

Desarrollo de un codificador para audio cardiaco

Roilhi F. Ibarra, Miguel A. Alonso, Salvador Villarreal, Roberto Conte

Depto. de Electrónica y Telecomunicaciones

Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE)

Carretera Ensenada-Tijuana No. 3918, CP 22860, Ensenada, B.C., México.

Email: {frajo, aalonso, svillar, conte}@cicese.edu.mx

Resumen—En este trabajo se describe el desarrollo un codificador para señales de audio cardiaco o fonocardiogramas (FCG), los cuales son representados por medio de un conjunto de formas de onda elementales llamadas átomos de Gabor. Mediante el algoritmo Matching Pursuit (MP) se seleccionan los átomos que mejor coincidan con la estructura local de la señal fonocardiográfica, sin embargo, en MP se tiene una señal residual que se considera aún representa información importante para el análisis del FCG. Este residuo se ha modelado con técnicas autorregresivas, en particular predicción lineal (LPC). La suma de los modelados en MP y LPC conforman el codificador, el cual es evaluado mediante pruebas de carácter objetivo, tales como el coeficiente de correlación, el porcentaje de distorsión y el porcentaje de compresión.

I. INTRODUCCIÓN

La auscultación cardiaca es un método no invasivo, económico y sencillo para el diagnóstico primario de enfermedades cardiovasculares, primera causa de defunciones en el mundo [1] y segunda en nuestro país. Por medio de la auscultación se obtiene una señal de audio llamada fonocardiograma (FCG), producto de la actividad mecánica del corazón, en particular de la apertura y cierre de las válvulas cardíacas. Varias anomalías cardíacas son detectables por medio de un proceso de auscultación precisamente ejecutado, lo cual se traducirá en la herramienta más efectiva a nivel costo al evitar realizar estudios más costosos y complejos [2].

Las técnicas tradicionales de auscultación involucraban solamente el juicio de los expertos en la salud por medio de la audición solamente, lo cual introduce una variabilidad en la percepción y apreciación de los sonidos afectando la precisión en el diagnóstico [3].

Sin embargo, los estetoscopios actuales cuentan con herramientas necesarias para adquirir y visualizar la forma de onda de la señal fonocardiográfica, extendiendo el análisis al intervalo de frecuencias donde la capacidad del oído humano es limitado. La concentración en energía permite caracterizar en dominio frecuencial al FCG en tareas de diagnóstico e identificación de patologías.

El análisis de sonidos cardiacos asistido por computadora ha despertado un gran interés en esta era informática, como consecuencia se han propuesto en el estado del arte algunos trabajos en el modelado de fonocardiogramas. La compresión de señal es un proceso siempre presente en el análisis asistido por computadora, es importante que la señal sea transmitida con la menor cantidad de recursos y extrayendo la mayor cantidad de información posible. El objetivo de este trabajo

es realizar un esquema de compresión para señales FCG que asegure estos dos lineamientos, con ello se economizarán los recursos en memoria sin perder información reelevante del audio cardiaco para ser usado en el diagnóstico por expertos de la salud.

Este artículo se organiza de la siguiente manera: en la sección II se describen los pasos y la metodología para desarrollar el codificador. Después, en la sección III se evalúa el rendimiento del codificador por medio de pruebas objetivas calculando los porcentajes de compresión y de distorsión además de los coeficientes de correlación generados al comparar las señales originales y reconstruidas por el codificador. Por último, la sección V muestra las conclusiones surgidas a partir de los resultados generados. También en esta sección se presentan algunas recomendaciones para el mejoramiento del codificador.

II. DESARROLLO DEL CODIFICADOR

En esta sección se describe el procedimiento de codificación propuesto para el FCG. El modelo consta de dos partes, una parte armónica y una parte estocástica, el fonocardiograma $x(t)$ se representa mediante la siguiente expresión:

$$x(t) = x_h(t) + x_n(t), \quad (1)$$

donde $x_h(t)$ corresponde a la parte armónica, producto de correlacionar los eventos principales con formas de onda de esta característica. Por otra parte, $x_n(t)$ corresponde a la resta entre la señal principal y la parte armónica, cuyo carácter es *estocástico* dada su baja correlación con las ondas elementales. Desde (II) se representa a $x(t)$ en energía, $x_h(t)$ contiene un 99 % y su modelado se basa en el método de Matching Pursuit (MP), $x_n(t)$ contiene el resto de la energía y es representado por técnicas autorregresivas (AR). El procedimiento de codificación propuesto para el FCG $x(t)$ se ilustra a manera de diagrama en la Figura 1.

