Revista Ingenierías Universidad de Medellín
ISSN: 1692-3324
Universidad de Medellín

Sepúlveda-Cano, Lina María; Quiza-Montealegre***, Jhon Jair; Gómez, Jorge Andrés

Análisis de la influencia de las técnicas de compresión de voz en la detección de anomalías vocales *

Revista Ingenierías Universidad de Medellín, vol. 16, núm. 30, 2017, Julio-Diciembre, pp. 49-66
Universidad de Medellín

DOI: https://doi.org/10.22395/rium.v16n30a3

Disponible en: https://www.redalyc.org/articulo.oa?id=75054207004

Análisis de la influencia de las técnicas de compresión de voz en la detección de anomalías vocales*

*Lina María Sepúlveda Cano***
*Jhon Jair Quiza Montealegre****
*Jorge Andrés Gómez*****

Recibido: 29/07/2015 • Aceptado: 15/09/2016
DOI: 10.22395/rium.v16n30a3

Resumen

En este artículo se comparan los resultados de utilizar señales de voz comprimidas frente a señales de voz sin comprimir para detectar de forma automática anomalías vocales. Las técnicas de codificación y compresión de voz usadas en este estudio son las mismas que se utilizan de forma estándar en los sistemas de telefonía fija, móvil e IP, y las técnicas de caracterización y clasificación usadas también están dentro de las más utilizadas para la detección automática de anomalías de voz. Los resultados obtenidos permiten concluir que es posible utilizar señales de voz comprimidas para detección automática de patologías vocales sin detrimento en el porcentaje de acierto en el diagnóstico, lo que haría posible la implementación de sistemas de telediagnóstico automático de patologías vocales.

Palabras clave: telediagnóstico, detección de patologías de voz, compresión de voz, análisis de bioseñales.

* El artículo se desarrolló en el marco de la investigación titulada "Validación de algoritmos de aprendizaje automático usados en sistemas de telecomunicaciones de radio cognitiva", financiada por la Universidad de Medellín, durante el periodo 2013-2, 2015-2.

** Ingeniera electrónica. PhD en Ingeniería - Automática. Investigadora Grupo Arkadius. Universidad de Medellín. Carrera 87 30-65, Medellín. Colombia. Dirección electrónica: lmsepulveda@udem.edu.co.

*** Ingeniero electrónico. MSc en Ingeniería - Telecomunicaciones. Investigador Grupo Arkadius. Universidad de Medellín. Carrera 87 30-65, Medellín. Colombia. Dirección electrónica: jhquiza@udem.edu.co.

**** Ingeniero electrónico. MSc en Ingeniería - Automatización. Investigador Grupo Bioingeniería y Optoelectrónica (ByO). Universidad Politécnica de Madrid. Carretera de Valencia Km 7. Madrid. España.Dirección electrónica: jorge.gomez.garcia@upm.es.

Analysis of the influence of signal compression techniques for voice disorder detection through filter-banked based features

Abstract:

This paper compares the results of using compressed voice signals versus uncompressed speech signals to automatically detect voice abnormalities. Coding techniques and voice compression used in this study are the same as those used by default in the fixed, mobile and ip telephony systems, and techniques of characterization and classification used are also among the most used for detecting automatic speech abnormalities. The results obtained indicate that it is possible to use compressed voice signals for automatic detection of vocal pathologies without compromising the success rate in the diagnosis, which would make the implementation of automatic remote diagnosis of vocal pathologies possible.

Key words: remote diagnostics, voice pathology detection, voice compression, biosignals analysis.

INTRODUCCIÓN

Los avances en telecomunicaciones y ciencias de la computación han posibilitado el desarrollo de herramientas que apoyan y potencian la labor de los profesionales del área de la salud, igual que ocurre con casi todas las demás profesiones. En el caso particular de los servicios de salud, desde hace algún tiempo es posible diagnosticar pacientes de forma remota, gracias al procesamiento y transmisión de señales biológicas y multimedia [1]. En general, el diseño de estos sistemas se enfrenta al desafío de transmitir la información que es relevante para el profesional que está haciendo el diagnóstico, considerando las restricciones propias de los sistemas de comunicación (ancho de banda, fiabilidad, seguridad, retardos, entre otras).

El diagnóstico basado en señales de voz se presenta como un caso de estudio interesante, dado que, por su naturaleza las señales de voz requieren de menor ancho de banda que otras señales (imágenes o vídeo) [2], y que pueden transmitirse haciendo uso de las redes de telefonía fija, telefonía móvil y VoIP existentes, lo que facilitaría la implementación de estos sistemas de tele-diagnóstico.

