# CUANTIFICACIÓN DE SEÑALES DE VOZ UTILIZANDO WAVELETS


**María Manuela Silva Zambrano**

*Universidad del Cauca*

**Facultad de Ingeniería Electrónica y Telecomunicaciones**
**Departamento de Telecomunicaciones**
**Grupo de Nuevas Tecnologías en Telecomunicaciones – GNTT**
**Línea de Señales y Sistemas de Telecomunicaciones**
**Popayán, 2022**

# CUANTIFICACIÓN DE SEÑALES DE VOZ UTILIZANDO WAVELETS


Trabajo de grado presentado como requisito para obtener el título de Magister en Electrónica y Telecomunicaciones

**María Manuela Silva Zambrano**

Director: Msc. Jesús Mauricio Ramírez Viáfara
Codirector: Msc. Harold Armando Romo Romero

*Universidad del Cauca*

**Facultad de Ingeniería Electrónica y Telecomunicaciones**
**Departamento de Telecomunicaciones**
**Grupo de Nuevas Tecnologías en Telecomunicaciones – GNTT**
**Línea de Señales y Sistemas de Telecomunicaciones**
**Popayán, 2022**

# TABLA DE CONTENIDO

# LISTA DE FIGURAS

# LISTA DE TABLAS

# LISTA DE ACRÓNIMOS

**ACELP** *Algebraic Code-Excited Linear Prediction,* Predicción Lineal Excitada por Códigos Algebraicos.

**ADC** *Analog to Digital Conversion,* Conversión Analógica-Digital.

**ADM** *Adaptative Delta Modulation* , Modulación Delta Adaptativa.

**ADPCM** *Adaptive Differential Pulse Code Modulation,* Modulación por Codificación de Pulsos Diferencial y Adaptativa.

**CDF** *Cumulative Distribution Function,* Función de Distribución Acumulativa.

**CT** *Cosine Transform,* Transformada Coseno.

**DAM** *Diagnostic Acceptability Measure,* Medida Diagnóstica de Aceptabilidad.

**DM** *Delta Modulation,* Modulación Delta.

**DPCM** *Differential Pulse Code Modulation,* Modulación por Codificación de Pulso Diferencial.

**DTW** *Dynamic Time Warping,* Deformación de Tiempo Dinámica.

**DWT** *Discrete Wavelet Transform,* Transformada *Wavelet* Discreta.

**ETSI** *European Telecommunications Standards Institute,* Instituto Europeo de Estándares de Telecomunicaciones.

**FT** *Fourier Transform,* Transformada de Fourier.

**FWT** *Fast Wavelet Transform,* Transformada Rápida *Wavelet.*

**GSM-EFR** *Global System for Mobile communications - Enhanced Full Rate,* Sistema Global para las Comunicaciones Móviles con Velocidad Máxima Mejorada.

**IDWT** *Inverse Discrete Wavelet Transform,* Transformada Inversa Discreta *Wavelet.*

**ITU-T** *International Telecommunications Union - Telecommunications*, Unión Internacional de Telecomunicaciones sector Telecomunicaciones.

**MDCT** *Modified Discrete Cosine Transform,* Transformada Discreta de Coseno Modificada.

**MFCC** *Mel Frequency Cepstral Coefficients,* Coeficientes Cepstrales en las Frecuencias de Mel.

**MOS** *Mean Opinion Score,* Nota Media de Opinión.

**MRA** *MultiResolution Analysis,* Análisis Multiresolución.

**MSE** *Mean Square Error,* Error Cuadrático Medio.

**MSSIM** *Mean Structural SIMilarity,* Índice de Similitud Estructural Promedio.

**MUSHRA** *MUltiple Stimuli with Hidden Reference and Anchor,* Múltiples Estímulos con Referencias Ocultas y Ancladas.

**PCM** *Pulse Code Modulation,* Modulación por Codificación de pulso.

**pdf** *Probability Density Function,* Función de Densidad de Probabilidad.

**PESQ** *Perceptual Evaluation of Speech Quality,* Evaluación Perceptual de la Calidad del Habla.

**QMF** *Quadrature Mirror Filter,* Filtro Espejo en Cuadratura.

**SNR** *Signal to Noise Ratio,* Relación Señal a Ruido.

**SPL** *Sound Pressure Level,* Nivel de Presión del Sonido.

**ST** *Sine Transform,* Transformada Seno.

**STFT** *Short Time Fourier Transform,* Transformada de Fourier de Tiempo Corto.

**TFTP** *Tight Framelet Packet Transform,* Transformada de Paquetes de Marco Estricto.

**TIA-EIA** *Telecommunications Industry Association - Electronic Industries Alliance,* Asociación de la Industria de las Telecomunicaciones y la Alianza de la Industria Electrónica.

**WP** *Wavelet Packet*, Paquetes *Wavelet.*
**WT** *Wavelet Transform,* Transformada *Wavelet.*

# INTRODUCCIÓN

La cuantificación de señales de voz, que por naturaleza son analógicas, se pude utilizar para su digitalización o compresión. Independientemente de su aplicación, la pérdida de información degrada la calidad de este tipo de señales y puede comprometer incluso la inteligibilidad del mensaje, debido a lo cual se hace necesario buscar alternativas para el diseño e implementación del cuantificador.

La cuantificación es el proceso de limitar las amplitudes de una señal a un conjunto finito de $N$ posibles amplitudes. En esencia, la función de un cuantificador consiste en redondear el valor de cada muestra de una señal al nivel de cuantificación más próximo, debido a lo cual se introduce distorsión, i.e., se pierde información. La idea es minimizar la diferencia entre la secuencia original y la secuencia cuantificada de acuerdo con algún criterio de distorsión. Sin embargo, sea cual sea ese criterio, la diferencia se minimiza en la medida que el número de niveles, $N$, se hace grande. Así mismo, un valor grande de $N$ implica un mayor consumo de recursos (mayor cantidad de bits para transmitir, procesar o almacenar la información). Esto hace que el número de niveles no se pueda incrementar deliberadamente (Gallager, 2008).

En general, un cuantificador se puede diseñar desde dos perspectivas. La primera de ellas busca maximizar la fidelidad de la señal (minimizar distorsión), manteniendo constante el número de niveles de cuantificación. La segunda intenta reducir el número de niveles de cuantificación, manteniendo constante la distorsión. Se tiene un compromiso entre calidad y eficiencia según las necesidades y limitaciones de una aplicación en particular.

La ventaja inherente de un esquema de cuantificación basado en la transformada *wavelet* radica en el hecho de que es posible contar con una noción de lo que ocurre tanto en tiempo como en frecuencia de manera simultánea, haciendo que el proceso de cuantificación/compresión se pueda amoldar más fácilmente a las características intrínsecas de la señal. Por otro lado, las operaciones de descomposición y reconstrucción de la señal son fácilmente implementables a través del algoritmo de Mallat, haciendo que este tipo de cuantificación sea una alternativa válida y atractiva en contraposición a las técnicas de cuantificación basadas únicamente en el comportamiento espectral de la señal (Fourier), o basadas únicamente en el comportamiento temporal de la señal (cuantificación con y sin memoria).

En este trabajo de grado de maestría se proponen dos algoritmos de cuantificación basados en la transformada *wavelet* y se evalúa su desempeño[1] con respecto a la cuantificación en el dominio del tiempo.

[1] Grado de similitud entre la señal original y la cuantificada según medidas objetivas y subjetivas.

# CAPÍTULO 1: GENERALIDADES

La voz humana, por su naturaleza, es analógica y su transformación a su versión digital implica pérdida de información, lo cual conlleva a un compromiso entre la calidad y el tamaño[2] de la voz digitalizada. La información prescindible es aquella que no puede ser captada por el sistema auditivo humano, no obstante, para su detección se requiere entender el funcionamiento de este sistema. Adicionalmente, se debe analizar la forma en la que se produce la voz y sus características.

## 1.1. OÍDO

Como primera medida, se busca responder a la pregunta ¿Qué es el sonido?

El sonido es una onda mecánica longitudinal producida por un cuerpo vibrante. Por tratarse de una onda mecánica, requiere de un medio elástico para su propagación. Por otro lado, el carácter longitudinal se refiere al hecho de que la dirección de propagación es la misma de la perturbación que ésta genera en el medio. Dicha perturbación consiste en cambios de presión, haciendo que el medio se comprima y se rarefaga[3] sucesivamente (Berg, 1982).

La Real Academia de la Lengua Española define al sonido como: “sensación producida por el órgano del oído por el movimiento vibratorio de los cuerpos, transmitido por un medio elástico, como el aire” (*Sonido | Definición de Sonido - Diccionario de La Lengua Española - Edición Del Tricentenario*, n.d.). Esta definición agrega que el sonido debe ser captado por el órgano del oído. Al respecto, los humanos y otros seres vivos han desarrollado la capacidad de detectarlas e interpretarlas, además de producir intencionalmente sus propias señales cargadas de información, como el habla.

El oído es un órgano del cuerpo humano que está asociado a los sentidos de la audición y el equilibrio. Este órgano está compuesto por tres partes a saber: el oído externo, el oído medio y el oído interno, como se ilustra en la Figura 1.1.

En el oído, las ondas acústicas son captadas y transformadas en impulsos eléctricos que se interpretan en el cerebro. Este proceso requiere del correcto funcionamiento de cada una de las partes mencionadas. El pabellón auricular (parte visible del oído) se encarga de captar los sonidos, pero además su forma ayuda a determinar la dirección en la que se encuentra la fuente sonora y realza el rango de frecuencias en el que se encuentra la voz. Las ondas acústicas viajan como cambios de presión por el canal auditivo y provocan que el tímpano vibre. El tímpano es la frontera entre el oído externo y el oído medio, por lo que su vibración se traspasa al martillo,

[2] Como tamaño se entiende a la cantidad de bits necesaria para representar una muestra de voz.
[3] La rarefacción consiste es el proceso por el cual un gas se hace menos denso (RAE, n.d.).

yunque y estribo. Estos tres pequeños huesos se encargan de amplificar las vibraciones y enviarlas a la cóclea (oído interno), en la que se crean ondas sobre el fluido que alberga en su interior -endolinfa-. La cóclea cuenta, además, con los cilios celulares que se encuentran ubicados en la membrana basilar y resuenan a diferentes frecuencias, enviando impulsos eléctricos al nervio auditivo que posteriormente son enviados al cerebro (Khetrapal, 2019).

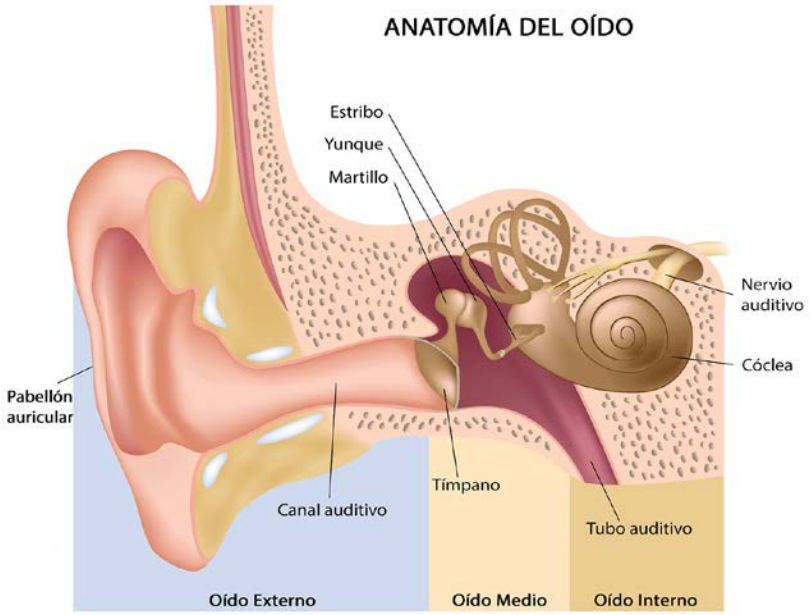


Figura 1.1. Anatomía del oído.
Tomado de: Khetrapal, 2019.

La información contenida en las ondas acústicas, al viajar a través de las diferentes partes del oído, cambia su representación. La información está contenida en los cambios de presión de la onda acústica, que se traducen en los movimientos del oído interno, el cual, a su vez, reproduce la información en la cóclea al provocar oscilaciones en su líquido interno. Las oscilaciones en líquido de la cóclea estimulan los cilios, que se encargan de la transducción al dominio eléctrico o electroquímico. Sin embargo, éstos no son los únicos cambios que sufre la información, ya que a lo largo de este proceso también cambia su naturaleza. Los cilios, encargados de discriminar las diferentes frecuencias presentes en el sonido y sus respectivas intensidades, realizan un proceso de conversión analógica-digital, puesto que su estimulación está supeditada a un rango de frecuencias y por lo tanto se tienen limitaciones de percepción que se explicarán con mayor profundidad en el modelo psicoacústico (Khetrapal, 2019).

Dado que el oído hace un proceso de conversión analógica-digital, se tiene una irremediable pérdida de información. Las técnicas de procesamiento de señales, especialmente aquellas enfocadas en la compresión de datos, realizan estudios sobre las limitaciones de los sentidos, de esta forma no se desperdician esfuerzos describiendo características que no pueden ser percibidas. En el caso de la audición

se tiene el modelo psicoacústico, el cual permite identificar qué puede ser escuchado y qué no (Sayood, 2006).

### 1.1.1. Modelo Psicoacústico

El proceso de audición depende de la frecuencia, es por esto que se trabaja con el Nivel de Presión del Sonido (SPL, *Sound Pressure Level*), el cual especifica el nivel de intensidad de un sonido necesario para ser escuchado según la frecuencia. Este parámetro se mide en decibeles (dB) y se calcula como $20\log_{10}\left(\frac{\rho}{\rho_o}\right)$, donde $\rho$ es la presión del estímulo auditivo, medida en Pascales (Pa), y $\rho_o$ es una presión de referencia, la cual se asume como $20\ \mu$Pa (Spanias et al., 2007).

Dado que la percepción de intensidad de un sonido varía dependiendo de la persona, las curvas de SPL corresponden a un promedio y sirven como guía para la percepción de la audición de las personas, ya que éstas muestran que las bajas frecuencias requieren de valores de SPL más elevados para ser escuchadas, en contraste con las altas frecuencias que requieren de valores de SPL menores. No obstante, esto se cumple dentro del rango de audición humano (20 a 20,000 Hz). Los valores de SPL ayudan a delimitar la frontera entre lo audible y lo inaudible, como se muestra en la Figura 1.2.

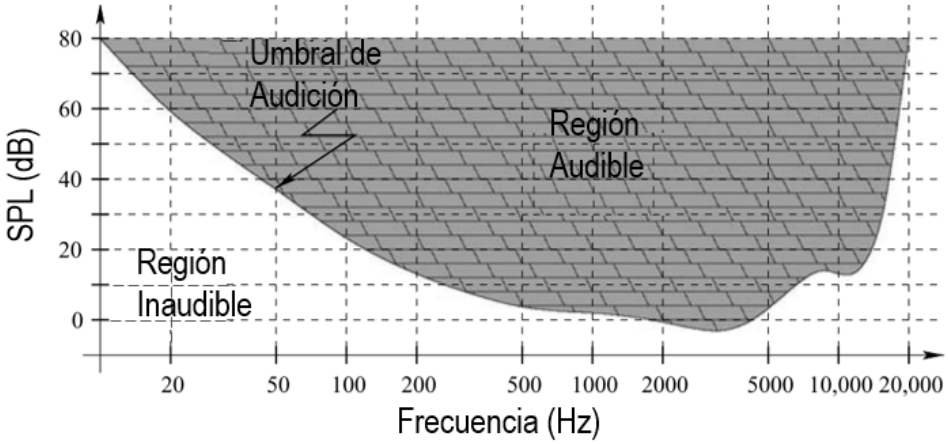


Figura 1.2. Curva del umbral de audición.
Adaptado de: Sayood, 2006.