II-A. Algoritmo Matching Pursuit

La *representación dispersa* es una técnica de compresión que bajo la suma ponderada de un conjunto de M formas de ondas elementales y un término residual $R_M(t)$ conforma la reconstrucción en energía de una señal $x(t)$ de la siguiente manera:

$$x(t) = \sum_{m=1}^M \alpha_m \cdot g_{\gamma_m}(t) + R_M(t), \quad (2)$$

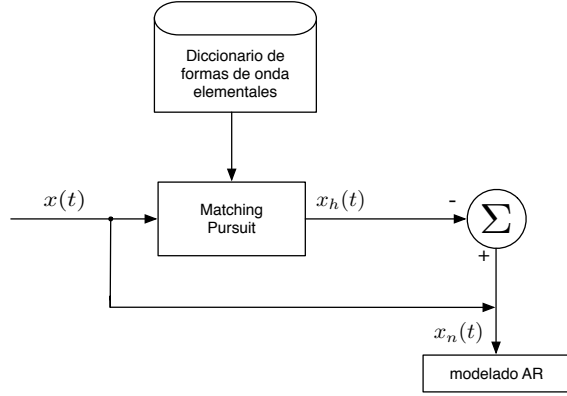


Figura 1. Diagrama a bloques del codificador para audio cardiaco propuesto.

donde cada forma de onda o átomo $g_{\gamma_m}(t)$ pertenece a un diccionario redundante de funciones $\mathcal{D} = \{g_{\gamma}(t), \gamma \in \Gamma\}$, y $g_{\gamma}(t) \in \mathcal{L}^2(\mathbb{R})$.

Seleccionar las formas de onda elementales óptimas para una representación dispersa es la tarea básica de Matching Pursuit (MP) [4], un algoritmo iterativo y codicioso cuyo objetivo es encontrar los átomos que mejor coincidan con la estructura local de la señal. Este procedimiento es realizado mediante el cálculo del valor máximo del producto interno entre la señal y un conjunto de ondas llamado *diccionario*. En cada una de las M iteraciones la selección de los átomos es realizada, además de el cálculo de la señal residual $R_M(t)$, producto de la resta entre la señal $x(t)$ y el átomo mayormente correlado seleccionado en la iteración. Debido a las características en tiempo-frecuencia de los elementos atómicos del diccionario la señal es bien representada en energía en ambos dominios. El procedimiento completo es presentado en el Algoritmo 1.

Un número infinito de iteraciones no garantiza la reconstrucción por completo en energía de la señal $x(t)$, como consecuencia, $R_M(t)$ nunca tendrá un valor de energía igual a cero. Después de un número significativo de M iteraciones en MP las formas de onda atómicas no tendrán un alto valor de correlación con la estructura local de la señal. Por esta razón, debe usarse un criterio para detener al algoritmo MP. En este trabajo se ha considerado tomar las M formas de onda elementales mayormente correladas tal que se logre retener en su suma ponderada el 99 % de la energía de la señal.

II-B. Selección del diccionario

La representación dispersa de una señal mediante representación escasa involucra la selección de un diccionario \mathcal{D} , el cual corresponde a un conjunto de onda elementales llamadas átomos. Una combinación lineal de átomos del diccionario seleccionados mediante MP representa en contenido energético a la señal $x(t)$. Las formas de onda atómicas seleccionadas que estén bien concentradas en el plano tiempo-frecuencia son imprescindibles en el procesamiento de señales de audio, esto

Algoritmo 1 Matching Pursuit

Entrada: $x(t)$, $\mathcal{D} = \{g_{\gamma}(t), \gamma \in \Gamma\}$

Salida: $\alpha_m, g_{\gamma_m}(t)$

$$R_M(t) = x(t)$$

$$\alpha_m = 0$$

repetir

$$g_{\gamma_m}(t) = \arg \max_{\gamma \in \Gamma} |\langle R_M(t), g_{\gamma}(t) \rangle|$$

$$\alpha_m = \langle R_M(t), g_{\gamma_m}(t) \rangle$$

$$R_M(t) = R_M(t) - \alpha_m \cdot g_{\gamma_m}(t)$$

hasta que Se haya alcanzado alguna SNR tope, porcentaje de energía substraída o la M -ésima iteración.

garantiza que la energía de la señal pueda representarse en ambos dominios con precisión.