Se hacen dos consideraciones acerca del diseño de estos sistemas de tele-diagnóstico basado en señales de voz: la forma en que es transmitida a través de la red de telecomunicaciones, y la forma en que es extraída la información relevante de la señal recibida. La primera tiene que ver con la técnica que es utilizada para digitalizar y comprimir la señal de voz, y la segunda, con las técnicas de análisis de datos y detección de patrones, utilizada ya en el destino. En este artículo se estudia el efecto que tienen las técnicas de compresión empleados por los sistemas de comunicaciones más utilizados para transmitir señales de voz como son las de telefonía fija, las redes de telefonía móvil y los sistemas de VoIP. En todos los casos se utilizó análisis de componentes principales (PCA) como técnica para clasificar e identificar señales que corresponden a anomalías de voz.

1. MARCO CONCEPTUAL

1.1 Técnicas de digitalización y compresión de voz en sistemas telefónicos

En el año 1937, el ingeniero inglés Alec Reeves propuso la primera técnica para digitalizar señales, conocida como modulación por pulsos codificados (PCM, por sus iniciales en inglés), compuesta por 3 etapas: muestreo, cuantización y codificación; esta técnica ha sido la base del proceso de digitalización de los sistemas de telefonía fija, a partir de los años 50 [3]. La técnica original, en la cual los intervalos de cuantización son uniformes, tiene el defecto de que para muestras pequeñas el ruido de cuantización es muy elevado y, por tanto, la señal original se recupera, pero distorsionada; una solución ineficiente es aumentar el número de intervalos de cuantización, lo que

implica un incremento en la velocidad de transmisión y, por ende, en el ancho de banda. Alternativas de solución más eficientes son el uso de técnicas de PCM no uniformes logarítmicas, en las cuales los intervalos de cuantización para muestras pequeñas son más pequeños que los intervalos para muestras grandes [4], de acuerdo (aproximadamente) con un conjunto de ecuaciones estandarizadas por la UIT en su recomendación G.711 [5], conocidas como Ley A y Ley *u*. Con estas técnicas, es posible alcanzar una muy alta calidad en la señal de voz recibida, lo que se suele medir a partir de pruebas subjetivas de percepción hechas a usuarios, conocidas como pruebas MOS [6].

A partir de esta técnica base, empezaron a desarrollarse otras técnicas de digitalización de señales de voz que permitieran su transmisión a velocidades menores, tratando de que el decremento en la calidad de la señal recibida fuera el menor posible. Entre estas se encuentran las técnicas PCM diferenciales como DPCM, ADPCM y la Modulación Delta [7], y las técnicas de análisis / síntesis o paramétricas, como los codificadores de voz (vocoders) y los codificadores predictivos lineales (LPC) [4]. Como se muestra en la tabla 1, con estas técnicas es posible alcanzar muy bajas velocidades de transmisión con una calidad aceptable, por lo que son utilizadas en redes de telefonía con restricciones en ancho de banda [8].

Tabla 1. Comparación de velocidad de transmisión vs. MOS de diferentes técnicas de codificación de voz [8].

Técnica de digitalización	*Velocidad de transmisión (Kbps)*	*Puntaje MOS*
G.711 PCM	64	4.1
G.726 ADPCM	32	3.85
G.728 Low Delay Code Excited Linear Predictive (LD-CELP)	15	3.61
G.729 Conjugate Structure Algebraic Code Excited Linear Predictive (CS-ACELP)	8	3.92
G.729a CS-ACELP	8	3.7
G.723.1 MP-MLQ	6.3	3.9
G.723.1 ACELP	5.3	3.65
iLBC Freeware	15.2	3.9

Fuente: elaboración propia

En este trabajo, se pretende determinar si estas mismas técnicas de digitalización que comprimen la señal de voz producen o no pérdida de información relevante para la detección de anomalías de voz, de manera similar a como degradan la calidad de la voz percibida.

1.2 Análisis de componentes principales PCA

El análisis de componentes principales es una técnica de extracción de características, que busca encontrar un espacio de menor dimensión que represente adecuadamente los datos de entrada. Según [9], es posible demostrar que el espacio de menor dimensión que mejor representa los datos originales viene definido por los vectores propios asociados a los q mayores valores propios de la matriz de covarianza de la matriz de datos. Se parte, entonces de la matriz de observaciones o de datos X con dimensiones $n \times p$, en donde n es la cantidad de elementos u observaciones, y p ($p>q$) es el número de variables o características. La matriz de covarianzas de X se puede calcular mediante $S=1/n\ X^TX$, donde X debe tener media cero. El cálculo de los valores propios de S estará determinado por la ecuación (1):

$$\left|S-\lambda I\right|=0, \tag{1}$$

y sus vectores asociados se dan por la ecuación (2):

$$\left(S-\lambda_i I\right)v_i=0, \tag{2}$$

donde i = 1,…,p, vi son los vectores propios, λ_i son los valores propios, e I es la matriz identidad. Finalmente, la matriz proyectada estará dada por la ecuación (3):

$$Z=XV, \tag{3}$$

donde V es la matriz conformada por los vectores propios asociados a los valores propios ordenados de forma descendente, y $V^TV=I$.

