ser identificados, midiendo de esta forma la efectividad en el proceso de identificación. Como parte del modelo, para los usuarios de control se toman tres registros en el sistema perteneciente a un mismo usuario (que pronunciará lo mismo varias veces), de esta forma se puede mejorar la calidad de los registros almacenados facilitando el proceso de identificación.

Para el proceso de identificación de coincidencias se consideran dos posibilidades, la primera solamente empleando la medición de las distancias de los coeficientes cepstrales, obteniendo los cinco registros con una menor distancia entre las características.

Para la segunda, adicional a la medición de distancias utilizada en la etapa anterior, se tiene en cuenta el cálculo de Shimmer, la forma de la señal y la frecuencia fundamental para obtener un puntaje total del registro. La combinación de las mediciones entrega como resultado el puntaje de la comparación. Un registro con menor puntaje significará una menor diferencia entre los registros, lo cual representa una coincidencia más aproximada.

Cada una de las mediciones mencionadas se realizará de la siguiente forma:

Distancias mediante coeficientes cepstrales: Para la comparación los coeficientes cepstrales se toman de un grupo de características que ha sido obtenido durante el proceso de enrolamiento. Estos se utilizan como referencia para calcular la respectiva distancia frente a uno de prueba [13]. La comparación que tenga como resultante una menor diferencia será considerada como coincidencia.

Shimmer: De acuerdo con las pruebas del análisis de la voz, se calcula el Shimmer para el registro durante el proceso de enrolamiento. Luego, en el proceso de identificación, se calculará el Shimmer para el registro que se va a identificar y se mide la diferencia absoluta.

Forma de la señal: Durante el proceso de enrolamiento se obtiene un vector de la posición de picos de la señal. Posteriormente, en el proceso de identificación, se calcula el vector para el registro que se va a identificar y se calcula la diferencia entre ellos [23].

Frecuencia fundamental: En el proceso de enrolamiento se obtiene el valor de frecuencia fundamental para la señal de voz ingresada. Luego, en el proceso de identificación, se calcula la frecuencia fundamental para el registro de entrada y se establece la respectiva diferencia [18].

8. RESULTADOS

Las pruebas se llevaron a cabo con registros de voz de terceros, los cuales, se cargan en el sistema como muestras de usuarios registrados. Fueron utilizados diferentes videos de libre distribución para obtener grabaciones extensas de audio, cada una de estas grabaciones fue cortada en fragmentos entre cuatro y seis segundos y almacenada en formato wav de un solo canal a 16 bits y 44100Hz.

Adicional a estas muestras, se obtuvieron grabaciones de usuarios de control sobre los cuales se realizaron mediciones para establecer el porcentaje de identificación del algoritmo. Estas tienen las mismas características de grabación (1 canal de audio, 16 bits y 44100Hz). Las grabaciones se obtienen mediante la utilidad record-voices, permitiendo grabar diferentes muestras de la misma persona pronunciando la misma frase “transmilenio biarticulado rojo”.

Para la medición por puntaje se realiza una prueba preliminar, donde todos los registros obtenidos en las pruebas fueron analizados para identificar los cinco registros que tuvieron un menor puntaje.

Los porcentajes utilizados para la primera prueba surgen del análisis de cada una de las características y su valor dentro del proceso de identificación, la asignación se realiza mediante estimación de significancia. En la Tabla II se encuentra el registro de porcentajes propuestos que fueron asignados a cada configuración.

TABLA II
PORCENTAJES ASIGNADOS PARA CADA CARACTERÍSTICA

Característica	Porcentajes asignados para cada prueba		
	Prueba1	Prueba 2	Prueba 3
Distancia de coeficientes cepstrales	70%	70%	60%
Forma de la señal	10%	10%	15%
Shimmer	15%	10%	15%
Frecuencia fundamental	5%	10%	10%

Fuente: Los autores.