Algunos diccionarios tiempo-frecuencia han sido reportados en la literatura para la descomposición MP de señales de audio, en [5] los paquetes de ondas de cosenos discretos modificados (MDCT), de ondículas (wavelets) y diccionarios de Gabor se emplean para la codificación de señales de audio y posteriormente se prueba su desempeño.

Los resultados de los trabajos reportados en [6, 7] y la experiencia previa han determinado para este trabajo la selección de diccionarios de Gabor en el análisis y síntesis de señales fonocardiográficas.

Los átomos de un bloque de Gabor son formas de onda *bien concentradas* en tiempo y frecuencia dadas por la dilatación, modulación y traslación de una ventana gaussiana $w(t)$, generalmente positiva, evaluada en los reales y de energía unitaria:

$$\int |w(t)|^2 dt = 1$$

$$g_{\gamma}(t) = \frac{1}{\sqrt{s}} w\left(\frac{t-u}{s}\right) e^{j2\pi\xi(t-u)}, \quad (3)$$

donde $j = \sqrt{-1}$, $w(t) = \sqrt[4]{2} e^{-\pi t^2}$ es la ventana, el factor de escala s controla la anchura de la envolvente, u es el desplazamiento en tiempo que define la locación del átomo modulado en la frecuencia ξ . Estos parámetros definen un elemento $\gamma_m = (s_m, u_m, \xi_m)$, donde cada índice γ_m pertenece al conjunto Γ .

II-C. Codificación predictiva lineal

Se considera que la señal residual $R_M(t)$ tiene un carácter aleatorio debido a su baja correlación con los átomos de la parte armónica. Existen varios modelados de tipo autorregresivo (AR) en la literatura para la representación de este tipo de señales.

La codificación predictiva lineal es un procedimiento de modelado AR ampliamente usado en señales de voz [8, 9], esta técnica tiene como suposición que una muestra de un proceso estocástico estacionario es representada por la combinación lineal de las p muestras anteriores:

$$\hat{r}_n = - \sum_{k=1}^p a_k r_{n-k}, \quad (4)$$

donde $n = 0, 1, \dots, N-1$ y los coeficientes a_k 's pertenecen a un conjunto de coeficientes de un filtro de respuesta al

impulso infinita de p -ésimo orden, calculado en términos del criterio del error cuadrático mínimo (MMSE) $e(n)$, dado por la diferencia entre el conjunto de muestras de la señal original y las muestras predichas por el procedimiento, esto es

$$e(n) = r(n) - \hat{r}(n).$$

Los p coeficientes a_k predictores se pueden calcular desde la autocorrelación o autocovarianza de la señal. Para esto, un conjunto de p ecuaciones lineales se resuelven mediante el algoritmo de recursión de Levinson [10]. El uso de la autocorrelación como punto de partida en la predicción lineal tiene como consecuencia que se realice un método para aproximar el espectro en frecuencia de la señal [11].

III. PRUEBAS OBJETIVAS

En esta sección se evalúa al codificador desarrollado mediante medidas objetivas tales como el coeficiente de correlación y el porcentaje de distorsión. Se ha tomado como referencia el conjunto de ruidos cardiacos correspondientes a diferentes murmullos. Las señales se encuentran disponibles en línea desde [12]. Se han tomado en cuanto a las partes armónica y estocástica las siguientes determinaciones en los parámetros:

- Una cantidad M de átomos tal que se extraiga el 99% de la energía en los eventos cardiacos. Se tuvieron en promedio 15 átomos por sonido (normal o murmullo).
- Un conjunto de $p = 10$ coeficientes del predictor lineal para el modelado del residual (1% en energía restante), cantidad suficiente debido al blanqueamiento producido por el filtro IIR generado.

En la Figura 2 se muestran las ondas generadas por el audio original y el audio reconstruido mediante el procedimiento de codificación propuesto. Es evidente la similitud entre ambas morfologías.