La audición tiene lugar gracias a una transformación de posición a frecuencia. Como ya se dijo, los cilios celulares presentes en la cóclea resuenan a diferentes frecuencias. Desde la perspectiva del procesamiento de señales, la cóclea trabaja como un banco de filtros pasa banda cuyas bandas de paso se traslapan entre sí. El tamaño de dichas bandas de paso varía con la frecuencia, esto es, a mayor frecuencia, mayor el rango de frecuencias que el filtro deja pasar, como se muestra en la Figura 1.3.

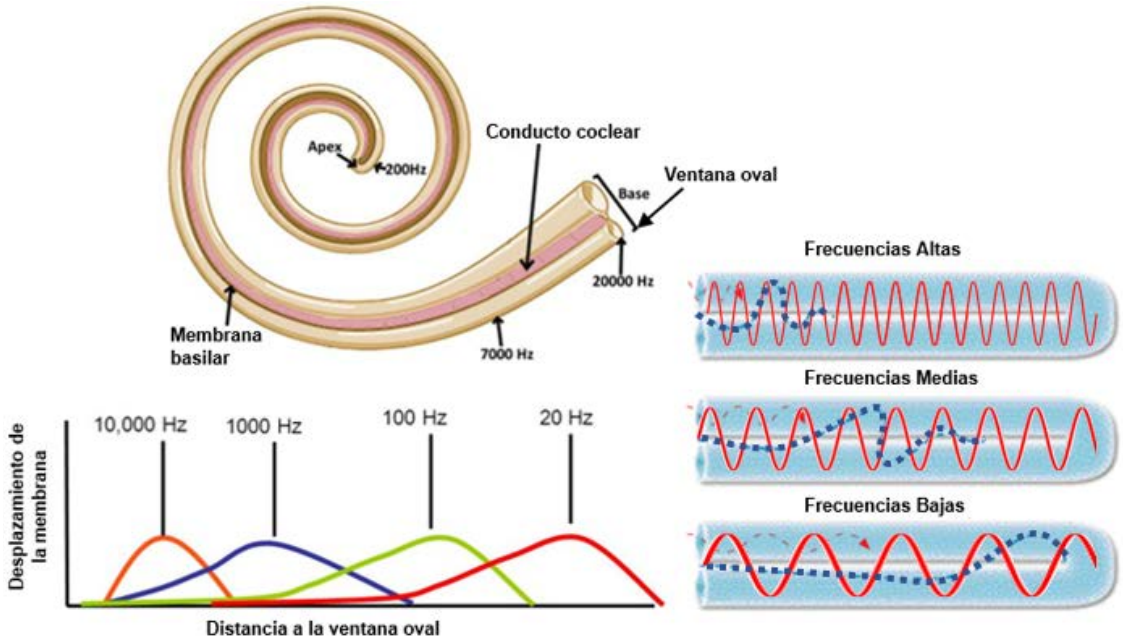


Figura 1.3. Funcionamiento de la cóclea.
Adaptado de: Alberta Education, 2009; Rich, 2015; scienceABC, 2017.

La detección de un tono en el oído se da porque, según su intensidad, se estimula una zona de la membrana basilar. Como consecuencia, se define la banda crítica como aquella zona de la membrana basilar que se excita ante la presencia de un tono de cierta intensidad, como se muestra en la Figura 1.4. Esta característica del sistema auditivo provoca el **enmascaramiento espectral de sonidos**, el cual consiste en la imposibilidad de distinguir dos tonos que ocurran dentro de la misma banda crítica, es decir, que el conjunto de tonos suma sus sonoridades y son captados como un único tono (Merino & Muñoz-Repiso, 2013).


Figura 1.4. Bandas críticas.
Tomado de: Merino & Muñoz-Repiso, 2013.

La excitación de la membrana basilar dura un cierto tiempo, por lo que existe otro tipo de enmascaramiento, el cual se presenta cuando los sonidos no ocurren de forma simultánea y se debe a la diferencia en los niveles de intensidad. En el **post-enmascaramiento**, un sonido de alta intensidad ocurre antes (30-60 ms) que el sonido enmascarado de menor intensidad, esto se debe a que después de percibir el sonido de alta intensidad, al cerebro le toma un tiempo adaptarse. Por otro lado, el enmascaramiento también puede ocurrir cuando el sonido enmascarado (de

menor intensidad) ocurre antes (5-10 ms) que el sonido enmascarador. Esto se conoce como **pre-enmascaramiento** y aunque es más difícil de explicar, dicho fenómeno se debe a la forma en la que el cerebro procesa los sonidos (Merino de la Fuente & Muñoz-Repiso, 2013).

### 1.1.2. Fase

Con el fin de complementar el estudio de la percepción del sonido, se debe hablar de la fase. Los primeros experimentos sobre el tema arrojaron que el oído era sordo a los cambios de fase, no obstante, estudios recientes han mostrado que el oído no es sordo frente a los cambios de fase, a pesar de que no tiene una alta sensibilidad, sí logra percibir dichos cambios (Alcántara et al., 2005; Kohlrausch & Sander, 2005; Laitinen et al., 2013).

La fase guarda relación con el desplazamiento temporal de las sinusoides, por lo que la alteración de la fase genera una forma de onda completamente diferente, lo cual se conoce como distorsión de fase. La Figura 1.5 muestra un ejemplo de distorsión de fase para una señal con cuatro (4) componentes espectrales.

$$x_1(t) = \cos\left(2\pi t + \frac{\pi}{3}\right) + 0.3\sin\left(3\pi t + \frac{\pi}{2}\right) - 0.5\cos\left(4\pi t + \frac{\pi}{4}\right)$$
$$x_2(t) = \cos(2\pi t) + 0.3\sin(3\pi t) - 0.5\cos(4\pi t)$$

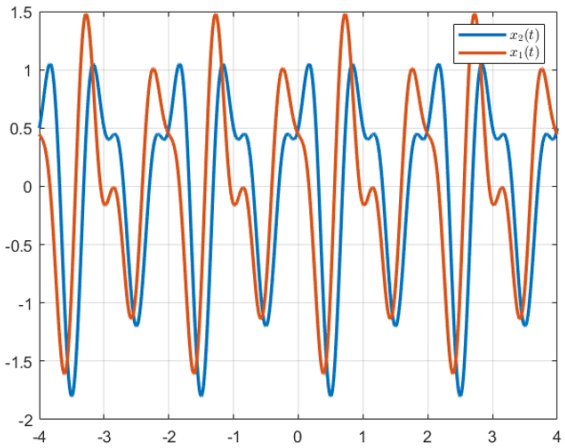


Figura 1.5. Distorsión de fase.
Elaboración propia.

Dentro de los resultados recientemente encontrados, se tiene que la distorsión de fase afecta la percepción de frecuencias cercanas (una octava menor y una octava mayor), pero no afecta la percepción de todas las frecuencias del sonido (Mowlaee et al., 2016). Adicionalmente, se han realizado estudios que demuestran que la precisión en el reconocimiento de sonidos mejora al aplicar un ecualizador de fase (Raitio et al., 2015).

## 1.2. VOZ

Este trabajo se concentra en el procesamiento de la voz humana, específicamente en su cuantificación, por lo que en esta sección se hablará de la generación de la voz desde un punto de vista anatómico y de sus características desde el punto de vista del procesamiento de señales.

### 1.2.1. Anatomía de la Voz

La generación de la voz en los humanos es responsabilidad del aparato fonador, que está compuesto por: el diafragma, los pulmones, la laringe, las cuerdas vocales, la faringe y las cavidades nasal y oral, como se muestra en la Figura 1.6.

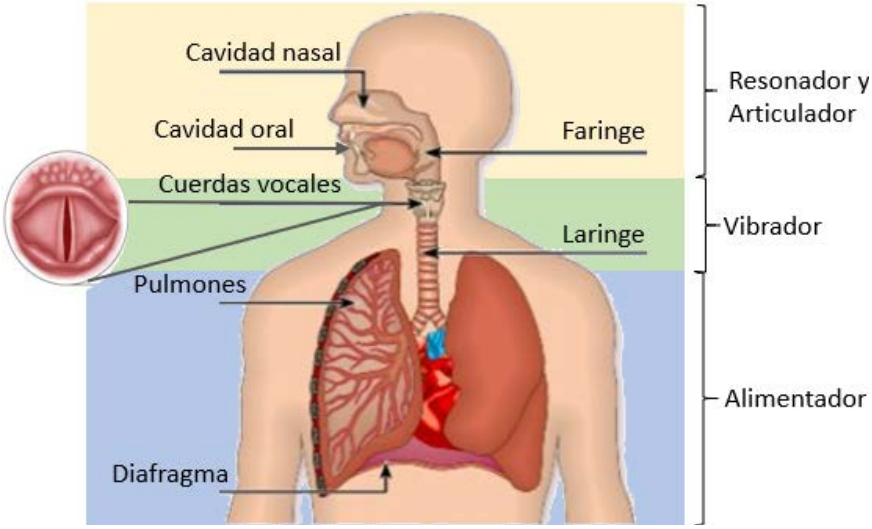


Figura 1.6. Aparato Fonador.
Adaptado de: Gonzáles, 2019.

La voz se compone de tres elementos: **tono**, **resonancia** y **articulación**. Se considera tonal al sonido que es producido por la vibración de las cuerdas vocales y es el resultado de la estimulación que la corriente de aire originada en los pulmones genera sobre éstas, las cuales entran en un ciclo de vibración en el que la presión del aire provoca la apertura de las cuerdas y el efecto Bernoulli las regresa a su estado original (cierre), generando un tren de pulsos de aire periódico (Figura 1.7) que corresponden a diferencias de presión, es decir, sonido (Loret, 2012; voice foundation, n.d.). La frecuencia del sonido se conoce como la frecuencia fundamental de la voz (*pitch*)[4], la cual depende de la longitud de las cuerdas vocales, por eso cambia en hombres y mujeres, ya que en los hombres su longitud está en el rango de 17 a 23 mm y en las mujeres está entre 12.5 y 17 mm.

[4] En adelante se utilizará el anglicismo por comodidad y abreviatura.

Figura 1.7. Cuerdas vocales.
Adaptado de: R Nave, n.d.

El tono está relacionado con dos características de la voz: intensidad y frecuencia. La intensidad de la voz depende de la corriente de aire que estimula las cuerdas vocales, por lo que su intensidad varía con el tiempo y está ligada a la voluntad de la persona que habla. El *pitch* depende de las características anatómicas de las cuerdas vocales, aunque éstas pueden variar su tensión y así conseguir pequeñas variaciones alrededor de dicha frecuencia, lo que se utiliza ampliamente en el canto.

La resonancia y la articulación son las encargadas de modular el tono puro para generar los diferentes sonidos, lo cual es controlado por el tracto vocal que está compuesto por: la faringe y las cavidades vocal y nasal. Por lo tanto, la apertura de la boca, la posición de la lengua y los labios, entre otros; determinan el sonido generado. Para analizar cómo es posible distinguir dichos sonidos se hace necesario observar lo que sucede en el dominio de la frecuencia, y es así como se cambia el enfoque anatómico por el de las señales y los sistemas.

### 1.2.2. Análisis de las Señales de Voz

Los bloques que representan la faringe y las cavidades nasal y vocal se modelan como funciones no lineales. Cuando la salida de un bloque contiene nuevas componentes de frecuencia (componentes adicionales a las de la entrada) se tiene entonces una relación no lineal entre la entrada, $x(t)$, y la salida, $y(t)$. Dicha relación no lineal equivale a que la salida es una función polinomial de la entrada, esto es, $y(t) = a_0 + a_1 x(t) + a_2 x^2(t) + \cdots + a_n x^n(t)$, lo que en el dominio de la frecuencia implica convoluciones del espectro de la entrada, $\tilde{y}(f) = a_0\delta(f) + a_1\tilde{x}(f) + a_2\tilde{x}(f) * \tilde{x}(f) + \cdots + a_n\tilde{x}(f) * \tilde{x}(f) * \cdots * \tilde{x}(f)$. Las nuevas componentes en frecuencia que se generan se pueden clasificar en dos categorías:

- Distorsión de armónicos: una componente en frecuencia (tono puro) corresponde a una señal sinusoidal y cualquier potencia $n$ de una señal sinusoidal equivale a una combinación lineal de senos y cosenos cuyas frecuencias son múltiplos enteros de la frecuencia del tono puro, tal como se presenta a continuación:

$$\cos^n(2\pi f_0 t) = \frac{1}{2^n}\left(e^{j2\pi f_0 t} + e^{-j2\pi f_0 t}\right)^n = \frac{1}{2^n}\sum_{k=0}^{n}\binom{n}{k} e^{j2\pi(n-2k)f_0 t}, \qquad (1.1)$$

$$\sin^n(2\pi f_0 t) = \frac{1}{(2j)^n}\left(e^{j2\pi f_0 t} - e^{-j2\pi f_0 t}\right)^n,$$
$$= \frac{1}{(2j)^n}\sum_{k=0}^{n}(-1)^k \binom{n}{k} e^{j2\pi(n-2k)f_0 t}, \qquad (1.2)$$

donde $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

- Distorsión de intermodulación: cuando la señal de entrada está compuesta por la combinación lineal de múltiples componentes en frecuencia y ésta se eleva a una potencia $n$, el resultado serán componentes armónicas y componentes mezcla. Con el fin de facilitar los cálculos teóricos se pueden emplear las identidades de la prostaféresis[5].

  De la mezcla de las componentes en frecuencia surgen componentes que no cumplen con la condición de ser un múltiplo entero de una frecuencia fundamental, haciendo que la señal resultante tienda a no ser periódica.

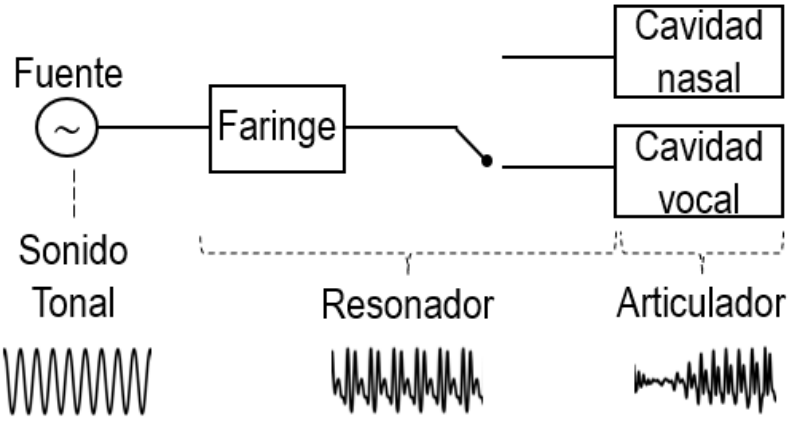


Figura 1.8. Diagrama de bloques tracto vocal.
Adaptado de: Fant, n.d.

La Figura 1.8 muestra un modelo por medio de un diagrama de bloques del aparato fonador:

- Fuente: salida de las cuerdas vocales, tren de pulsos de aire periódico que equivale a un tono puro.

- Faringe: conducto que conecta las cuerdas vocales con las cavidades nasal y vocal. En este punto ocurre una conmutación entre la cavidad nasal y la vocal para permitir tanto la respiración como el habla. Adicionalmente, se genera el proceso de resonancia, en el que el tono puro experimenta una no linealidad que crea nuevas componentes en frecuencia. Dado que la

[5] La prostaféresis es el conjunto de identidades trigonométricas que representan los productos de senos y cosenos mediante combinaciones lineales de los mismos.

entrada a este bloque es un tono puro, a la salida se tendrá un conjunto discreto de componentes en frecuencia -armónicos- que son múltiplos enteros de la frecuencia del tono (señal periódica en el tiempo). La apertura de la boca y la cavidad nasal influyen en la resonancia y por tanto determinarán los armónicos generados, resultando en la modulación de la señal de voz, que al cambiar su envolvente se deja de considerar periódica.