Para esta medición, primero se obtuvo la diferencia de puntaje entre el tercer y cuarto lugar para cada uno de los procesos de identificación realizados; posteriormente se determina el valor promedio. La Tabla III contiene la información referente a la diferencia encontrada.

TABLA III
DIFERENCIA PROMEDIO DE PUNTAJES

Prueba número	Prueba 1	Prueba 2	Prueba 3
Diferencia obtenida	1,5721	1,5502	0,7442

Fuente: Los autores.

Lo anterior permite establecer que la mejor asignación de porcentajes corresponde a los utilizados en la primera prueba: 70% para la distancia, 10% para la forma de la señal, 15% para el Shimmer y 5% para la frecuencia fundamental. Adicionalmente se puede establecer un umbral de puntaje bajo, en el cual se identifican coincidencias: una medición que sea menor o igual a la diferencia obtenida en la prueba 1, se puede tomar como una coincidencia.

Con el fin de mostrar el funcionamiento del sistema se lleva a cabo una prueba piloto con 11 usuarios diferentes, cada uno con 4 muestras de voz (considerando lo reportado en [8] donde se realiza una prueba con 12 usuarios). Para los procesos de medición y análisis de resultados, tanto por distancia como por puntaje, solo se tienen en cuenta los 5 primeros registros con mejor resultado.

La métrica para observar el desempeño del algoritmo consiste en el porcentaje de aciertos, el cual se puede calcular como el número de aciertos sobre la cantidad total de registros tal como se muestra en la ecuación 4.

$$J = \frac{N_{Aciertos}}{N_{Total}} 100\% \qquad (4)$$

Al realizar el respectivo proceso de identificación se obtuvieron los siguientes porcentajes de identificación según el tipo de implementación empleada:

- Implementación con coeficientes cepstrales: 77.78%
- Implementación con ponderación de características: 96.67%

Como es de apreciar, se tiene un mejor desempeño con la medición por puntaje (ponderación) al incorporar más información de la voz del usuario que se quiere identificar.

9. CONCLUSIONES

En este trabajo se establecen las características para ser empleadas en el algoritmo y se propone una posible ponderación de estas, en trabajos futuros se puede emplear optimización para encontrar la mejor ponderación.

El modelo propuesto de identificación integral para la medición de una serie de características de la voz, las cuales pueden variar debido a afectaciones fisiológicas, lo cual permite generar tolerancia a dichas variaciones. Según las pruebas realizadas, al incorporar las mediciones de la voz consideradas se logra una mejor identificación en relación con el caso donde solo se emplean coeficientes cepstrales.

Una limitante en este trabajo fue la baja cantidad de pacientes para realizar las pruebas, por lo que, en trabajos posteriores, se puede incluir un número mayor de personas.

Al analizar diferentes rasgos de la voz durante el proceso de extracción de características, fue posible identificar que la medición de Shimmer tiende a ser estable frente a diferentes afectaciones de la voz, lo cual permitió mejorar la identificación de usuarios con afectaciones de la voz.

REFERENCIAS

[1] S. Vasuhi, V. Vaidehi, B. N. Nare, T. Treesa, ''An efficient multi-modal biometric person authentication system using Fuzzy Logic'', in *IEEE Second International Conference in Advanced Computing (ICoAC)*, December, 2010.

[2] H. Hollien, ''Barriers to Progress in Speaker Identification'', *Linguistic Evidence in Security, Law and Intelligence*, vol. 1, no. 1, pp. 1-23, 2013.

[3] T. Kinnunen, H. Li, ''An Overview of Text Independent Speaker Recognition: from Features to Supervectors'', *Speech Communication*, vol. 52, no. 1, pp. 12-40, 2010.

[4] R. Peacocke, ''An Introduction to Speech and Speaker Recognition'', *Computer*, vol. 26, no. 8, pp. 26-33, 1990.