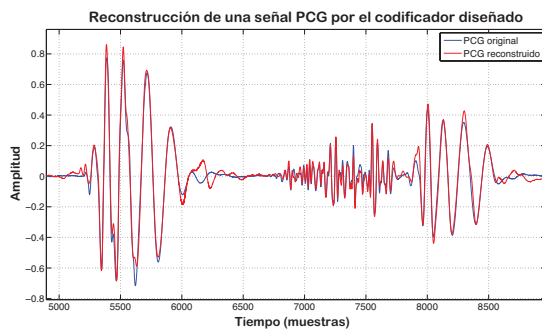


Figura 2. Comparación en forma de onda de un murmullo sistólico tardío del audio original y el reconstruido por el codificador.

III-A. Porcentaje de compresión

La tasa o porcentaje de compresión es la primera medida de evaluación. Esta operación está definida como la razón entre el número de bits de la señal comprimida (codificada) N_{bits_c}

y el número de bits de la señal original N_{bits_o} , de la siguiente manera:

$$P_c = \frac{N_{bits_o}}{N_{bits_c}}. \quad (5)$$

En la tabla I se muestran los resultados del cálculo del porcentaje de compresión para las referencias de la base de datos analizada. Debido a que los fonocardiogramas se presentaron en formato .wav se consideraron los casos de 8 y 16 bits por muestra para una frecuencia de muestreo $f_s = 8KHz$, resultando en velocidades de transmisión de 64 y 128 kilobits por segundo.

Cuadro I
PORCENTAJES DE COMPRESIÓN OBTENIDOS PARA LAS SEÑALES DE LA BASE DE DATOS [12].

Nombre de la señal	P_c @124kbps	P_c @64kbps
Soplo diastólico	93.24	86.47
Clic de eyección	93.88	87.68
Murmullo sistólico temprano	93.35	86.70
Murmullo sistólico tardío	93.30	86.61
Chasquido de apertura	94.06	88.13
S3	93.99	87.98
S4	94.06	88.16
Murmullo pansistólico	93.00	86.00
División normal de S1	94.49	88.98
División normal de S2	94.21	88.43

III-B. Coeficiente de correlación

El grado de intensidad que guardan dos variables $x(t)$ y $\hat{x}(t)$ matemáticamente es descrito mediante el coeficiente de correlación $\rho_{x,\hat{x}}$. Esta medida que cuantifica y describe esta relación por medio de la razón de las covarianzas entre ambas variables, esto es:

$$\rho_{x,\hat{x}} = \frac{\sigma_{x\hat{x}}}{\sigma_x \sigma_{\hat{x}}} = \frac{\mathbb{E}[(x - \mu_x)(\hat{x} - \mu_{\hat{x}})]}{\sigma_x \sigma_{\hat{x}}}, \quad (6)$$

donde $\sigma_{x\hat{x}}$ es la covarianza de las variables x y \hat{x} ; σ_x es la desviación estándar de la variable x y $\sigma_{\hat{x}}$ es la desviación estándar de la variable \hat{x} respectivamente. $\mathbb{E}(x)$ es el valor esperado de la variable aleatoria x y μ_x , $\mu_{\hat{x}}$ son las medias de las variables x y \hat{x} respectivamente.

La magnitud de $\rho_{x,\hat{x}}$ varía entre 0 y 1, representando una mayor grado de intensidad al acercarse a la unidad. Los resultados del cálculo del coeficiente de correlación entre las señales de audio de la base de datos y los audios reconstruidos por el codificador se muestran en la Tabla II.

IV. PORCENTAJE DE DISTORSIÓN (PRD)

El proceso de compresión-descompresión con pérdidas de la señal de audio involucra la introducción de distorsión con respecto a la señal original. Una medida que cuantifica el porcentaje de esta degradación es el porcentaje dado por la raíz cuadrada de la diferencia cuadrática media (*Percent Root*

Cuadro II
COEFICIENTES DE CORRELACIÓN CALCULADOS ENTRE LAS SEÑALES DE REFERENCIA Y RECONSTRUIDAS MEDIANTE EL CODIFICADOR.

Nombre de la señal	Coefficiente de correlación
Soplo diastólico	0.9697
Clic de eyección	0.9724
Murmulo sistólico temprano	0.9335
Murmulo sistólico tardío	0.9330
Chasquido de apertura	0.9406
S3	0.9614
S4	0.9638
Murmulo pansistólico	0.9648
División normal de S1	0.9634
División normal de S2	0.9675

Mean Square Difference, PRD)[13], la cual es calculada de la siguiente manera:

$$PRD = \sqrt{\frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{\sum_{i=1}^N (x_i - \mu_x)^2}} \times 100, \quad (7)$$

donde x_i es la secuencia o señal original, \hat{x}_i es la señal reconstruida y μ_x la media de la secuencia x .