- Cavidad vocal: es la encargada de la articulación, la cual también genera nuevas componentes en frecuencia, no obstante, en este caso la entrada está compuesta por más de una componente en frecuencia generando una señal no periódica. La señal no periódica tiene ciertas características relevantes tanto en el dominio del tiempo como en el de la frecuencia. En el dominio del tiempo consiste en una señal periódica multiplicada por una envolvente que determina el inicio y fin del sonido. En el dominio de la frecuencia consiste en componentes de frecuencia agrupadas alrededor de los impulsos que corresponden a los armónicos. Estas agrupaciones se conocen como formantes.

En la Figura 1.9 se muestran con mayor claridad los efectos de la resonancia y la articulación, donde la señal ideal corresponde a aquella resultante de la resonancia, por lo que es una señal periódica con un espectro discreto cuyas componentes son múltiplos enteros del *pitch* (aproximadamente 239 Hz). Por otro lado, la señal de voz real es aquella que experimenta tanto los efectos de la resonancia, como de la articulación, gracias a esto es posible apreciar la envolvente en el dominio del tiempo y los formantes en el dominio de la frecuencia. Es necesario destacar que los picos más altos de cada uno de los formantes corresponden a los armónicos de la señal ideal, esto no significa que la relación de amplitud de los armónicos se mantenga después de la articulación.

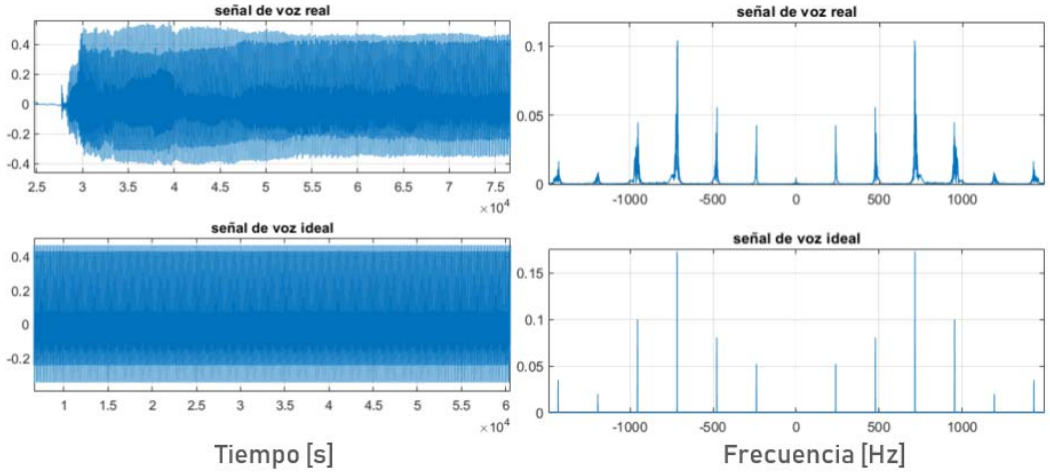


Figura 1.9. Efectos de la resonancia y la articulación.
Elaboración propia.

Los diferentes sonidos son entonces señales con formas diferentes en el dominio del tiempo, cuyas componentes en frecuencia también se diferencian. Dado el carácter cuasi periódico de los sonidos (formantes de la voz), es más simple analizar

las diferencias espectrales. En la Figura 1.10 se muestran los espectros del sonido de las diferentes vocales. En la izquierda se muestran los sonidos de «a», «e» e «i»; cuya principal diferencia al pronunciarse es la apertura de la boca, por lo que los armónicos que se generan cambian de una vocal a la otra. Por otro lado, en la parte derecha se muestran los sonidos correspondientes de «o» y «u», en los cuales la apertura es aproximadamente igual, por lo que se generan los mismos armónicos y la diferencia está dada por la articulación mediante la posición de los labios (modulación), que cambia la relación de amplitud de los formantes.

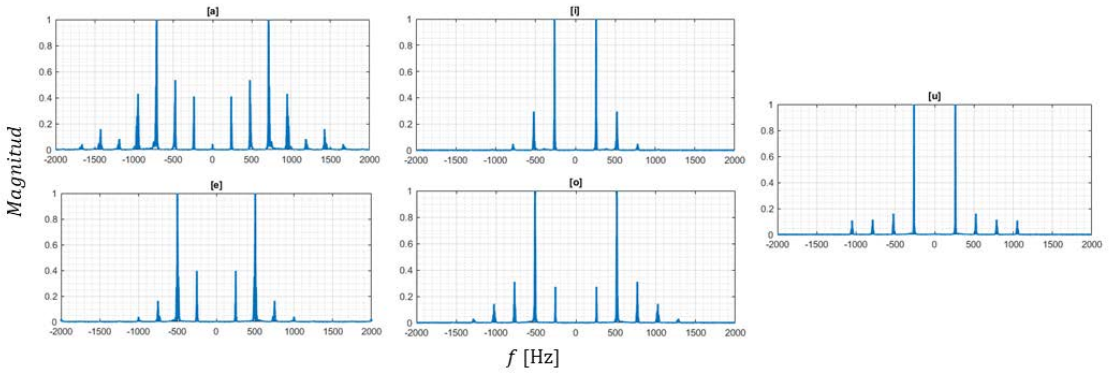


Figura 1.10. Sonidos vocales en el dominio de la frecuencia.
Elaboración propia.

Estos sonidos vocales (ver Figura 1.10) fueron pronunciados por la misma persona, no obstante, se presentan variaciones en el *pitch* cercanas a los $30$ Hz, el cual toma valores entre $237$ y $266$ Hz. Fisiológicamente es posible que el *pitch* cambie debido a variaciones en la tensión de las cuerdas vocales.

Se han llevado a cabo numerosas investigaciones relacionadas con los significados de dichas variaciones en frecuencia. En *Breitenstein et al., Burkhardt y Sendlmeier* y *Lee et al*. muestran que existe una relación entre las emociones y las variaciones del *pitch* (Breitenstein et al., 2001; Burkhardt & Sendlmeier, n.d.; Lee et al., 2005). En *Apple et al.* realizan un sondeo para determinar las opiniones de los diferentes oyentes sobre las emociones de las personas a partir de sus grabaciones de voz, emitiendo juicios sobre su carácter y sus características físicas (Apple et al., 1979). Dada la importancia de la percepción del *pitch,* se ha estudiado el cambio en éste que los padres inducen al momento de hablar con sus hijos - principalmente en sus primeros años de vida - (Broesch & Bryant, 2018), además, dado que existen frecuencias consideradas como más “agradables” que otras, se ha estudiado su relación con la percepción de que una persona como “atractiva”. En *Fraccaro et al.* concluyen que aunque las frecuencias “agradables” no tienen influencia significativa en la percepción de una persona como “atractiva”, las frecuencias “desagradables” sí disminuyen las posibilidades de alcanzar dicho calificativo (Fraccaro et al., 2013). Finalmente, se ha demostrado que existe relación entre ciertas enfermedades y variaciones “no naturales” del *pitch*, como las parálisis cerebrales (Chen et al., 2016) y el Alzheimer (Zhu et al., 2017).

Existen idiomas tonales y no tonales y su pronunciación tiene asociadas variaciones en el *pitch* más o menos pronunciadas (Metze et al., 2013), sin embargo, todos los idiomas usan variaciones en el *pitch*. En el caso de los idiomas tonales dichas variaciones están asociadas con el léxico y para los idiomas no tonales se utilizan para dar énfasis (Bent et al., 2006). Otra relación entre los idiomas y el *pitch* es que, según *Hanjun Liu et al.* existe una correlación entre el idioma y la facilidad con la que se altera el *pitch* frente a estímulos externos. Esta correlación no depende de si el idioma es tonal o no (H. Liu et al., 2010).

Hasta el momento, se ha analizado de forma general el proceso de creación de la voz utilizando como ejemplo los sonidos vocales. Dichos sonidos se diferencian por sus componentes en frecuencia y la relación de amplitudes entre las mismas, aunque las frecuencias de tales componentes no sólo dependen de la persona (características anatómicas) sino también del idioma que dicha persona se encuentre hablando y varían con el tiempo debido a que se alteran frente a diferentes estímulos (como la emoción que se esté experimentando). Por lo anterior, es posible diferenciar entre el habla[6] natural y el habla sintética mediante el análisis de las variaciones del *pitch* (Pal et al., 2018).

La producción del habla, por naturaleza, tiene una esencia dinámica, pues se compone de la combinación de diferentes sonidos y periodos de silencio. La variación de los movimientos articulatorios en el tiempo causa el cambio entre sonidos y hace del habla una forma efectiva para transmitir información (Huffman, 2016).

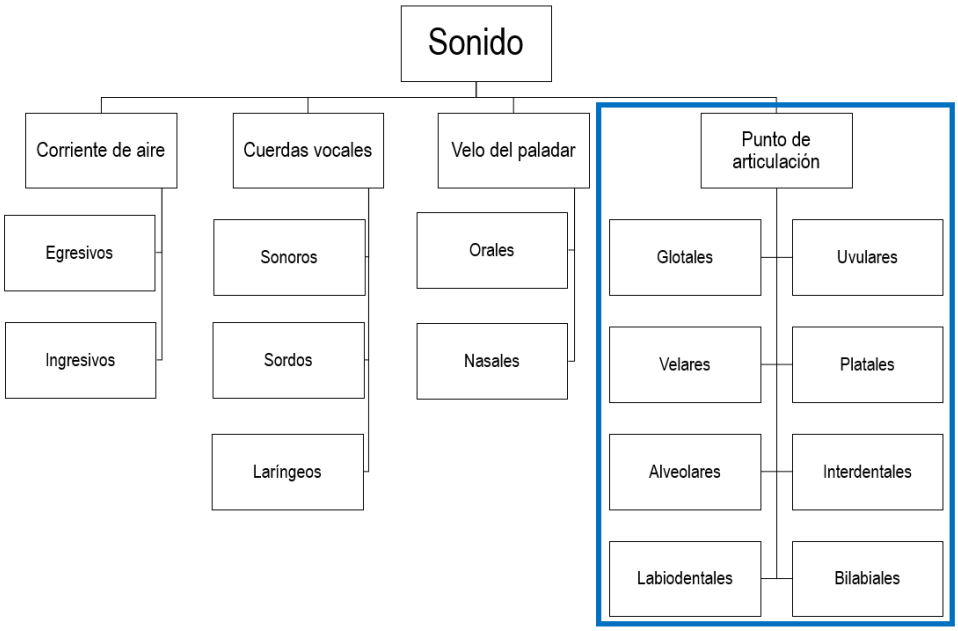


Figura 1.11. Clasificaciones del sonido.
Elaboración propia.

[6] El habla corresponde a la realización de una lengua, es decir, la producción de sonidos regidos por ciertas convenciones gramaticales. La voz describe de forma general los sonidos generados por el aparato fonador de los humanos.

En la Figura 1.11 se muestran algunas de las formas en las que se pueden clasificar los sonidos de la voz humana. La fonética articulatoria estudia la producción del sonido mediante las formas en las que se afecta el flujo del aire que viaja entre los pulmones y la boca.

Dentro de los sonidos que componen el habla se tienen las vocales, las cuales no generan obstrucciones en el flujo del aire, y las consonantes, las cuales generan obstrucciones en el flujo del aire. Los sonidos correspondientes a las consonantes se consiguen mediante la articulación, así que se pueden agrupar mediante el lugar en el que se genera la restricción del flujo del aire; es decir, el punto de articulación, como se muestra en la Figura 1.12 (Szczegielniak, n.d.):

- Glotales: Sonidos producidos al restringir el flujo de aire por la glotis abierta [h] o obstruirlo por completo [ʔ][7].
- Uvulares: Sonidos producidos al subir la parte de atrás de la lengua hacia la úvula [r, q, G].
- Velares: Sonidos producidos al subir la parte trasera de la lengua al velo [k, g, ŋ].
- Palatales: Sonidos producidos al subir la parte delantera de la lengua al paladar [ʃ, ʒ, ʧ, ʤ, j].
- Alveolares: Sonidos producidos al subir la lengua al alveolar rígido. En este caso la lengua se puede subir de diferentes formas
  - Parte delantera de la lengua [t, d, n].
  - Lados de la lengua [s, z].
  - Punta de la lengua [l].
  - Punta de la lengua curva [r].
- Interdentales: Sonidos producidos al ubicar la punta de la lengua entre los dientes [θ, ð].
- Labiodentales: Sonidos producidos al tocar el labio inferior con los dientes superiores [f, v].
- Bilabiales: Sonidos producidos al juntar los labios [p, b, m].

La Figura 1.12 muestra diferentes formas con las que se puede hacer la articulación y que permite realizar una clasificación de las consonantes, no obstante, la producción de los sonidos no depende únicamente de un parámetro, como el punto de articulación, sino que es el resultado de muchos parámetros (Gick et al., n.d.), algunos de los cuales se muestran en la Figura 1.13. Al analizar los sonidos asociados a las vocales y las consonantes se observa que las consonantes tienen un efecto envolvente sobre los sonidos de las vocales (más armónicos) (Woods et al., n.d.), así desde el punto de vista de la fonética articulatoria, es posible diferenciar estos sonidos como se muestra en la Tabla 1.1.

[7] Los símbolos ʔ, ŋ, ʃ, ʒ, ʧ, ʤ, j, θ, ð representan diferentes sonidos fonéticos.

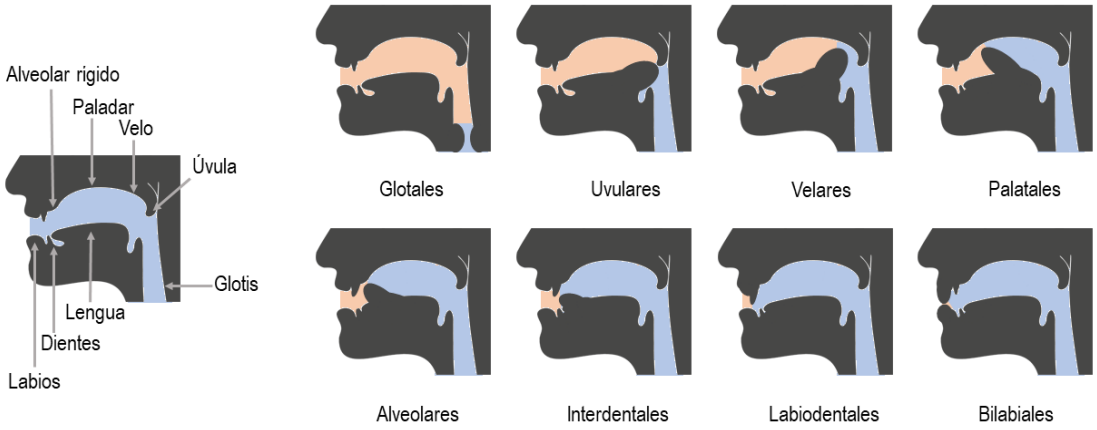


Figura 1.12. Punto de articulación.
Elaboración propia.

Tabla 1.1. Fonética articulatoria de las vocales y consonantes.
Adaptado de: *Articulatory Phonetics Speech Sound Form*, n.d.

| Vocales | Consonantes |
|---|---|
| Restricción del tracto vocal sin relevancia. | Restricción del tracto vocal significativa. |
| Sonidos abiertos. | Sonidos constrictivos. |
| Sonoros. | Sonoros o sordos. |
| Acústicamente más intensos. | Acústicamente menos intensos. |

Con el fin de analizar espectralmente cuáles son los efectos de la conjunción de vocales y consonantes se utilizarán palabras que utilicen las mismas vocales, pero en las que cambien sus consonantes. La pronunciación de una vocal, al ser un sonido sonoro y abierto (se debe a la vibración de las cuerdas vocales y no restringe el paso del flujo de aire) genera un espectro aproximadamente discreto (formantes de la voz), por otro lado, las consonantes son sonidos constrictivos (restringen el paso del flujo de aire por medio de la articulación), los cuales no generan armónicos en el dominio de la frecuencia, sino componentes intermedias. De esta forma, se tendrán señales periódicas moduladas por una envolvente que cambia lentamente.