[5] R. Price, J. Willmore, W. Roberts, K. Zyga, ''Genetically optimized feed forward neural networks for speaker identification'', in *Fourth International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologie*s, Sept. 2000.

[6] T. Gannert, ''A Speaker Verification System Under the Scope: Alize'', Master's Thesis in Speech Technology at TMH, 2007.

[7] J. P. Teixeira, A. Gonçalves, ''Algorithm for jitter and Shimmer measurement in pathologic voices'', *Procedia Computer Science*, vol. 100, pp. 271-279, 2016.

[8] E. San Segundo, A. Tsanas, P. Gómez, ''Euclidean Distances as measures of speaker similarity including identical twin pairs: A forensic investigation using source and filter voice characteristics'', *Forensic Science International*, vol. 270, pp. 25-38, 2017.

[9] H. Beigi, ''Speaker Recognition'', *InTech*, 2012.

[10] A. Boles, P. Rad, ''Voice biometrics: Deep learning-based voiceprint authentication system'', in *12th System of Systems Engineering Conference (SoSE)*, pp. 1-6, June, 2017.

[11] E. Ramdinmawii, V. K. Mittal, ''Gender identification from speech signal by examining the speech production characteristics'', in *International Conference on Signal Processing and Communication (ICSC)*, pp. 244-249, December, 2016.

[12] H. Pérez, J. Martínez, I. Espinosa, J. Rodríguez, H. Ávila, ''Using acoustic paralinguistic information to assess the interaction quality in speech-based systems for elderly users'', *International Journal of Human-Computer Studies*, vol. 98, pp. 1-13, 2017.

[13] H. Beigi, ''Fundamentals of Speaker Recognition'', *Springer US*, 2011.

[14] M. Fröhlich, H. Michaelis, H. Werner, ''Acoustic Breathiness Measures in the description of patologic voices'', in *IEEE International Conference on Acoustics, Speech and Signal Processing*, May, 1998.

[15] D. Michaelis, T. Gramss, H. W. Strube, ''Glottal-to-Noise Excitation Ratio - a New Measure for Describing Pathological Voices'', *ACUSTICA - acta acústic*a, vol. 83, pp. 700-706, 1997.

[16] A. Mendoza, G. Mansilla, ''Rinitis alérgica'', *Revista de la Sociedad Boliviana de Pediatría*, vol. 41, no.1, pp. 50-53, 2002.

[17] K. Aubry, A. El Sanharawi, A. Pommier, ''Laringitis agudas del adulto'', *EMC – Otorrinolaringología*, vol. 46, no. 1, pp. 1-9, 2017.

[18] J. C. Casado, A. Pérez, ''*Trastornos de la voz: del diagnóstico al tratamiento*'', Ediciones Aljibe, 2009.

[19] M. Farrús, J. Hernando, P. Ejarque, ''Jitter and Shimmer Measurements for Speaker Recognition'', in *8th Annual Conference of the International Speech Communication Association ISCA*, pp. 778-81, Belgium, Aug. 2007.

[20] M. Shahbakhi, D. Taheri, E. Tahami, ''Speech Analysis for Diagnosis of Parkinson's Disease Using Genetic Algorithm and Support Vector Machine'', *Journal of Biomedical Science and Engineering,* vol. 7, no. 4, 2014.

[21] A. Zewoudie, J. Luque, F. Hernando, ''Jitter and Shimmer Measurements for Speaker Diarization'', in VIII Jornadas en Tecnología del Habla and IV Iberian SLTech Workshop, 2014, pp. 21-30.

[22] K. Messer, J. Matas, J. Kittler, K. Jonsson, ''The extended M2VTS Database'', in *Second International Conference on Audio- and Video-based Biometric Person Authentication AVBPA*, 1999, pp. 72-77.

[23] E. D. Ellis, ''Design of a Speaker Recognition Code using MATLAB'', Department of Computer and Electrical Engineering - University of Thennessee, Knoxville Tennessee, May, 2001.