Detección automática de patologías de voz

Como alternativa a los métodos tradicionales de detección de patologías de voz utilizadas por los médicos–audición de la voz del paciente por parte del especialista e inspección visual de las cuerdas vocales por medio de técnicas de laringología [10]– desde hace alrededor de dos décadas se empezaron a desarrollar técnicas automáticas aprovechando las tecnologías de información y comunicaciones. En un trabajo pionero en este campo [11], haciendo uso de técnicas de procesamiento digital, se estima la capacidad de discriminación de señales de voz entre normales y patológicas, de acuerdo con el análisis multidimensional de características extraídas. En [12] se presenta el desarrollo de un sistema no invasivo para el análisis de patologías de voz, que hace análisis de las señales de voz en tiempo, frecuencia y cepstrales, y combina cuatro técnicas de clasificación: PDM (*Prototype Distribution Maps*), K-NN (*K-Nearest Neighbourgh*), LDA (*Linear Discriminant Analysis*) y SOM (*Self-Organizing Map*).

En [10] se propone el uso de parámetros no clásicos, como la interferencia del índice de biocoherencia (IB), el ruido glotal para la identificación de patologías en señales de voz, junto con el uso de redes neuronales como clasificadores para la identificación de patologías vocales. En [13] se presenta un sistema no invasivo con una tasa de acierto del 99.44%, usando bancos de filtros de coeficientes cepstrales en escala de Mel (MFCC) y un clasificador HMM (*Hidden Markov Model*); la base de datos utilizada fue la Massachusetts Eye and Ear Infirmary (MEII). En [14] se propone un sistema que hace uso de la transformada de paquetes de wavelets (WPT) para parametrización, y de LDA y PCA (*Principal Component Analysis*) como métodos de clasificación, que optimiza el proceso de detección automática de patologías en señales de voz; los autores reportan una precisión del 100%.

Como se observa, se pueden encontrar muchos trabajos sobre sistemas automáticos de detección de patologías vocales, pero en estos la señal de voz ha sido codificada con alta calidad, generalmente muestreada a 44100 Hz y codificada a 16 bits, condiciones que son factibles para diagnósticos in-situ, pero no para telediagnósticos. En este último caso los autores solo han encontrado los siguientes trabajos: [15] propone un sistema para detectar de forma remota patologías de voz usando señales de voz con la calidad de un sistema de telefonía fija; los resultados de este trabajo muestran que mientras para señales de voz de alta calidad grabadas en ambientes controlados, la precisión en la detección de patologías de voz alcanza el 89,1%, la precisión alcanzada para señales de voz con calidad telefónica desciende al 74,2%. En [16] se reporta una precisión del 89,7% usando la metodología EMD (*Empirical Mode Decomposition*) para clasificación, asumiendo una relación señal–ruido del canal de 30 dB. En [17] se analizan los efectos de la compresión de señales de voz en formato mp3 en la detección de patologías, usando como métodos de clasificación GMM (*Gaussian Mixture Models*) y SVM (*Support Vector Machines*); se obtiene como resultado precisiones del 87,05% usando SVM y 85,37% usando GMM, para señales comprimidas a 8 kpbs. En [18] se propone un método para mejorar la precisión de un sistema que transmite las señales de voz haciendo uso de redes inalámbricas. Finalmente, en [19] se evalúa un sistema que utiliza los micrófonos de los teléfonos inteligentes como transductor de entrada, obteniendo una precisión del 84,6%

2. DISEÑO EXPERIMENTAL

La metodología propuesta tiene las siguientes fases: *i)* Preprocesamiento, *ii)* Codificación - Compresión, *iii)* Caracterización, y *iv)* Clasificación.

En este trabajo se utilizó la base de datos de desórdenes en la voz, grabada en el Massachusetts Eye and Ear Infirmary [20]. Tanto los registros patológicos (175

observaciones), como los de la clase normal (53 observaciones) tienen una frecuencia de muestreo de 44100Hz, y fueron remuestreados a 25000Hz con una cuantización de 16 bits. Los registros con duración mucho menor a 1 segundo fueron eliminados (5 observaciones patológicas). La longitud de los registros es de 21.700 muestras (0.868 segundos).

2.1 Preprocesamiento

Los registros se normalizaron bajo la expresión de la ecuación (4) y se multiplicaron posteriormente por una ventana tipo Tukey con el fin de suavizar los extremos de la muestra.

$$x(t)=\frac{2\left(x(t)-E\{x\}\right)}{\max\limits_{\forall t}\{x\}-\min\limits_{\forall t}\{x\}}, t\in T \tag{4}$$

2.2 Codificación-compresión

Se probaron cuatro métodos de codificación de voz: Codificación no lineal ley A, codificación no lineal ley *u*, LPC y DCT [2], a los cuales se les hicieron las variaciones de parámetros presentadas en la tabla 2.

Tabla 2. Variación de parámetros para las técnicas de compresión de voz usadas en el experimento

Técnica	*Parámetros*	*Valores*
LPC	Longitud de ventana *m*	750 bins
	Orden de filtro *p*	10, 20, 30, 40
DCT	Longitud de ventana *m*	12500, 6250, 3125
	Factor de compresión *p*	2, 4, 8
PCM	A	87.6
	u	255

Fuente: elaboración propia

A continuación se explica en qué consisten estos métodos.