Con el propósito de mostrar lo anteriormente mencionado, se realizó un experimento en el cual se grabaron las señales de una persona diciendo las palabras: ‘cama’, ‘cata’, ‘mama’, ‘mata’, ‘capa’, ‘gata’, ‘mapa’ y ‘rata’. En la Figura 1.13 se muestran los resultados obtenidos mediante tres columnas, donde la primera columna representa la grabación total de la palabra y muestra que las ocho (8) palabras tienen envolventes diferentes y que dichas envolventes dependen de las consonantes puesto que, por ejemplo, el inicio de las señales correspondientes a ‘cama’ y ‘cata’ es muy similar. La segunda columna corresponde a un tiempo de observación menor, para mostrar que la presencia de una señal periódica de mayor frecuencia se ve modificada en amplitud por la envolvente de baja frecuencia. Finalmente, la tercera columna muestra los componentes espectrales de cada una de las señales de voz, donde es posible identificar los formantes de la voz y además

apreciar cambios en las frecuencias intermedias y en las relaciones de amplitud de dichos formantes.

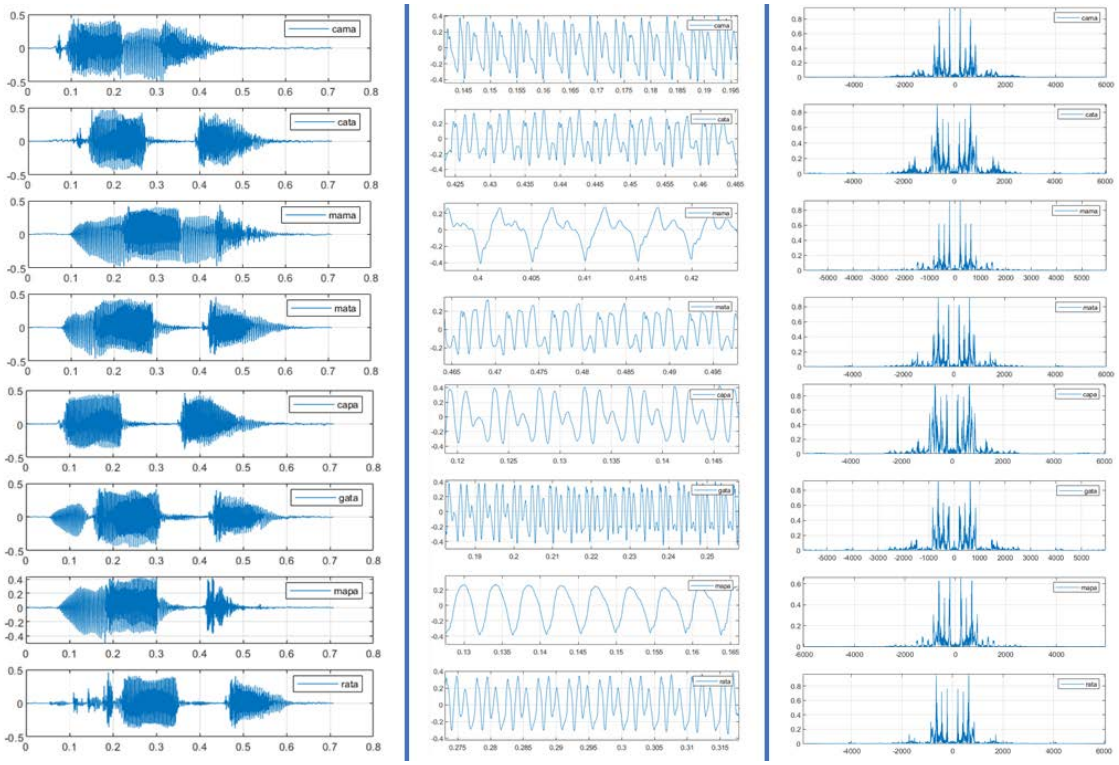


Figura 1.13. Fonética articulatoria de las vocales y consonantes.
Elaboración propia.

### 1.2.3. Comportamiento Estadístico de las Señales de Voz

Las señales de voz tienen una naturaleza aleatoria con muchos grados de libertad[8], lo cual hace que sea extremadamente complejo encontrar un modelo que abarque su comportamiento en general, sin embargo, a pequeña escala (cortos instantes de tiempo) es posible asumir que sus valores de amplitud se rigen bajo una distribución de Laplace, no obstante, los parámetros de la distribución varían entre las diferentes realizaciones de la señal de voz. Matemáticamente, la frecuencia relativa de aparición de una amplitud $x$ está dada por la siguiente ecuación:

$$f_X(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}, \tag{1.3}$$

donde $X$ es una variable aleatoria continua que representa la amplitud, $x$ es una realización de la variable aleatoria, $\mu$ es la media y $b$ la dispersión.

[8] Se entiende como grados de libertad a las variables de carácter aleatorio que generan incertidumbre sobre el comportamiento de la señal. Una señal de voz depende, por ejemplo, del tono de voz, la emoción, la velocidad y el acento de una persona.

En la Figura 1.14 se muestran las Funciones de Densidad de Probabilidad (pdf, *Probability Density Function*) de 3 señales aleatorias con distribución Laplaciana, en las cuales varían sus valores de $\mu$ y $b$.

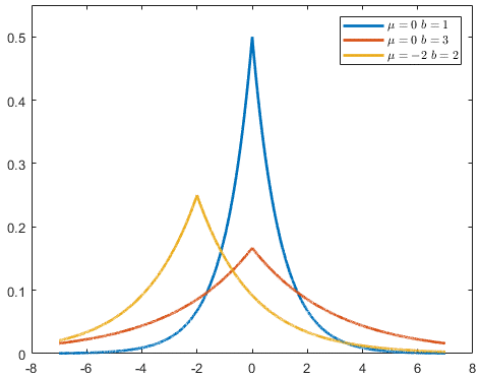


Figura 1.14. Distribución Laplaciana.
Elaboración propia.

Con las señales de voz de corta duración, del orden de los milisegundos, se trabaja bajo la suposición de que el comportamiento de la señal de voz es ergódico[9], al menos en primer orden. En la Figura 1.15 se muestran los histogramas de los valores de amplitud de dos tramas consecutivas de una señal de voz con duración de cinco milisegundos cada una, con los cuales es posible evidenciar la variabilidad de este tipo de señales, ya que cambian los dos parámetros de la distribución, i.e., la media y la dispersión.

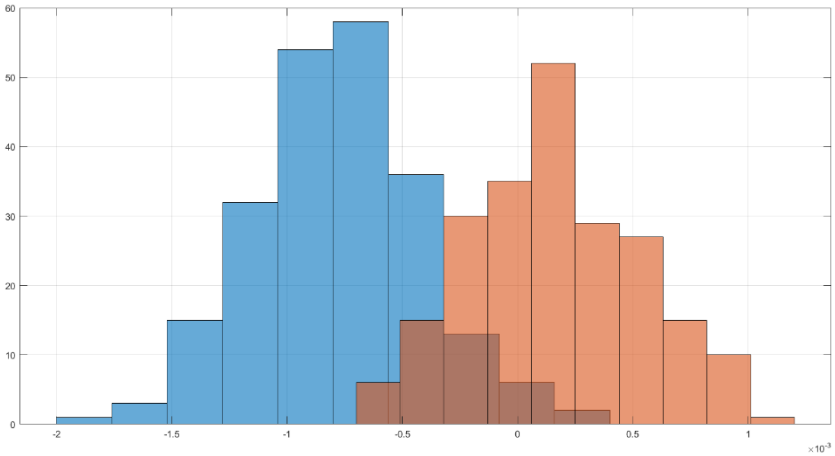


Figura 1.15. Distribución de amplitudes de señales de voz consecutivas.
Elaboración propia.

Por lo anterior, para el procesamiento de señales de voz se asume que la amplitud de éstas sigue una distribución de amplitud Laplaciana. Algunas de las técnicas de procesamiento en las que más se utiliza esta suposición son: el mejoramiento (Jukic & Doclo, 2014; Mahmmod et al., 2017; Rashidi-nejad & Abutalebi, 2012), la

[9] Un comportamiento ergódico indica que los parámetros estadísticos del proceso estocástico se pueden determinar a partir de una única realización.

clasificación (Cai et al., 2007; Kaya & Arioz, 2015; Tinati & Mozaffary, 2006) y marcas de agua (Akhaee et al., 2009). No obstante en *Gazor y Zhang* y *Kokkinakis y Nandi* se realizan pruebas de hipótesis para determinar cuál es la pdf que mejor modela la amplitud de las señales de voz y encontraron que esto depende de la duración de la señal, puesto que a mayor duración se tendrá una mayor proporción de periodos de silencio y en éstos se verá el ruido de fondo que se modela mediante una distribución Gaussiana (Gazor & Zhang, 2003; Kokkinakis & Nandi, 2005).

En la Figura 1.16 se muestran los histogramas, con igual número de barras, de los valores de amplitud para señales de voz de diferente duración y con diferentes frecuencias de muestreo. El caso en el que menos se aproxima a una distribución Laplaciana es aquel con una duración aproximada de $40\ s$ y una frecuencia de muestreo de $44100\ \text{Hz}$ (esquina inferior izquierda).

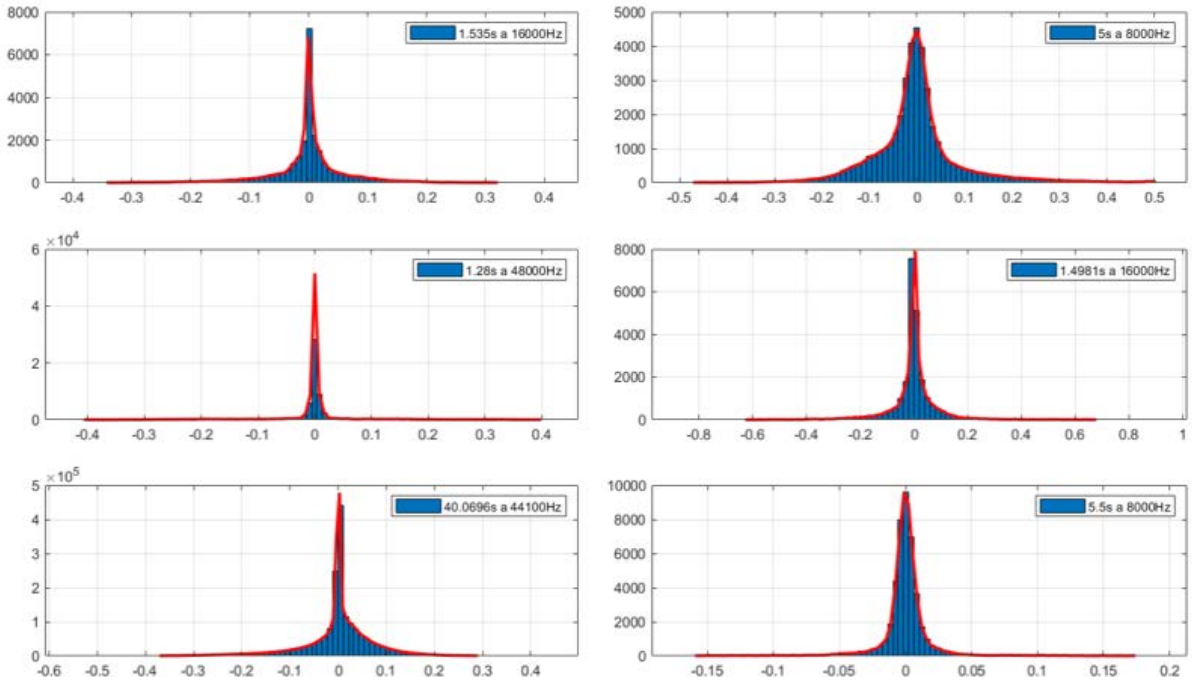


Figura 1.16. Histogramas de los valores de amplitud de la voz.
Elaboración propia.

### 1.2.4. Componentes Espectrales de las Señales de Voz

La versión más aceptada sobre los límites de la audición humana establece que éstos se encuentran entre los $20\ \text{Hz}$ y los $20000\ \text{Hz}$, no obstante es algo sobre lo que no hay un consenso, puesto que se tienen estudios que definen el límite superior en $12000\ \text{Hz}$ (Pumphrey, 1950) y otros que defienden que las frecuencias de ultrasonido (superiores a $20\ \text{KHz}$) estimulan la cóclea y generan resonancia en el cerebro, por lo que deben considerarse dentro del rango de audición (M. Lenhardt et al., 1991; M. L. Lenhardt, 2003).

En la Figura 1.17 se muestran los diferentes anchos de banda considerados en el procesamiento de señales de voz. Los estudios que involucran la opinión de diferentes personas muestran que, entre mayor sea el ancho de banda considerado, mayor será la calidad percibida de la voz (Hanzo et al., 2007); sin embargo, no es uniforme el aporte que realizan las subbandas[10] de frecuencias a la inteligibilidad de la señal de voz (Letowski & Scharine, 2017).

El ancho de banda con el que operan los codificadores de voz (*vocoders*) depende de la aplicación, aunque no es la única diferencia entre éstos, ya que existen diferentes formas a partir de las cuales se puede generar una secuencia digital con información semejante a la de la señal de voz.

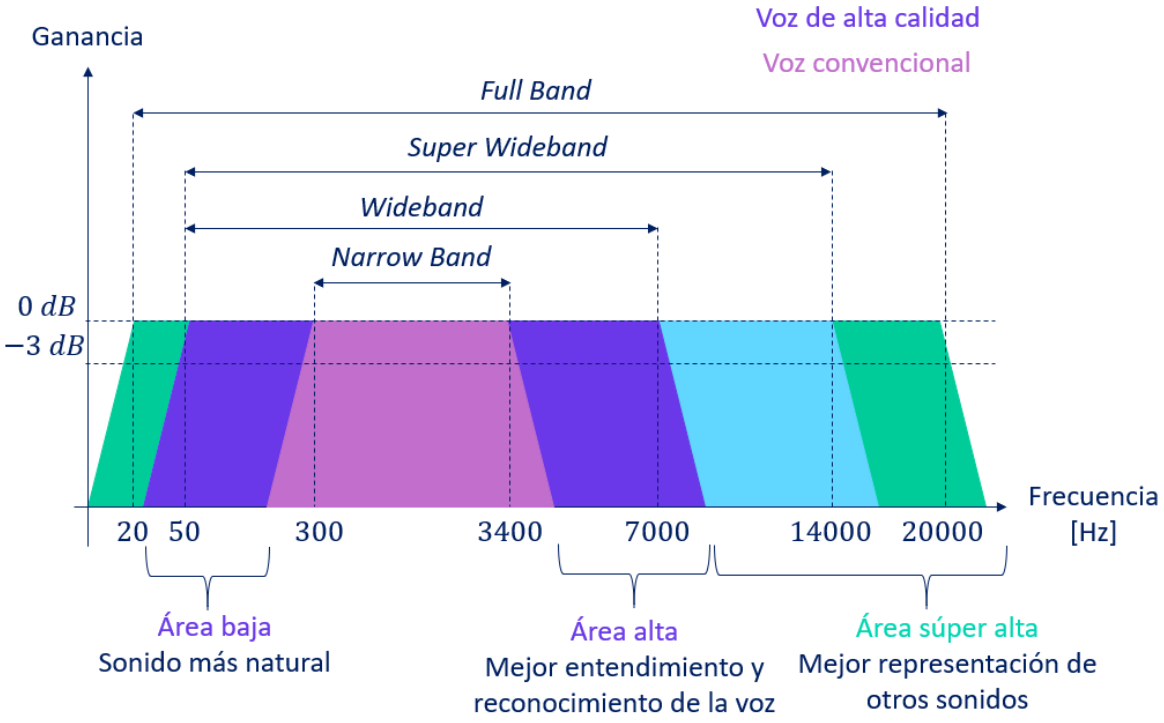


Figura 1.17. Bandas de frecuencia de la voz.
Elaboración propia.