En la tabla III se muestran los resultados obtenidos para el cálculo de la PRD que se introdujo en el procedimiento de compresión a los fonocardiogramas por el codificador propuesto.

Cuadro III
PORCENTAJE DE DISTORSIÓN PRD CALCULADO PARA LAS SEÑALES FCG ANALIZADAS TRAS EL PROCESO DE COMPRESIÓN-DESCOMPRESIÓN.

Nombre de la señal	PRD
Soplo diastólico	2.50
Clic de eyección	2.38
Murmulo sistólico temprano	2.33
Murmulo sistólico tardío	2.41
Chasquido de apertura	2.15
S3	2.87
S4	2.77
Murmulo pansistólico	2.74
División normal de S1	2.77
División normal de S2	2.61

V. CONCLUSIONES

En este trabajo se propone la metodología para el desarrollo de un codificador para señales audio cardiaco con pérdidas, basado en un modelado paramétrico mediante las técnicas de Matching Pursuit y Predicción Lineal. Hasta el momento en la literatura no se ha encontrado algún codificador propuesto basado en un modelo similar. El codificador presenta resultados satisfactorios en cuanto a reconstrucción de la forma de onda, alta tasa de compresión (más del 90% considerando 16 bits por muestra), alto coeficiente de correlación y bajo porcentaje de distorsión (menor al 3%) tras ser aplicado a una base de datos de murmullos cardiacos.

Como futuro de esta investigación se completará el proceso mediante la cuantificación y la presentación de los bits. Se pretende evaluar el desempeño del codificador en redes de bajas tasas de datos y corto alcance que sean aplicables al entorno clínico.

RECONOCIMIENTOS

El primer autor agradece al Consejo Nacional de Ciencia y Tecnología (CONACYT) por financiar la presente publicación bajo el CVU No. 477876.

REFERENCIAS

- [1] W. H. Organization *et al.*, *Global status report on alcohol and health-2014*. World Health Organization, 2014.
- [2] C. B. Mahnke, "Automated heartsound analysis/computer-aided auscultation: A cardiologist's perspective and suggestions for future development," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. IEEE, 2009, pp. 3115–3118.
- [3] A. K. Abbas and R. Bassam, "Phonocardiography signal processing," *Synthesis Lectures on Biomedical Engineering*, vol. 4, no. 1, pp. 1–194, 2009.
- [4] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *Signal Processing, IEEE Transactions on*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [5] E. Ravelli, G. Richard, and L. Daudet, "Union of mdct bases for audio coding," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 16, no. 8, pp. 1361–1372, 2008.
- [6] C. I. Nieblas, M. A. Alonso, R. Conte, and S. Villarreal, "High performance heart sound segmentation algorithm based on matching pursuit," in *Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE), 2013 IEEE*. IEEE, 2013, pp. 96–100.
- [7] X. Zhang, L.-G. Durand, L. Senhadji, H. C. Lee, and J.-L. Coatrieux, "Analysis-synthesis of the phonocardiogram based on the matching pursuit method," *Biomedical Engineering, IEEE Transactions on*, vol. 45, no. 8, pp. 962–971, 1998.
- [8] J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, 1975.
- [9] P. Vaidyanathan, "The theory of linear prediction," *Synthesis lectures on signal processing*, vol. 2, no. 1, pp. 1–184, 2007.
- [10] N. Levinson, "The wiener rms (root mean square) error criterion in filter design and prediction," 1947.
- [11] J. Makhoul, "Spectral analysis of speech by linear prediction," *Audio and Electroacoustics, IEEE Transactions on*, vol. 21, no. 3, pp. 140–148, 1973.
- [12] Litmann, "Ruidos cardiacos y pulmonares," 2013. [Online]. Available: http://solutions.3m.com.mx/wps/portal/3M/es_MX/3M-Litmann-LA/home/Education/SoundLibrary/
- [13] R. J. Ruiz Merino *et al.*, "Comprensión de fonocardiogramas basada en eventos," 2005.