2.2.1 Codificación no lineal Ley A y Ley u

Para la codificación de audio, la IUT bajo la recomendación G.711 establece dos estándares: la ley μ, usada en Estados Unidos y Japón, y la ley A, usada en Europa y el resto del mundo. Ambos mecanismos son no–lineales y las escalas de cuantización son

similares; la única diferencia es la ecuación matemática que se utiliza para producir la curva de cuantización [21]. En la ley A, el codificador convierte muestras de 13 bits codificadas en PCM lineal, a muestras comprimidas de 8 bits en forma logarítmica; mientras en la ley μ se convierten muestras de 14 bits [22]. En las ecuaciones (5) y (6) se describe matemáticamente el efecto de compresión para la Ley A y la ley μ, respectivamente.

$$\hat{x} = \begin{cases} \dfrac{A|x|}{1+ln\,A} & 0 \le |x| \le \dfrac{1}{A} \\ \dfrac{sgn\,(x)\left(1+ln\,(A|x|)\right)}{1+ln\,(A)} & \dfrac{1}{A} \le |x| \le 1 \end{cases} \tag{5}$$

$$\hat{x} = \frac{\text{sgn}\,(x)\ln\,(1+\mu|x|)}{\ln\,(1+\mu)}, \tag{6}$$

donde x está definido entre 0 y 1.

2.2.2 Compresión por codificación predictiva lineal (LPC)

El principio detrás de la codificación predictiva lineal (compresión por síntesis) es la minimización de la suma de las diferencias al cuadrado entre la señal original y la señal estimada durante una duración finita [23]. La descomposición del fenómeno físico que determina la producción de la voz en instantes breves donde sus características permanezcan estáticas determina la duración de la ventana de análisis. De esta manera, LPC determina los coeficientes de un filtro digital de longitud finita (FIR) (ecuación (7)) que pueda predecir el siguiente valor a partir de las entradas pasadas y presentes [24].

$$H(z) = \frac{G}{A(z)}, \tag{7}$$

donde G es la ganancia y $A(z) = 1 - \sum_{k=1}^{p} a_k z^{-} k.$ Si a la entrada se tiene un espectro de excitación U(z), el espectro del habla estaría determinado por la ecuación (8).

$$Y(z) = H(z)U(z) = \frac{G}{A(z)}U(z) \tag{8}$$

De la ecuación (8) se puede determinar que la función del habla en el tiempo discreto estará determinada por la ecuación en diferencias (ecuación (9)).

$$y(n) = \sum_{k=1}^{p} a_k y(n-k) + Gu(n) \tag{9}$$

Para determinar el error entre la señal original y la señal reconstruida se hace uso del error cuadrático medio, empleando la ecuación (11):

$$E = \sum_{m=-\infty}^{\infty} \left(y(m) - \sum_{k=1}^{p} a_k y(n-k) \right)^2 \tag{10}$$

Con el fin de determinar los coeficientes del filtro a_k se pueden aplicar dos métodos: en el primer caso se recurre a las ecuaciones de Yule-Walker para solucionar las ecuaciones normales de autocorrelación (ecuación (11)), que resultan al minimizar el error cuadrático medio de la ecuación (8).

$$R_n(i) = \sum_{k=1}^{p} (a_k R_n(|i-k|)), \tag{11}$$

donde,

$$R_n(i) = \sum_{m=-\infty}^{\infty} y_n(m) y_n(|m-i|) \tag{12}$$

En el segundo caso se recurre al algoritmo recursivo de Levinson-Durbin (algoritmo 1) en donde se resuelven las ecuaciones normales para la autocorrelación que determinan el error de predicción mínima E_p para un modelo de p polos, como se muestra en la ecuación (13).

$$E_p = R(0) - \sum_{k=1}^{p} a_k R(k) \tag{13}$$

Entrada: R,m

Salida: a_k

Inicialización de variables

$$E_0 = R(0), a_0 = 0$$

Para cada m haga:

$$a_m^m = k_m$$

$$a_i^m = a_i^{m-1} - k_m a_{m-i}^{m-1}$$

$$E_m = \left(1 - {}_m^2\right) E_{m-1}$$

Fin

Algoritmo 1. Algoritmo de Levinson - Durbin

2.2.3 Compresión por transformada discreta del coseno (DCT)

La transformada discreta del coseno es un método usado ampliamente en aplicaciones de compresión (compresión por transformación), debido a que la energía de la señal de voz se concentra principalmente en pocos coeficientes de la transformada dando como resultado un factor alto de compresión [2]. Los coeficientes de la transformada están determinados por la ecuación (14), siendo para este caso $s_m = x(n)$ la señal de entrada.

$$X(k) = \sum_{n=0}^{N-1} x(n) \cos\left[\frac{\pi}{n} j\left(k + \frac{1}{2}\right)\right] \tag{14}$$

2.3 Caracterización

Las señales de voz fueron caracterizadas usando técnicas de representación tiempo-frecuencia y coeficientes cepstrales. A continuación se describen estas técnicas.