## 1.3. CODIFICADORES DE VOZ

Los *vocoders* ejecutan los tres procesos relacionados con la conversión analógica-digital: muestreo, cuantificación y codificación de fuente. Los *vocoders* se pueden clasificar como *vocoders* de: forma de onda; subbanda y transformaciones; y de análisis y síntesis (Ogunfunmi et al., 2015). La cuantificación de las señales de voz a partir de su representación *wavelet* corresponde a la categoría de *vocoders* de subbanda y transformaciones.

[10] En este trabajo de grado se entiende por subbanda a una fracción de las componentes de frecuencia.

Los primeros experimentos relacionados con la digitalización de la voz fueron efectuados en el siglo XX, motivados por la necesidad de transmitir mensajes de voz de manera inteligible a través de la red telefónica existente (Ogunfunmi et al., 2015).

Los primeros *vocoders* fueron los de análisis y síntesis (Dudley, 1940), los cuales fueron inicialmente olvidados debido a que la voz resultante carecía de naturalidad, sin embargo, posteriormente resurgieron debido a que nuevos logros tecnológicos permitieron mejorar los resultados obtenidos. Las primeras versiones comerciales fueron *vocoders* de forma de onda. Finalmente, se implementaron los *vocoders* de subbanda y transformación, enfoques que habían sido utilizados con antelación en el procesamiento de audio e imágenes.

Los organismos de estandarización como la Asociación de la Industria de las Telecomunicaciones y la Alianza de la Industria Electrónica (TIA-EIA, *Telecommunications Industry Association - Electronic Industries Alliance*), el Instituto Europeo de Estándares de Telecomunicaciones (ETSI, *European Telecommunications Standards Institute*), y la Unión Internacional de Telecomunicaciones sector Telecomunicaciones (ITU-T, *International Telecommunications Union - Telecommunications*), han especificado diferentes *vocoders* que buscan cumplir los requerimientos de los nuevos servicios, no obstante, existen actualmente en el mercado *vocoders* que no se encuentran avalados por ningún organismo de estandarización (Ogunfunmi et al., 2015).

En la Figura 1.18 se muestra una línea de tiempo con algunos de los vocoders más conocidos[11], representados en azul los *vocoders* de forma de onda, en rojo los de análisis y síntesis, y en verde los de subbanda y transformación. Los casos en los que el mismo estándar ésta representado en dos colores se deben a que en dichos estándares se contemplan dos tipos diferentes de *vocoders*.

El *vocoder* de forma de onda G.726 se basa en la Modulación por Codificación de Pulsos Diferencial y Adaptativa (ADPCM, *Adaptive Differential Pulse Code Modulation*), funciona a una tasa de 32 Kbps y es utilizado en los teléfonos inalámbricos (ITU, 1990), por otra parte, dentro de los estándares para telefonía móvil celular se encuentra el *vocoder* del Sistema Global para las Comunicaciones Móviles con Velocidad Máxima Mejorada (GSM-EFR, *Global System for Mobile communications - Enhanced Full Rate*), el cual es un *vocoder* de análisis y síntesis que se basa en Predicción Lineal Excitada por Códigos Algebraicos (ACELP, *Algebraic Code-Excited Linear Prediction*) y opera a una tasa de 12.2 Kbps (Hanzo et al., 2001). Finalmente, dentro de los *vocoders* de subbanda y transformación se encuentra el G.719 que puede basar su funcionamiento en la Transformada Discreta de Coseno Modificada (MDCT, *Modified Discrete Cosine Transform*) y operar a tasas entre 32 y 128 Kbps (W. Jiang et al., 2013).

[11] Las fechas mostradas en la Figura 2 corresponden a la primera versión de los estándares que cobijan a cada uno de los *vocoders* mencionados.

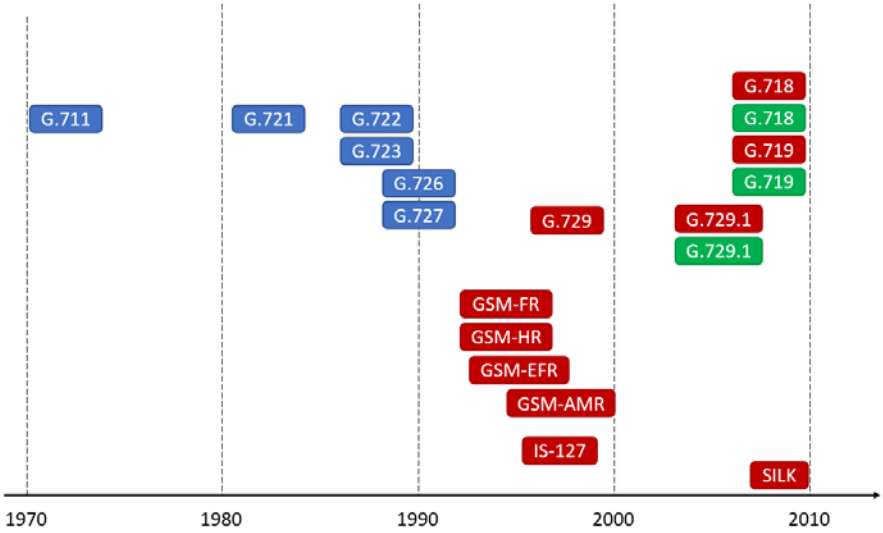


Figura 1.18. Evolución de los codificadores de voz.
Elaboración propia.

Cada uno de los métodos empleados en los *vocoders* tiene como objetivo encontrar un compromiso entre la calidad de la voz resultante y su compresión. Los métodos basados en la forma de onda o el análisis y síntesis permiten alcanzar grandes tasas de compresión sacrificando la calidad de la voz resultante, por lo cual, se ha dado paso a métodos de mayor costo computacional y menor eficiencia en cuanto a la tasa de compresión, pero con los que se mejora la calidad.

# CAPÍTULO 2: CONVERSIÓN ANALÓGICA DIGITAL

La codificación de fuente busca la representación más eficiente de la señal sin afectar su calidad, i.e., la representación más corta de la señal manteniendo gran parte de su información.

La Conversión Analógica-Digital (ADC, *Analog to Digital Conversion*) tiene como objetivo pasar el conjunto denso de amplitudes de la señal analógica al número finito de valores (en tiempo y amplitud) que mejor la representen. Esta es la razón por la cual el proceso de conversión analógica digital se puede entender como una forma de codificación de fuente para señales analógicas (Gallager, 2008).

El proceso de conversión analógica digital consta de tres etapas, a saber: el muestreo, en el que se pasa de una señal de tiempo continuo $u(t)$ a una secuencia de muestras (tiempo discreto) $u(nT_s)$; la cuantificación, en el cual se discretizan los valores de amplitud y se tiene una secuencia $v(nT_s)$ discreta en tiempo y amplitud; y finalmente la codificación de fuente discreta, que consiste en representar dichos valores discretos de amplitud a través de una secuencia de dígitos binarios $\{b_i\}$ (ver Figura 2.1).

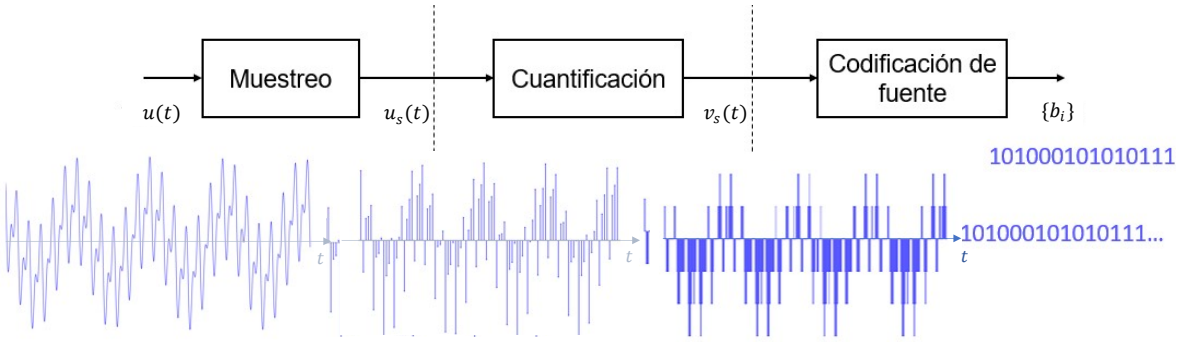


Figura 2.1. Conversión analógica digital.
Elaboración propia.

## 2.1. MUESTREO

El proceso de convertir una señal de tiempo continuo en una secuencia de muestras debe garantizar la conservación de la información. En 1928 Nyquist enunció el teorema de muestreo (Nyquist, 1928), en el que se establecieron las condiciones para representar fielmente una señal de tiempo continuo a través de sus muestras (Ramírez Viáfara et al., 2020).

Sea $u(t)$ una señal de tiempo continúo limitada en banda a $W$ Hz[12], el teorema de muestreo establece que, para su fiel representación a través de una secuencia de muestras, la versión muestreada de la señal, $u_s(t)$, debe ser una secuencia de

[12] Las señales susceptibles a un proceso de muestreo deben ser limitadas en banda, i.e., para frecuencias superiores a una frecuencia $W$ sus componentes espectrales deben ser nulas o cercanas a cero.

muestras impulsivas tomadas cada $T_s$ segundos (ver Figura 2.2), donde $T_s = {}^{1}/_{F_s}$ es el periodo de muestreo y $F_s$ la frecuencia de muestreo, la cual debe cumplir con $F_s \geq 2W$. En (2.1) se observa la expresión de la señal muestreada.

$$u_s(t) = u(t)\sum_{n=-\infty}^{\infty}\delta(t - nT_s) = \sum_{n=-\infty}^{\infty}u(nT_s)\delta(t - nT_s). \tag{2.1}$$

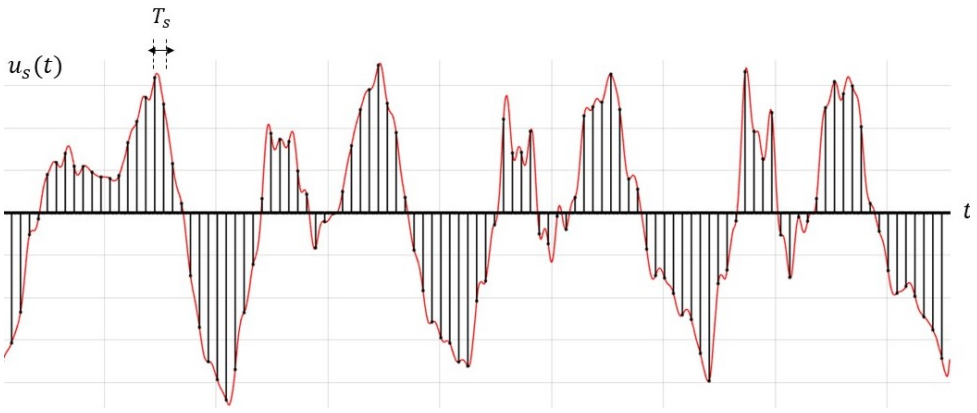


Figura 2.2. Señal muestreada en el dominio del tiempo.
Elaboración propia.

Del análisis de Fourier se sabe que: 1) La Transformada de Fourier (FT, *Fourier Transform*) de un tren periódico de impulsos es otro tren periódico de impulsos. 2) La multiplicación de dos señales en el dominio del tiempo equivale a la convolución de sus respectivos espectros en el dominio de la frecuencia. En el dominio de la frecuencia se tienen entonces versiones desplazadas (desfazadas) de la señal en múltiplos enteros de $F_s$ (2.2). Para que no exista traslape entre las réplicas espectrales, y de esta manera evitar el *aliasing*, se debe garantizar que $F_s \geq 2W$.

$$\tilde{u}_s(f) = \tilde{u}(f) * \left[F_s\sum_{l=-\infty}^{\infty}\delta(f - lF_s)\right] = F_s\sum_{l=-\infty}^{\infty}\tilde{u}(f - lF_s), \tag{2.2}$$

donde $\tilde{u}_s(f)$ representa la transformada de Fourier de $u_s(t)$, es decir, $\mathcal{F}\{u_s(t)\}$.

En la Figura 2.3 se muestran los efectos del valor de $F_s$ sobre el espectro de la señal muestreada. Las réplicas del espectro de la señal que se generan al emplear $F_s = 2W$ no se traslapan, por lo que no existe alteración en su forma y se conserva la información de ésta; no obstante, este caso requiere del uso de filtros ideales (sin bandas de transición) para la reconstrucción de $u(t)$. En la práctica, se muestrea por encima de este límite inferior, esto es $F_s > 2W$, con el fin de lograr reconstruir correctamente la señal original a partir de sus muestras, utilizando filtros reales. Los valores de $F_s < 2W$ generan *aliasing*, i.e., traslape entre las réplicas espectrales,

provocando que el espectro de la señal original se distorsione e impidiendo su recuperación.

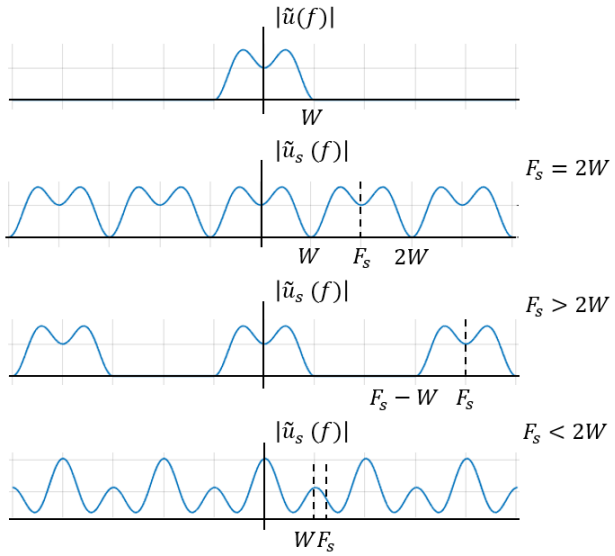


Figura 2.3. Teorema de muestreo en el dominio de la frecuencia.
Elaboración propia.

La señal recuperada, $u'(t)$, es el resultado de la interpolación de sus muestras utilizando un filtro pasa bajas que elimina las réplicas en frecuencia que no se encuentran alrededor del origen. En (2.3) se plantea la forma teórica por medio de la cual es posible realizar la reconstrucción de una señal a partir de sus muestras (Shannon, 1948a).

$$u'(t) = \sum_{n=-\infty}^{\infty} u(nT_s)\mathrm{sinc}\left(\frac{t - nT_s}{T_s}\right). \tag{2.3}$$

Cuando se muestrea por encima del límite de Nyquist ($F_s = 2w$), el proceso se denomina sobremuestreo. El sobremuestreo, además de permitir la reconstrucción de la señal con filtros reales, dota de robustez a la señal muestreada ante muestras perdidas o una posible distorsión, puesto que las muestras están más correlacionadas y por tanto comparten información (Eldar, 2014).