2.3.1 Representaciones de tiempo-frecuencia

Las distribuciones de tiempo-frecuencia son basadas en el principio de incertidumbre, donde la frecuencia de una señal en un instante particular de tiempo no puede ser determinada. En señales que presentan alta no-estacionariedad, por lo general se hace necesaria la información de ambos parámetros al mismo tiempo. La transformada corta de Fourier (STFT) introduce el concepto de localización utilizando una ventana deslizante ϕ. El espectrograma es una transformación tiempo-frecuencia utilizada regularmente en diferentes aplicaciones, y se calcula como la magnitud al cuadrado de la STFT a través de la ecuación (15), así:

$$|x,\phi|^2 = \left| \int_T x(\tau)\phi(\tau - t) e^{-j2\pi f \tau} d\tau \right|^2 \triangleq S_x(t,f), \tag{15}$$

El espectrograma generado para cada una de las observaciones en los diferentes esquemas de compresión tiene como parámetros: 1.024 puntos en frecuencia y una ventana de Hamming de 513 puntos. Las figuras 1 a 8 muestran ejemplo de espectrogramas de cada una de los esquemas de compresión.

2.3.2 Coeficientes cepstrales

Con el fin de sintetizar la información contenida en las representaciones tiempo-frecuencia, algunos parámetros se han aceptado ampliamente para la caracterización de señales biológicas [25]. Estos parámetros se pueden calcular a partir de una descomposición

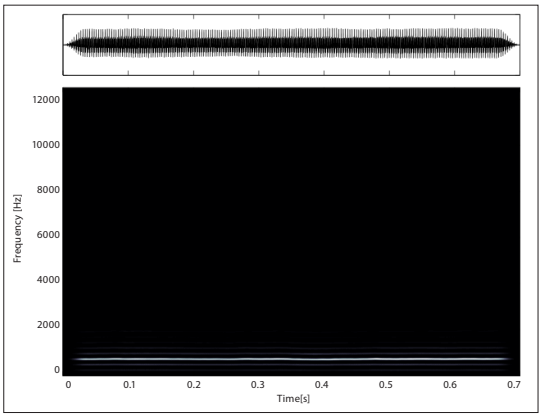

Figura 1. Espectrograma de señal de voz normal sin comprimir.
Fuente: elaboración propia

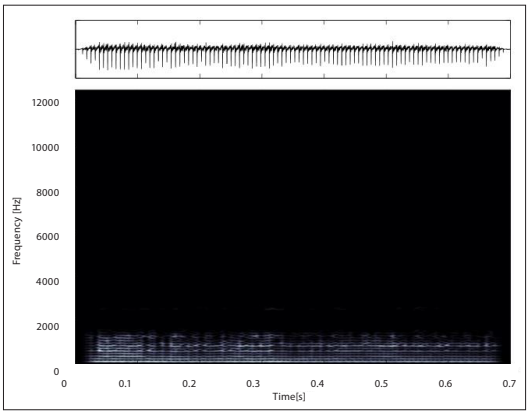

Figura 2. Espectrograma de señal de voz con anomalías sin comprimir.
Fuente: elaboración propia

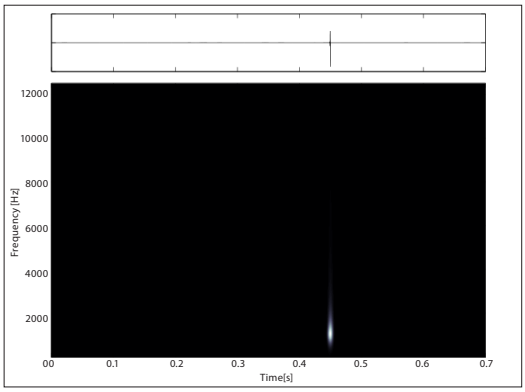

Figura 3. Espectrograma de señal de voz normal comprimida en LPC.
Fuente: elaboración propia

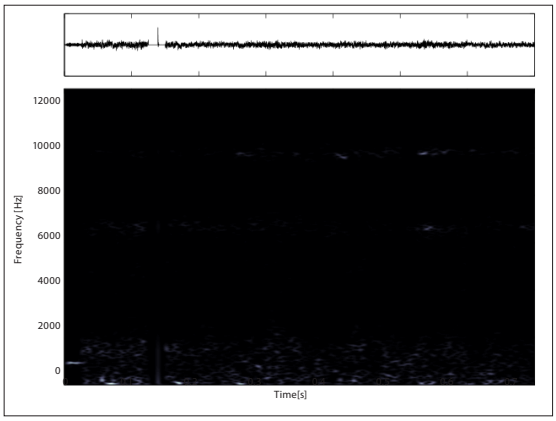

Figura 4. Espectrograma de señal de voz con anomalías comprimida en LPC.
Fuente: elaboración propia