Suponiendo que $T_s = \frac{1}{2W}$, el sobremuestreo se puede expresar como $T_s' = \frac{T_s}{k}$ donde $k \epsilon \mathbb{N}, k > 1$, de esta forma, la señal muestreada queda como:

$$u_s(t) = \sum_{n=-\infty}^{\infty} u\left(n\frac{T_s}{k}\right)\delta\left(t - n\frac{T_s}{k}\right). \tag{2.4}$$

Sin importar cuál sea el valor del periodo de muestreo, éste indica que se toma una muestra cada $T_s/k$ segundos; sin embargo, el intervalo de tiempo dado por $T_s$ admite un desfase, $t_o$, siempre y cuando dicho desfase satisfaga $0 \leq t_o < T_s$, así:

$$u_s(t) = \sum_{n=-\infty}^{\infty} u\left(n\frac{T_s}{k} + t_o\right)\delta\left(t - n\frac{T_s}{k} - t_o\right). \quad (2.5)$$

La expresión matemática (2.5) se puede expresar como se muestra en (2.6), lo cual se muestra gráficamente en la Figura 2.4, para la cual se toma un valor de $k = 4$.

$$u_s(t) = \sum_{n=-\infty}^{\infty}\sum_{m=0}^{k-1} u\left(nT_s + \frac{mT_s}{k}\right)\delta\left(t - nT_s - \frac{mT_s}{k}\right),$$

$$= \sum_{j=0}^{k-1} u_s{}^{(j)}(t)\,. \quad (2.6)$$

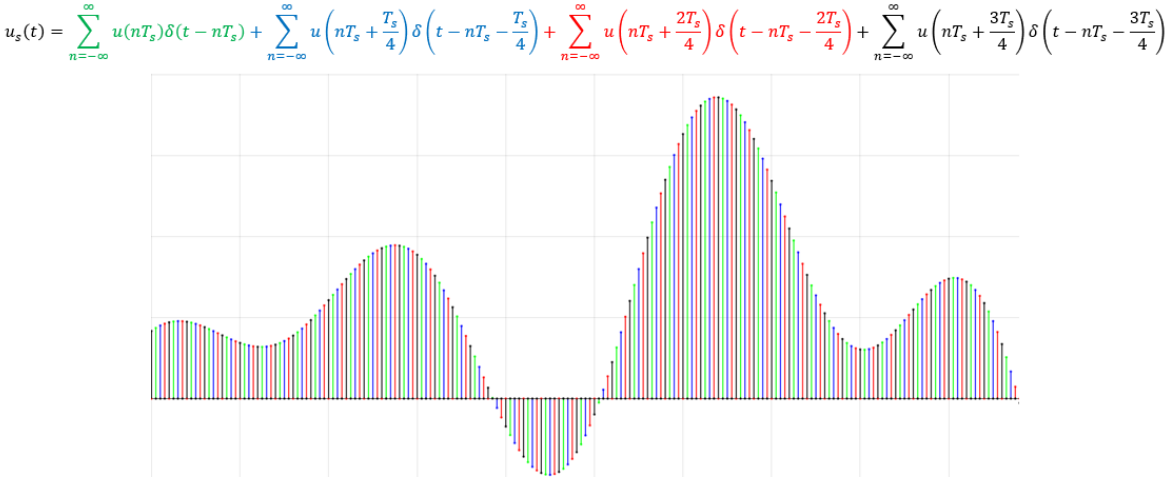


Figura 2.4. Sobremuestreo como una sumatoria de señales muestreadas con corrimientos en el tiempo. Elaboración propia.

El filtro interpolador utilizado para recuperar únicamente la réplica del espectro de la señal centrada en el origen es un sistema Lineal e Invariante en el Tiempo (LTI), por lo que, si a la entrada se tiene una sumatoria de señales, a la salida se tendrá la sumatoria de las salidas individuales afectadas por constantes de proporcionalidad y cambios de fase (2.7).

$$u'(t) = \sum_{n=-\infty}^{\infty}\sum_{j=0}^{k-1} u\left(nT_s + \frac{jT_s}{k}\right)\mathrm{sinc}\left(\frac{t - nT_s - \frac{jT_s}{k}}{T_s}\right). \quad (2.7)$$

En la Figura 2.5 se muestra en azul la reconstrucción de las $k$ señales de manera individual y en violeta la sumatoria de dichas $k$ señales. En este caso, como cada versión desplazada $u_s{}^{(j)}(t)$ cumple con el teorema de muestreo y se está utilizando el filtro interpolador correcto, se tiene que aproximadamente $u'(t) = u(t)$. Este enfoque permite ver que sobremuestrear una señal a $F_s' = kF_s$, donde $F_s = 2W$, es equivalente a tener $k$ señales muestreadas a $F_s$.

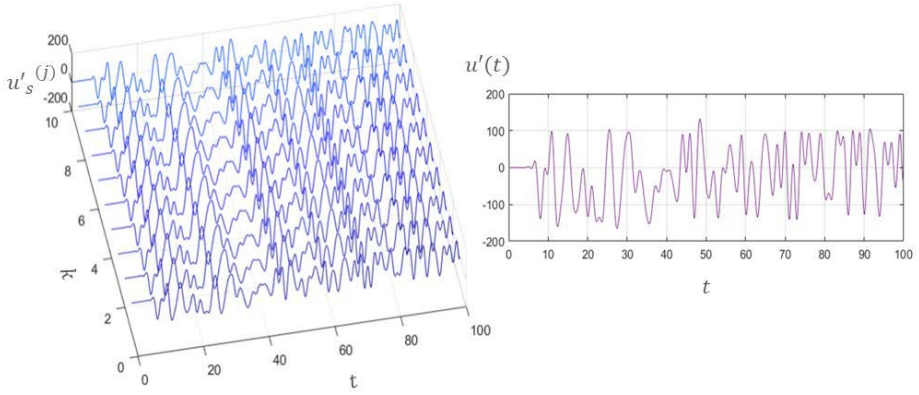


Figura 2.5. Interpolación de las señales muestreadas.
Elaboración propia.

## 2.2. CUANTIFICACIÓN

Cuando se emplea la cuantificación, como parte del proceso de conversión analógica digital, se realiza una transformación de variables aleatorias, donde la entrada del cuantificador es una secuencia de variables aleatorias continuas $U$, que al pasar por el cuantificador se convierten en variables aleatorias discretas $V$. En síntesis, el cuantificador limita un número infinito de valores a un conjunto finito conocido como **alfabeto de cuantificación**. Los elementos de dicho conjunto son los **niveles de cuantificación,** denotados con $a_i$.

Un cuantificador se describe por medio de su característica de transferencia (relación entre la entrada y la salida), porque no es un sistema lineal, en ésta se especifican las **regiones de cuantificación** $R_i$ (intervalos de amplitud sobre los cuales todos sus valores se clasifican como un nivel de cuantificación particular), como se muestra en la Figura 2.6. Las regiones de cuantificación deben cubrir perfectamente todo el rango de valores de amplitud de la señal original.

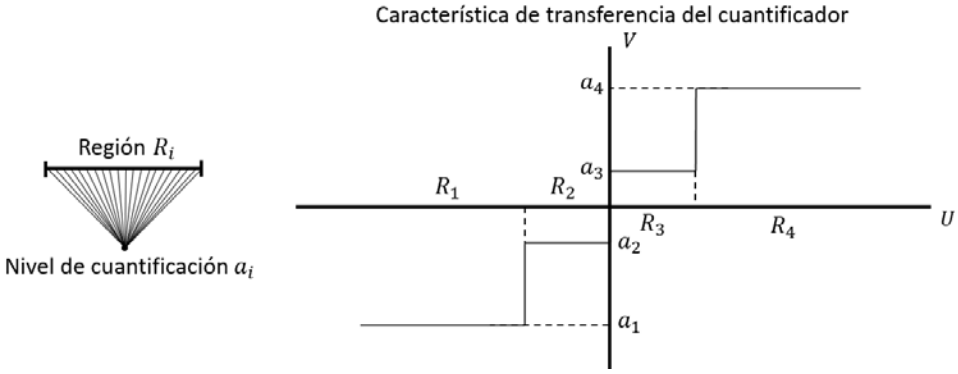


Figura 2.6. Característica de transferencia.
Elaboración propia.

En la Figura 2.6 se muestra la característica de transferencia de un cuantificador escalar, es decir, un cuantificador que discretiza los valores de amplitud de la señal de entrada de forma individual; no obstante, ésta no es la única forma en la que se

puede abordar el proceso de cuantificación. La Figura 2.7 muestra una forma de clasificar los cuantificadores:

- Según su dimensión o número de muestras, el cuantificador puede discretizar las amplitudes de las muestras de forma individual (escalar) o de forma conjunta por medio de *n*-tuplas de muestras (vectorial).
- Según el tamaño de sus regiones, el cuantificador debe crear una partición de la recta real[13], para esto puede optar por divisiones del mismo tamaño (uniforme) o por divisiones con tamaños diferentes (no uniforme).
- Según la representación de las muestras, se tienen muestras de la señal en el dominio del tiempo, las cuales se pueden cuantificar directamente (directa), o se pueden someter a un proceso de transformación (indirecta). La transformación de las muestras puede ser una modificación de su escala por medio de una función del dominio temporal o en otro dominio (transformadas).

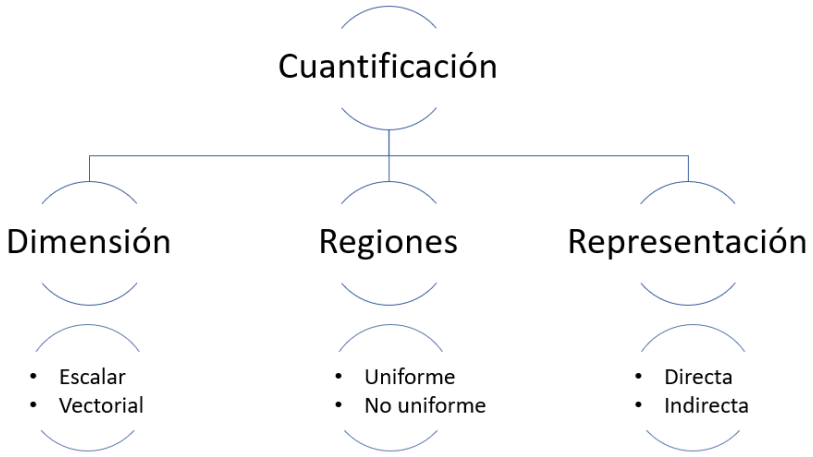


Figura 2.7. Tipos de cuantificadores.
Elaboración propia.

El análisis del proceso de cuantificación, desde la perspectiva de la teoría de la información, es muy útil para entender de forma intuitiva el proceso, por lo que en adelante se recurrirán sus definiciones para la fundamentación del proceso de cuantificación.

El proceso de cuantificación implica pérdida de información, por tanto, el cuantificador debe diseñarse buscando minimizar dicha pérdida. Desde la perspectiva de la teoría de la información, la pérdida de información está relacionada con la información mutua entre la entrada y la salida del cuantificador, la cual está dada por:

$$I(U;V) = H(V) - H(V|U),$$

[13] Conjunto de todos los posibles valores de amplitud.

donde $I(U;V)$ es la información mutua entre las variables aleatorias $U$ y $V$, $H(V)$ representa la entropía de $V$, entendida como la información promedio que aporta $V$ y $H(V|U)$ es la entropía condicional de $V$ dado $U$, la cual representa la incertidumbre que surge sobre $V$ al observar $U$, y en este caso, el conocimiento de $U$ implica el conocimiento de $V$, por lo que dicha incertidumbre no existe, haciendo que $H(V|U) = 0$, de ese modo:

$$I(U;V) = H(V). \tag{2.8}$$

En (2.9) se define la información promedio de la salida en función de la entropía diferencial de la entrada, $h(U)$, y la distorsión que introduce el cuantificador $D$.

$$I(U;V) = H(V) = h(U) - h(U|V)$$

$$= h(U) - D, \tag{2.9}$$

donde $D = h(U|V)$, i.e., la distorsión es la entropía diferencial condicional de $U$ dado $V$.

La entropía diferencial de una variable aleatoria continua se define matemáticamente, como se muestra a continuación:

$$h(U) = \mathrm{E}\left[-\log_2\left(f_U(U)\right)\right],$$

$$= \int_{-\infty}^{\infty} f_U(u) \log_2\left(\frac{1}{f_U(u)}\right) du, \tag{2.10}$$

donde $f_U(u)$ es la pdf de $U$.

No obstante, esta definición por si sola carece de sentido, puesto que las probabilidades puntuales de una variable aleatoria continua son cero, haciendo la información asociada a este tipo de variables infinita. Matemáticamente, los problemas de la entropía diferencial se ven reflejados en que ésta toma valores tanto negativos como positivos y que su valor no es independiente de la escala, puesto que el resultado de (2.10) no es adimensional (Marsh, 2013).

La distorsión $D$, que reduce la información mutua entre la entrada y la salida del cuantificador, depende del diseño del cuantificador. El caso ideal es aquel en el que se dispone de un número de niveles de cuantificación, $N$, muy grande.

### 2.2.1. Cuantificación Ideal

Robert Gallager realizó el planteamiento matemático para el escenario en el que se tiene un número de niveles de cuantificación, $N$, muy grande (Gallager, 2008). Bajo dicha suposición, sin importar la forma de la pdf de $U$, se tiene que sobre cada región

de cuantificación la pdf es aproximadamente uniforme, como se presenta en la Figura 2.8.

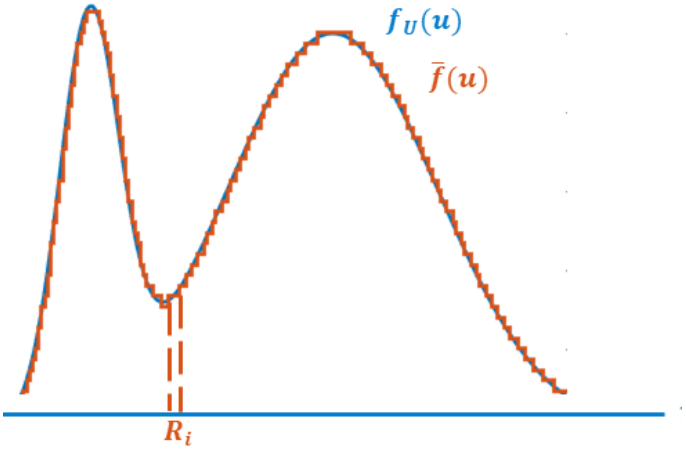


Figura 2.8. pdf aproximada sobre una región.
Elaboración propia.

Bajo el supuesto de una pdf aproximadamente uniforme, la información promedio a la salida del cuantificador, $H(V)$, se puede representar en función de la longitud del intervalo. La información mutua entre la entrada y la salida se define como:

$$I(U;V) = h(U) - h(U|V),$$

pero para $R_i$ se tiene que

$$U|V = a_i \sim \mathcal{U}\left[a_i - \Delta/2, a_i + \Delta/2\right],$$

luego

$$h(U|V = a_i) = \log_2 \Delta,$$

así

$$h(U|V) = \sum_{i=1}^{N} p_i h(U|V = a_i),$$

$$= \log_2 \Delta \sum_{i=1}^{N} p_i = \log_2 \Delta,$$

Finalmente,

$$I(U;V) = h(U) - \log_2 \Delta\ . \qquad (2.11)$$

La expresión matemática (2.11) se puede generalizar para un cuantificador vectorial con un número $d$ de dimensiones, donde las regiones $R_i$ pasan de ser intervalos a particiones del espacio $\mathbb{R}^d$. De esta forma, se utilizará su volumen $\mathcal{V}$ como la medida geométrica del espacio, así:

$$p_i = \int \int_{R_i} \cdots \int f_{\boldsymbol{U}}(\boldsymbol{u}) du \approx f_{\boldsymbol{U}}(a_i)\mathcal{V}(R_i), \qquad (2.12)$$

donde $f_{\boldsymbol{U}}(\boldsymbol{u})$ es la pdf conjunta del vector aleatorio $\boldsymbol{U} = [U_1, U_2, \ldots, U_d]$.

La entropía del vector aleatorio $\boldsymbol{V}$ está acotada por la expresión que se presenta a continuación:

$$H(\boldsymbol{V}) = H(V_1, V_2, \dots, V_n) \leq \sum_{i=1}^{d} H(V_i), \tag{2.13}$$

donde $V_1, V_2, \dots, V_d$ son las componentes de $\boldsymbol{V}$. La igualdad se cumple si las componentes de $\boldsymbol{V}$ son independientes e idénticamente distribuidas (iid).

El resultado de (2.13) muestra que, para un número de niveles de cuantificación elevado, la distorsión que introduce el cuantificador depende del tamaño de la región de cuantificación. Se considera este caso como el de *cuantificación ideal* puesto que, la distorsión que introduce el cuantificador se hace despreciable ante un número muy grande de niveles de cuantificación.

Según el teorema de codificación de fuente de Shannon, se necesitarán en promedio $H(\boldsymbol{V})$ bits para representar cada nivel de cuantificación (Shannon, 1948b). Dado que conforme aumenta el número de niveles de cuantificación, la distorsión disminuye, el costo a pagar es que se necesita un mayor número de bits para representar la secuencia de valores discretizados de la señal. En general, el número de niveles de cuantificación es finito y es un limitante en el diseño del cuantificador.

En la realidad, al tener un número de niveles de cuantificación finito, el cuantificador introduce distorsión significativa, la cual se debe medir de forma práctica. Es así como en la mayoría de los casos se utiliza el Error Cuadrático Medio (MSE, *Mean Square Error*) para calcular dicha distorsión. Matemáticamente el MSE se define como:

$$\begin{aligned} MSE &= \mathrm{E}\left[\big(U - V(U)\big)^2\right], \\ &= \int_{-\infty}^{\infty} \big(u - v(u)\big)^2 f_U(u)du, \\ &= \int_{R_1} \big(u - v(u)\big)^2 f_U(u)du + \cdots + \int_{R_N} \big(u - v(u)\big)^2 f_U(u)du, \end{aligned} \tag{2.14}$$

donde, $N$ es el número de niveles de cuantificación y $V(U) = a_i$ para cada región $R_i$, luego:

$$MSE = \sum_{i=1}^{N} \int_{R_i} (u - a_i)^2 f_U(u)du. \tag{2.15}$$

Para el caso planteado por Gallager[14], el MSE del cuantificador escalar es aproximadamente (Gallager, 2008):

$$MSE \approx \frac{\Delta^2}{12} \tag{2.16}$$

### 2.2.2. Cuantificación Basada en Características Estadísticas

Si se deja de lado la suposición de que el número de niveles es grande, entonces se debe buscar la forma de disminuir la distorsión. Maximizar la entropía de $V$ es equivalente a disminuir la distorsión. En el caso de una variable aleatoria discreta con recorrido finito de tamaño $N$, su entropía se maximiza cuando los $N$ elementos del recorrido tienen la misma probabilidad.

Cuando $U$ tiene una pdf uniforme, las regiones de cuantificación serán del mismo tamaño; sin embargo, cuando esto no ocurre, las regiones deben ser de diferente tamaño, haciendo más pequeñas las más probables y más grandes las menos probables. En la Figura 2.9 se muestra como ejemplo una distribución normal con 8 regiones de cuantificación.

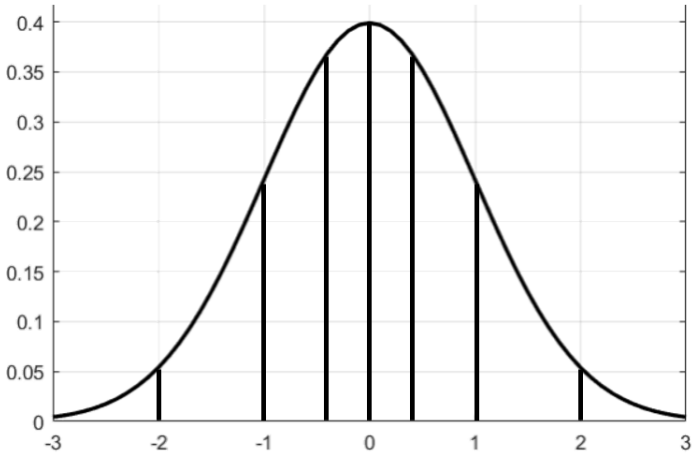


Figura 2.9. Cuantificador pdf no uniforme.
Elaboración propia.

Partiendo de la expresión matemática (2.15) y asumiendo que las regiones de cuantificación están dadas por $R_i: b_{i-1} < u < b_i$, entonces el MSE está dado por:

$$MSE = \sum_{i=1}^{N} \int_{b_{i-1}}^{b_i} (u - a_i)^2 f_U(u) du\,.$$

Los valores de los niveles de cuantificación se deben seleccionar con el propósito de reducir la distorsión. Para encontrar los valores $a_i$ que minimizan dicha distorsión, se iguala el gradiente del $MSE$ a 0, como se presenta a continuación:

[14] Ante un número de niveles de cuantificación suficientemente grande es posible asumir que dentro de cada región de cuantificación la pdf es uniforme.

$$\nabla(MSE) = \begin{bmatrix} \frac{\partial}{\partial a_i} MSE \\ \vdots \\ \frac{\partial}{\partial a_N} MSE \end{bmatrix} = \begin{bmatrix} -2\int_{b_0}^{b_1} (u - a_1) f_U(u) du \\ \vdots \\ -2\int_{b_{N-1}}^{b_N} (u - a_N) f_U(u) du \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

Despejando $\nabla(MSE) = 0$ se llega a que los valores óptimos para los niveles de cuantificación son las medias condicionales de cada región de cuantificación, entonces:

$$a_i = \frac{\int_{b_{i-1}}^{b_i} u f_U(u) du}{\int_{b_{i-1}}^{b_i} f_U(u) du}, \quad i = 1,2, \dots, N \tag{2.17}$$

El mismo método es aplicable para encontrar los límites de las regiones de cuantificación. En (2.18) se muestra la expresión para calcular los $b_i$ que minimizan el MSE

$$b_i = \frac{a_{i+1} + a_i}{2}, \qquad i = 1,2, \dots, N-1 \tag{2.18}$$

De las expresiones matemáticas (2.17) y (2.18), se observa que el cálculo de $a_i$ depende de los valores $b_i$ y que los $b_i$ dependen de $a_i$, por lo que no existe una solución analítica directa que minimice el MSE. Algoritmos como el de Lloyd-Max proponen acercamientos iterativos que de manera heurística buscan encontrar la configuración de $a_i$ y $b_i$ que logre un valor pequeño de MSE, pero esto no conduce necesariamente al mínimo valor de MSE posible (Lloyd, 1982).

Hasta el momento, se ha analizado el diseño del cuantificador partiendo del conocimiento de la pdf de la señal de entrada; sin embargo, la pdf no es la única información relacionada con el proceso aleatorio que representa a la señal de entrada. Cuando existe correlación entre las componentes de $\boldsymbol{U}$, la entropía conjunta de las variables aleatorias es menor o igual a la sumatoria de las entropías individuales (2.19). El caso de la igualdad se da únicamente cuando las variables aleatorias son independientes, por lo que la dependencia de las variables aleatorias implica que conjuntamente aportan menos información que la que aportan individualmente (2.20), esto es:

$$h(U_1, U_2, \dots, U_N) \leq \sum_{j=1}^{N} h(U_j). \tag{2.19}$$

$$I(\boldsymbol{U}; \boldsymbol{V}) \leq \sum_{i=1}^{N} I(U_i; V_i). \tag{2.20}$$

Dado que las variables aleatorias correlacionadas comparten información, cuantificarlas de forma independiente es ineficiente debido a la redundancia. Esto justifica el uso de la cuantificación vectorial (vectores aleatorios).

La cuantificación vectorial aumenta la complejidad del diseño, puesto que ya no se deben definir únicamente los valores de los niveles de cuantificación y el tamaño de las regiones, sino también la forma de éstas.

Teniendo en cuenta que las regiones del cuantificador deben particionar perfectamente el espacio $\mathbb{R}^d$, es decir, no deben dejar espacios en blanco ni tener traslapes, según el número de dimensiones se tendrán diferentes formas válidas. Para el caso 2D (vectores con dos componentes) las formas más comunes son: cuadrado, rectángulo, hexágono y triángulo (ver Figura 2.10); sin embargo, algunas de estas formas son más eficientes que otras. En (Gallager, 2008) se demuestra que el hexágono es la forma con la que se consigue el menor MSE (sin contar el círculo).

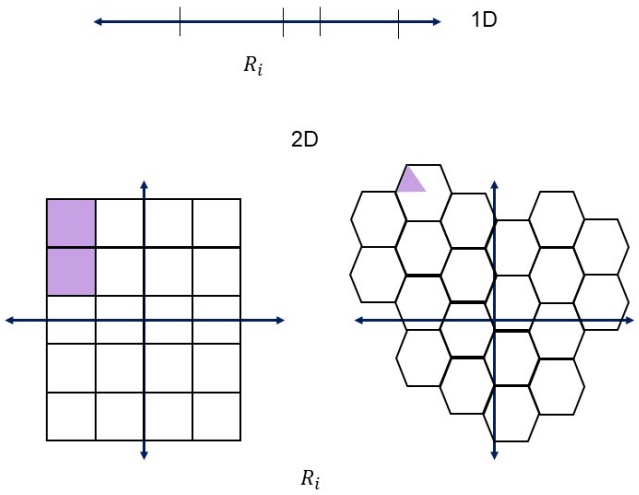


Figura 2.10. Formas válidas de las regiones de cuantificación para 1 y 2 dimensiones. Elaboración propia.

Aunque con un cuantificador vectorial es posible obtener una distorsión menor que con un cuantificador escalar (para un número de niveles fijo), las consideraciones que se deben tener para su implementación no se limitan al número de niveles de cuantificación, puesto que esto conduce a un aumento en la complejidad, el cual no justifica la reducción de la distorsión asociada.

Conocer las características estadísticas de las señales de entrada del cuantificador no es garantía para diseñar el mejor cuantificador y el costo computacional que demanda un diseño iterativo es alto. No obstante, éstas no son las únicas cuestiones que se deben considerar al momento de diseñar el cuantificador, puesto que por lo general no es posible conocer perfectamente las características estadísticas de las señales de entrada, generando errores provocados al asumir modelos estadísticos erróneos.

Los errores provocados al asumir modelos estadísticos erróneos se pueden observar cuando se hace una transformación de la pdf que rige la variable aleatoria que representa la señal de entrada. Conociendo la pdf de la variable aleatoria del proceso aleatorio es posible realizar una transformación para que la nueva variable aleatoria tenga una distribución uniforme y de esta forma utilizar un cuantificador cuyas regiones de cuantificación sean iguales. Bajo la suposición de que la variable aleatoria a transformar es continua, su transformación, $\ddot{U}$, es:

$$\ddot{U} = \mathrm{g}(U) = F_U(U), \tag{2.21}$$

donde $F_U$ es la Función de Distribución Acumulativa (CDF, *Cumulative Distribution Function*) de $U$; así, la nueva variable aleatoria es $\mathcal{U} \in [0,1]$ (ver Apéndice A).

Asumiendo que las señales de voz se rigen bajo una distribución Laplaciana, para su transformación es necesario conocer la CDF de esta distribución, a saber:

$$F_U(u) = \frac{1}{2}\left[1 + \mathrm{sgn}(u-\mu)\left(1 - e^{-\frac{|u-\mu|}{b}}\right)\right]. \tag{2.22}$$

Así, (2.23) muestra la transformación que se debe realizar sobre $U$ para que $\ddot{U}$ ahora tenga una distribución uniforme entre -1 y 1, esto es:

$$\ddot{U} = 2F_U(u) - 1, \tag{2.23}$$

esta transformación requiere que se conozcan con total certeza los parámetros de la distribución que rige a $U$. En caso de que se desconozcan tales parámetros, deben estimarse, lo cual conlleva a cierto grado de error.

Para realizar la transformación de la pdf existen otros enfoques. Dado que los valores de amplitud cercanos a cero son los más probables en las señales de voz, se utilizan funciones logarítmicas para amplificar en mayor medida los valores pequeños de amplitud, con el fin de compensar las diferencias en la distribución de amplitudes y de esta manera realizar una cuantificación más eficiente; sin embargo, dado que con este proceso se distorsionan los valores originales, es necesario realizar el proceso inverso después de que se ha realizado la cuantificación, esto hace parte del proceso conocido como *compansión* (compresión y expansión).

Las formas más utilizadas para transformar la señal y por consiguiente la pdf de una señal de voz antes de realizar el proceso de cuantificación son la Ley A y la Ley $\mu$, descritas mediante (2.24) y (2.25) y representadas en la Figura 2.11; no obstante, como se alejan de (2.23), el resultado de la transformación no es necesariamente una distribución uniforme.

$$g_A(U) = \begin{cases} \dfrac{\operatorname{sgn}(u)A|u|}{1+\ln(A)}, & |u| < \dfrac{1}{A} \\ \dfrac{\operatorname{sgn}(u)(1+\ln(A|u|))}{1+\ln(A)}, & \dfrac{1}{A} \le |u| \le 1 \end{cases} \quad ; \quad A = 87.6 \qquad (2.24)$$

$$g_\mu(U) = \frac{\operatorname{sgn}(u)\ln(1+\mu|u|)}{\ln(1+\mu)}, -1 \le u \le 1\,; \qquad \mu = 255 \qquad (2.25)$$

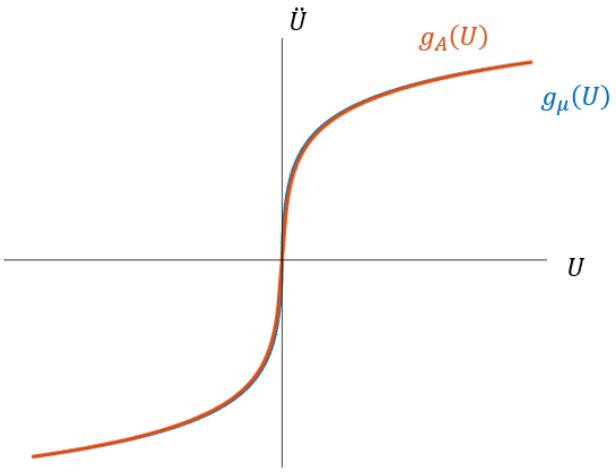


Figura 2.11. Transformación de la entrada por medio de la ley A y la ley μ.
Elaboración propia.

### 2.2.3. Codificación Diferencial

Los codificadores diferenciales cuantifican la diferencia entra las muestras consecutivas en lugar de las muestras individuales. El esquema básico de un codificador diferencial es el definido en la Modulación por Codificación de Pulso (PCM, *Pulse Code Modulation*), que consiste básicamente en discretizar la señal tanto en tiempo como en amplitud y representar el resultado mediante una secuencia binaria (ver Figura 2.12).

El esquema de PCM toma cada muestra de la señal de forma independiente; sin embargo, cuantificar la diferencia entre las muestras de la señal tiene como ventaja que el rango dinámico de la señal a cuantificar es menor (Sayood, 2006). Así, surgen variantes como la Modulación por Codificación de Pulso Diferencial (DPCM, *Differential Pulse Code Modulation*), ADPCM, la Modulación Delta (DM, *Delta Modulation*) y la Modulación Delta Adaptativa (ADM, *Adaptative Delta Modulation*). DPCM fue patentado en 1952 (Cutler, 1950), y su funcionamiento se sintetiza mediante el diagrama de bloques mostrado en la Figura 2.12.

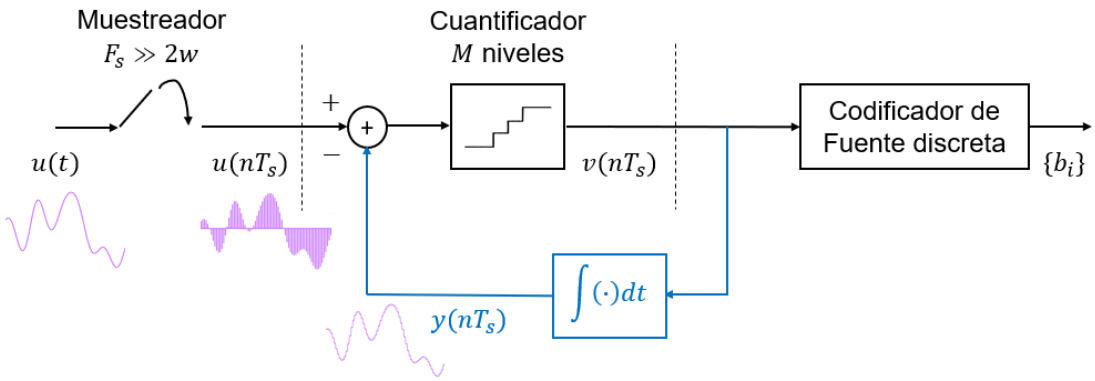


Figura 2.12. Codificador DPCM.
Elaboración propia.