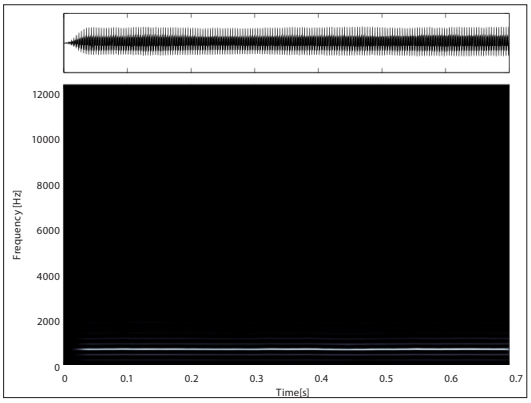

Figura 5. Espectrograma de señal de voz normal comprimida en DCT.
Fuente: elaboración propia

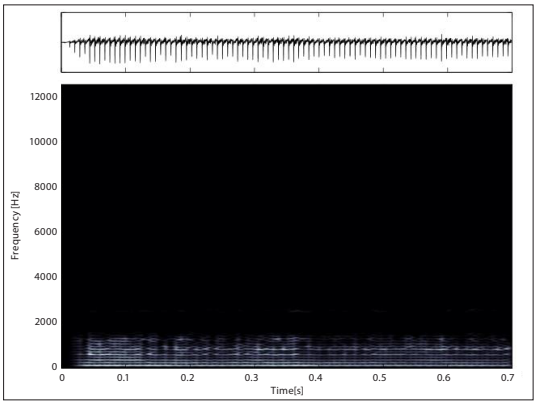

Figura 6. Espectrograma de señal de voz con anomalías comprimida en DCT.
Fuente: elaboración propia

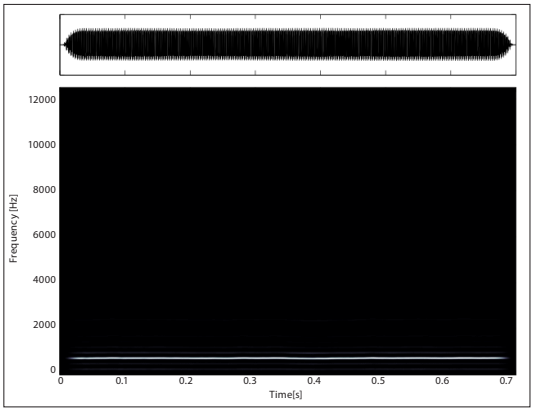

Figura 7. Espectrograma de señal de voz normal codificada en PCM Ley A.
Fuente: elaboración propia

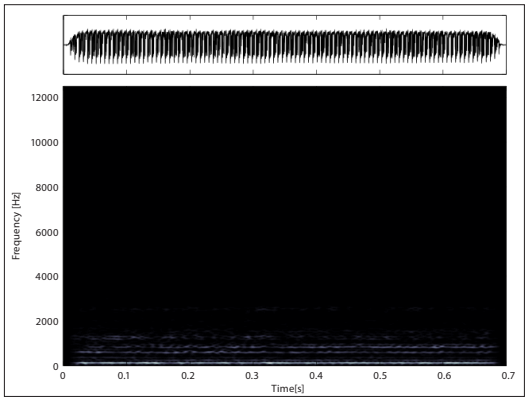

Figura 8. Espectrograma de señal de voz con anomalías codificada en PCM Ley A.
Fuente: elaboración propia

por bancos de filtros, en donde la señal resultante del proceso de filtrado combina la información tanto de la magnitud del espectro como de la frecuencia. Para el caso de las señales de voz, se utiliza una distribución de filtros en escala Mel, i.e., una distribución logarítmica. El conjunto de características $\{y_i\}$ extraídas de esta forma se denomina coeficientes cepstrales en escala Mel (MFCC) y se calculan mediante la transformada discreta del coseno de un banco de filtros triangulares, $\left\{F_m(f) : m = 1,\ldots,n_M\right\}$, espaciados en escala logarítmica en el dominio de la frecuencia como se muestra en la ecuación (16).

$$y_i = \sum_{m=1}^{n_M} \log\left(s_m\right) \cos\left(i\left(m - \frac{1}{2}\frac{\pi}{p}\right)\right) \tag{16}$$

donde p es el número deseado de características y $s_m \in \mathbb{R}^T$ es la sumatoria ponderada de la respuesta de cada conjunto de filtros, $s_m = \sum_{\lambda=1}^{F} S(f) H_m(f)$, siendo m y f los índices de los ejes ordinales y de frecuencia, respectivamente; F es el número de muestras en el dominio de la frecuencia, y H_m el conjunto de filtros pasa–banda distribuidos en escala Mel [26].

Para los coeficientes cepstrales en escala Mel, se tomaron 11 coeficientes de un banco de 32 filtros como se recomienda en [26].

2.4 Clasificación

Para la validación de la metodología propuesta se utiliza la técnica de validación cruzada de 10 particiones (10-folds), mediante la cual se suelen evaluar modelos predictivos. La idea básica es hacer una partición de las muestras totales de las que se dispone, de tal manera que se pueda obtener un conjunto de entrenamiento y un conjunto de validación para el clasificador. La partición se debe hacer de manera aleatoria durante varias iteraciones (en este caso 10 iteraciones), manteniendo el mismo porcentaje de muestras en el conjunto de entrenamiento y el mismo porcentaje de muestras en el conjunto de validación. Los porcentajes de muestras comúnmente utilizados son 80% o 70%, para el conjunto de entrenamiento, y 20% o 30% para el conjunto de validación. Para la metodología propuesta se recurrió a la partición 70%-30%. La finalidad de una validación de este tipo es intentar evitar que el clasificador se sobre-entrene, y adicionalmente, obtener información sobre la distribución de los datos a lo largo de las muestras a través de la desviación estándar del acierto de clasificación de las 10 iteraciones. El clasificador utilizado es un *K-NN* o de *k* vecinos más cercanos. La principal razón para la elección de este clasificador es que la simplicidad de su funcionamiento permite establecer que el mayor esfuerzo en la metodología propuesta no se cargue sobre esta etapa, sino sobre las etapas anteriores que se quieren comprobar.

3. RESULTADOS Y DISCUSIÓN

En la tabla 3 se muestran los resultados obtenidos en términos de acierto de clasificación para los diferentes esquemas de compresión. Como se puede apreciar, la técnica que presenta menor desempeño es la compresión por LPC, exceptuando el caso del filtro de grado 2. Esto se puede explicar a partir de las gráficas 1(a) a 1(h). Las señales de audio quedan caracterizadas por un tono en un momento determinado, al ser un tipo de compresión por síntesis, y gran parte de la información espectral, principalmente en lo que se refiere a potencia, se pierde a lo largo del tiempo, haciendo difícil identificar un fenómeno físico normal de uno patológico. Respecto a la compresión por

DCT es importante resaltar que bajo algunos parámetros el acierto de clasificación es mayor que en el caso de la señal sin compresión, de lo cual se puede concluir que en el proceso de compresión se elimina información innecesaria de la señal de voz. Este comportamiento se puede apreciar igualmente en las técnicas PCM no lineales que alcanzan un mayor porcentaje de acierto. En ambas técnicas (DCT y PCM no lineal) se llevan a cabo operaciones logarítmicas que acentúan las bandas bajas de frecuencia, en el primer caso, y las señales de baja intensidad, en el segundo, para así hacer una mejor definición a los fenómenos vocales.

Tabla 3. Precisión de la clasificación de patologías de voz

Técnica	*Parámetros*	*Precisión (%)*
Sin comprimir		85,07 ± 2,98
LPC	$m = 750, p = 10$	84,17 ± 3,31
	$m = 750, p = 20$	75,52 ± 1,04
	$m = 750, p = 30$	75,82 ± 3,70
	$m = 750, p = 40$	77,74 ± 2,38
DCT	$m = 12500, p = 2$	83,43 ± 3,88
	$m = 12500, p = 4$	88,06 ± 4,55
	$m = 12500, p = 8$	85,97 ± 3,53
	$m = 6250, p = 2$	87,01 ± 2,63
	$m = 6250, p = 4$	87,01 ± 3,52
	$m = 6250, p = 8$	86,11 ± 1,99
	$m = 3125, p = 2$	85,52 ± 2,33
	$m = 3125, p = 4$	88,65 ± 2,74
	$m = 3125, p = 8$	87,31 ± 2,35
PCM	A = 87,6	90,74 ± 4,02
	$u = 255$	89,70 ± 2,76

Fuente: elaboración propia

CONCLUSIONES

El mayor porcentaje de acierto fue de 90.74±4.02% obtenido usando PCM no lineal ley A; este tipo de codificación en realidad no comprime las señales de voz, sino que trata de mejorar la relación señal a ruido de estas, particularmente para muestras pequeñas;

sin embargo, frente a la forma en que fueron codificadas las señales de voz en la base de datos utilizada (25 kHz a 16 bits), sí existe una reducción significativa del tamaño de los datos, lo que paradójicamente da como resultado una mejoría pequeña en la precisión (90.74±4.02% frente a 85,07 ± 2,98%). En el caso de señales comprimidas por DCT también se observa que, aunque en esta caso también existe una reducción significativa en el volumen de datos, el porcentaje de acierto mejora nuevamente un poco (88,06 ± 4,55). Esto permite concluir que es posible utilizar señales de voz comprimidas para detectar patologías, lo que abriría la posibilidad de hacer telediagnóstico de patologías de voz a través de redes telefónicas.Como trabajo futuro se propone la expansión de la metodología propuesta a otros tipos de técnicas de compresión de voz y estudiar la viabilidad de implementar un sistema de telediagnóstico de patologías de voz.

AGRADECIMIENTOS

Los autores agradecen a la Universidad de Medellín por la financiación del proyecto "Validación de algoritmos de aprendizaje automático usados en sistemas de telecomunicaciones de radio cognitiva (Código No. 737)", de cuyos resultados se desprende el presente trabajo y al programa "Ayudas para la realización del Doctorado (RR01/2011) de Universidad Politécnica de Madrid y TEC2012-38630-C04-01" del Ministerio de Educación de España

REFERENCIAS BIBLIOGRÁFICAS

[1] S. Kadambe and P. Srinivasan, "Adaptive wavelets for signal classification and compression," *AEU-International Journal of Electronics and Communications,* vol. 60, pp. 45-55, 2006.

[2] G. Rajesh*, et al.*, "Speech compression using different transform techniques," in *Computer and Communication Technology (ICCCT), 2011 2nd International Conference on*, 2011, pp. 146-151.

[3] I. Otung, *Communication engineering principles*: Palgrave Macmillan, 2001.

[4] B. Sklar, *Digital communications* vol. 2: Prentice Hall NJ, 2001.

[5] T. ITU, "Recommendation G. 711," *Pulse Code Modulation (PCM) of voice frequencies, November,* 1988.

[6] R. ITU-T and I. Recommend, "P. 800," *Methods for subjective determination of transmission quality,* 1996.

[7] S. Haykin, *Communication systems*: John Wiley & Sons, 2008.

[8] J. Davidson, *Voice over IP fundamentals*: Cisco press, 2006.

[9] D. Peña, *Análisis de datos multivariantes* vol. 24: McGraw-Hill Madrid, 2002.

[10] J. B. Alonso*, et al*., "Automatic detection of pathologies in the voice by HOS based parameters," *EURASIP Journal on Applied Signal Processing,* vol. 4, pp. 275-284, 2001.

[11] G. Banci*, et al*., "Vocal fold disorder evaluation by digital speech analysis," *Journal of Phonetics,* vol. 14, pp. 495-499, 1986.

[12] B. Boyanov and S. Hadjitodorov, "Acoustic analysis of pathological voices. A voice analysis system for the screening of laryngeal diseases," *Engineering in Medicine and Biology Magazine, IEEE,* vol. 16, pp. 74-82, 1997.

[13] A. A. Dibazar*, et al*., "Feature analysis for automatic detection of pathological speech," in *Engineering in Medicine and Biology, 2002. 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, 2002. Proceedings of the Second Joint*, 2002, pp. 182-183 vol.1.

[14] M. K. Arjmandi and M. Pooyan, "An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine," *Biomedical Signal Processing and Control,* vol. 7, pp. 3-19, 2012.

[15] R. J. Moran*, et al*., "Telephony-based voice pathology assessment using automated speech analysis," *Biomedical Engineering, IEEE Transactions on,* vol. 53, pp. 468-477, 2006.

[16] M. F. Kaleem*, et al*., "Telephone-quality pathological speech classification using empirical mode decomposition," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 7095-7098.

[17] N. Saenz-Lechon*, et al*., "Effects of Audio Compression in Automatic Detection of Voice Pathologies," *Biomedical Engineering, IEEE Transactions on,* vol. 55, pp. 2831-2835, 2008.

[18] D. Arifianto, "Enhancement of speech over wireless network using sinusoidal modeling and synthesis," in *Signal Processing Systems (SiPS), 2013 IEEE Workshop on*, 2013, pp. 301-305.

[19] V. Uloza*, et al*., "Exploring the feasibility of smart phone microphone for measurement of acoustic voice parameters and voice pathology screening," *European Archives of Oto-Rhino-Laryngology,* pp. 1-9, 2015/07/11 2015.

[20] M. Eye and E. Infirmary, "Voice disorders database, version. 1.03 (cd-rom)," *Lincoln Park, NJ: Kay Elemetrics Corporation,* 1994.

[21] G. Smillie, *Analogue and digital communication techniques*: Butterworth-Heinemann, 1999.

[22] S. Karapantazis and F.-N. Pavlidou, "VoIP: A comprehensive survey on a promising technology," *Computer Networks,* vol. 53, pp. 2050-2090, 2009.

[23] A. R. Madane*, et al*., "Speech compression using Linear predictive coding," in *proceedings International workshop on Machine Intelligence Research MIR labs*, 2009.

[24] M. Hasegawa-Johnson, "Lecture notes in speech production, speech coding, and speech recognition," *class notes, University of Illinois at Urbana-Champaign, Fall,* 2000.

[25] L. M. Sepúlveda Cano, «Análisis dinámico de relevancia en bioseñales," Universidad Nacional de Colombia-Sede Manizales, 2013.

[26] A. F. Quiceno Manrique, "Análisis tiempo-frecuencia por métodos no paramétricos orientado a la detección de patologías en bioseñales," Universidad Nacional de Colombia-Sede Manizales, 2009.