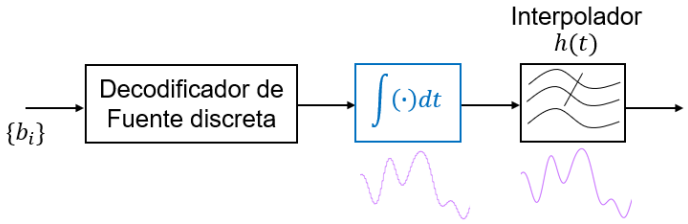


Figura 2.13. Decodificador DPCM.
Elaboración propia.

En la Figura 2.12 se resalta en azul el bloque necesario para estimar la muestra cuantificada disponible en el decodificador (Figura 2.13), esto se hace con el fin de evitar la acumulación de ruido de cuantificación (ver Apéndice A).

Los esquemas de cuantificación diferencial varían dependiendo de la complejidad del predictor utilizado, el número de niveles de cuantificación y la adaptabilidad del cuantificador. Los requerimientos en cuanto a calidad provocaron que el esquema básico se hiciera cada vez más complejo.

### 2.2.4. Cuantificación en un Dominio Transformado

El último enfoque de cuantificación es aquel donde las muestras de la señal se pueden transformar para ser cuantificadas en un dominio transformado alternativo (e.g. Fourier, coseno o *wavelet*) o en el mismo dominio por medio de una función no lineal (como el caso de los codificadores diferenciales). En ambos casos se cambia la representación de las muestras.

La cuantificación en el domino transformado se basa en que la distorsión inducida por el cuantificador en el dominio transformado no es la misma que induce el cuantificador sobre las muestras originales. Bajo este principio, en el procesamiento de imágenes y audio se han propuesto diferentes algoritmos de cuantificación/compresión (Sayood, 2006).

Para el caso de las señales de voz, las transformaciones más utilizadas se derivan del análisis de Fourier.

# CAPÍTULO 3: TRANSFORMADA WAVELET

En la Figura 3.1 se muestran las diferentes alternativas para abordar el análisis de las señales desde la perspectiva de los dominios del tiempo y la frecuencia. Del análisis de Fourier (Apéndice C) nace la posibilidad de conocer perfectamente el comportamiento de una señal en el dominio del tiempo (amarillo) y en el dominio de la frecuencia (azul), por lo que se puede entender como la forma más directa de generar una equivalencia entre éstos; sin embargo, su limitante es que no es posible tener alguna noción de lo que está pasando de forma simultánea en los dos dominios.

La alternativa en color rojo de la Figura 3.1 muestra el caso ideal, desde la perspectiva del procesamiento de señales, puesto que en éste se tiene una perfecta resolución tanto en el dominio del tiempo como en el dominio de la frecuencia, no obstante, el principio de incertidumbre de Heisenberg[15] es la limitante que vuelve a este caso irrealizable; por lo que, con el fin de encontrar un compromiso en la cantidad de incertidumbre que se tiene sobre cada uno de estos dos dominios, nacen alternativas como la Transformada de Fourier de Tiempo Corto (STFT, *Short Time Fourier Transform*) y la Transformada *Wavelet* (WT, *Wavelet Transform*).

La STFT limita la señal en el dominio del tiempo y analiza qué componentes de frecuencia ocurren en ese periodo de tiempo; sin embargo, cortar la señal en el dominio de tiempo tiene repercusiones en el dominio de la frecuencia, no sólo en las componentes de frecuencia que se pueden analizar[16], sino también en la forma del espectro, puesto que para cortar la señal en el dominio del tiempo se requiere de la multiplicación de la señal por una ventana que tiene asociado un espectro, así, el proceso de limitar la señal en el dominio del tiempo conlleva a una convolución en el dominio de la frecuencia.

La distorsión que sufre la señal en el dominio de la frecuencia, al implementar la STFT, corrobora que la FT es la forma más simple de relacionar la información en el dominio del tiempo con la información en el dominio de la frecuencia. La implementación de la STFT implica una decisión con respecto a la duración y el tipo de ventana, lo cual depende del contexto de aplicación (Kadambe, 1992).

[15] El principio de incertidumbre de Heisenberg en el análisis tiempo-frecuencia establece que no es posible conocer con completa exactitud lo que está ocurriendo en el tiempo y en la frecuencia de manera simultánea (Vošvrda & Schürrer, 2015), en otras palabras, entre mayor sea la precisión con la que se mida en un dominio mayor será la incertidumbre en el dominio complementario.

[16] Cortos periodos de tiempo únicamente permiten identificar altas frecuencias, conforme aumenta el periodo de tiempo aumenta la resolución en frecuencia y es posible identificar cada vez frecuencias más bajas.

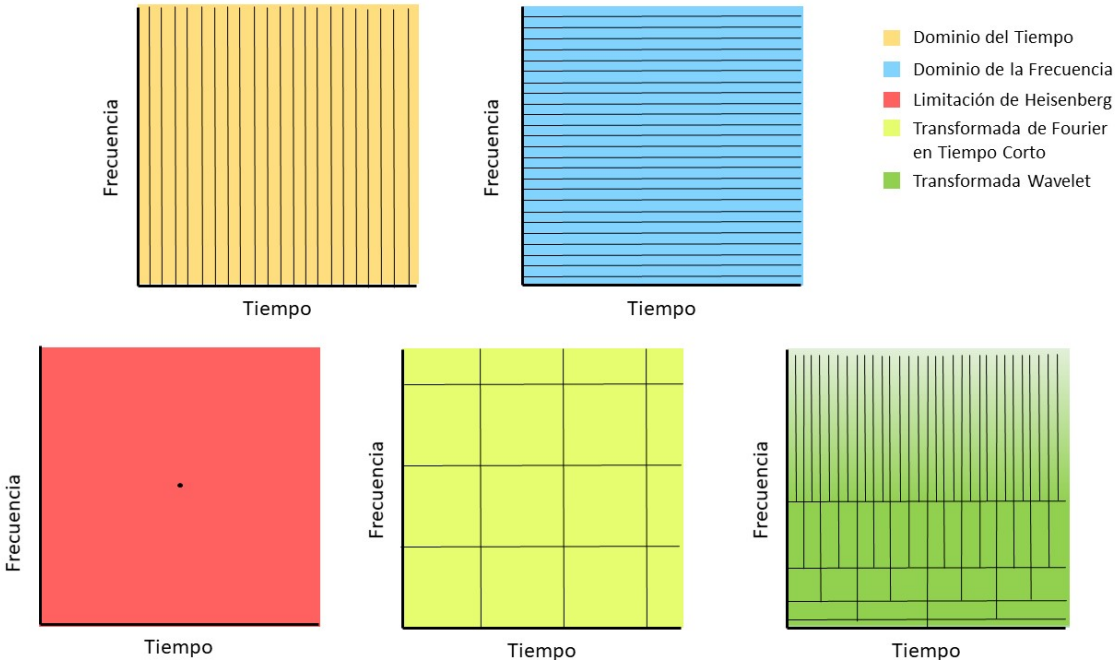


Figura 3.1. Resolución tiempo-frecuencia.
Elaboración propia.

La WT surge como una evolución natural de la STFT, dado que estas dos alternativas generan compromisos entre la resolución en el dominio del tiempo y en el dominio de la frecuencia. La historia del origen de la WT se puede contar desde diferentes perspectivas; sin embargo, Ingrid Daubechies muestra la creación del análisis en el dominio *wavelet* como una suma de esfuerzos de personas de diferentes disciplinas, las cuales notaron el potencial de modificar la forma con la que se trabajaba hasta el momento con la STFT (Daubechies, 1994b). En la STFT el tamaño de la ventana permanece fijo, ocasionando que la resolución en frecuencia sea constante, por lo que en la década de 1970 el ingeniero geofísico J. Morlet planteó la posibilidad de realizar cambios de escala y desplazamiento para de esta forma poder identificar los cambios bruscos que ocurren en cortos periodos de tiempo – asociados a altas frecuencias – y los cambios suaves que ocurren en grandes periodos de tiempo – asociados a bajas frecuencias -.

## 3.1. TRANSFORMADA WAVELET

La WT consiste en proyectar la señal sobre un subespacio generado por una familia bi-paramétrica de funciones, conocida como familia *wavelet*. Las funciones que hacen parte de la familia son versiones escalonadas y desplazadas de una forma de onda común, conocida como *wavelet* madre, la cual se simboliza con $\psi(\mathrm{t})$.

Las *wavelets* deben cumplir 3 condiciones de *admisibilidad*:

- Área bajo la curva igual a cero[17].

$$\int_{-\infty}^{\infty} \psi(t)dt = 0.$$

- Duración finita.

$$\psi(t) = 0, \forall\, t \notin [t_1, t_2],$$

donde $-\infty < t_1 < t_2 < \infty$. Luego, su energía finita se puede calcular de la forma:

$$\mathcal{E}_\psi \approx \int_{t_1}^{t_2} |\psi(t)|^2 dt\,, \qquad \mathcal{E}_\psi < \infty.$$

- Buena localización espectral[18].

$$\mathcal{E}_\psi \approx \int_{f_1}^{f_2} \left|\tilde{\psi}(f)\right|^2 df\,, \qquad -\infty < f_1 < f_2 < \infty.$$

Existen numerosas familias *wavelet* que además cumplen la característica deseable de ser ortogonales, entre las más conocidas se encuentran las de *Daubechies*, las *symlet*, las *coiflet*, y la de *Meyer*, entre otras (Daubechies, 1994a). Si las funciones de la familia son ortogonales entre sí, la WT permite representar con exactitud toda la información de la señal (Burrus et al., 1998).

Los dos parámetros utilizados en la WT son la escala, $a$, y la traslación, $b$, así, un miembro particular de la familia *wavelet*, $\psi_{a,b}(t)$, se define según (3.1).

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right). \tag{3.1}$$

La aplicación de la WT resulta entonces en una superficie, $w_x(a,b)$, dada por:

$$w_x(a,b) = \int_{-\infty}^{\infty} x(t)\psi_{a,b}(t)\,dt. \tag{3.2}$$

[17] A partir de esta condición se asume que la *wavelet* es una señal pasa banda.
[18] Dado que la *wavelet* es finita en el tiempo, no puede ser simultáneamente finita en la frecuencia.

Cabe resaltar que la WT es invertible, por lo que se garantiza la conservación de la información contenida en la señal $x(t)$. El proceso inverso se presenta en la siguiente expresión matemática:

$$x(t) = C \int_{0}^{\infty} \int_{-\infty}^{\infty} w_x(a,b)\psi_{a,b}(t)dbda, \qquad (3.3)$$

donde $C$ es una constante cuyo valor depende principalmente de la familia *wavelet* utilizada, pero que en la práctica no tiene mucha importancia.

Hasta el momento, se ha detallado la forma teórica por medio de la cual se define la WT, así como la ventaja que representa su carácter bi-paramétrico en comparación con alternativas como la STFT, sin embargo, la aplicación práctica de (3.2) no es posible. Por lo anterior, surge la necesidad de pensar en alternativas realizables. Una de ellas es la Transformada *Wavelet* Discreta (DWT, *Discrete Wavelet Transform*), la cual se define a continuación.

## 3.2. TRANSFORMADA WAVELET DISCRETA

La superficie resultante de aplicar la WT, $w_x(a,b)$, contiene información redundante, por lo que para su implementación práctica es necesario determinar los valores de escala y traslación que sean estrictamente necesarios para conservar la información, lo cual es equivalente a realizar un "muestreo" de dicha superficie, como se observa en la Figura 3.2.

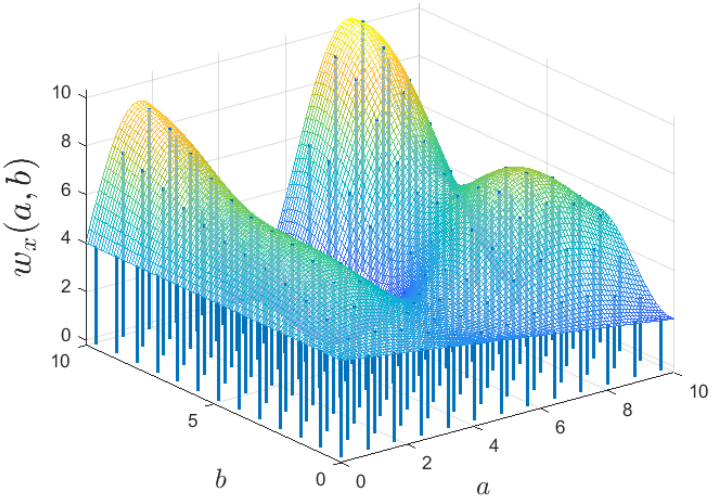


Figura 3.2. Discretización WT.
Elaboración propia.

*Mallat y Zhong* y *Daubechies* presentan explicaciones sobre la forma en la que se deben restringir los posibles valores de los parámetros de escala y traslación (ver Apéndice A sección A.4), lo cual lleva a que estos dos parámetros deben estar relacionados conforme (3.4) y (3.5) (Daubechies, 1994a; Mallat & Zhong, 1992).

$$a = 2^{-j}, \qquad j \in \mathbb{Z} \tag{3.4}$$

y

$$\begin{aligned} b &= ak \\ &= 2^{-j}k, \qquad k \in \mathbb{Z}\,. \end{aligned} \tag{3.5}$$

Así, los parámetros de la DWT son: $j$, conocido como nivel de resolución, y $k$, denominado nivel de traslación. La creación de la familia *wavelet* en la DWT se obtiene al reemplazar (3.4) y (3.5) en (3.1). El resultado obtenido se describe a continuación:

$$\psi_{j,k}(t) = 2^{\frac{j}{2}}\,\psi\big(2^j t - k\big); \quad j,k = 0, \pm 1, \pm 2, \cdots \tag{3.6}$$

En (3.6) se observa que, aunque se han restringido a desplazamientos *diádicos*[19], los niveles de resolución y traslación aún toman infinitos valores. La aplicación de la DWT para la obtención de los coeficientes *wavelet*, $w_x^j(k)$, se muestra en (3.7) y su proceso inverso en (3.8).

$$w_x^j(k) = 2^{\frac{j}{2}} \int_{-\infty}^{\infty} x(t)\psi\left(2^j t - k\right)dt, \tag{3.7}$$

$$x(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} 2^{\frac{j}{2}} w_x^j(k)\psi\big(2^j t - k\big). \tag{3.8}$$

En (3.8) se considera que la familia *wavelet* utilizada es ortogonal, i.e., cumple la condición descrita en (3.9).

$$\int_{-\infty}^{\infty} \psi_{j,k}(t)\psi_{m,n}(t) = \delta[j-m, k-n] = \begin{cases} 1, & j = m \text{ y } k = n \\ 0, & en\ cualquier\ otro\ caso \end{cases}, \tag{3.9}$$

donde $\delta[\cdot]$ es la función delta de kronecker.

Con la DWT se están creando particiones del espectro utilizando la familia *wavelet*, las cuales están relacionadas con el parámetro de escala o nivel de resolución; sin embargo, por más de que estas particiones sean muy pequeñas no hay ninguna que se encuentre alrededor del origen, así como no hay ninguna lo suficientemente grande como para abarcar toda la parte alta del espectro. En la Figura 3.3 se ejemplifica el caso ideal de estas particiones. La implementación de la DWT se logra a partir del Análisis Multiresolución (MRA, *MultiResolution Analysis*), el cual se explica a continuación.

[19] Se entiende como desplazamiento diádico al cambio de valores en función de potencias enteras de dos.

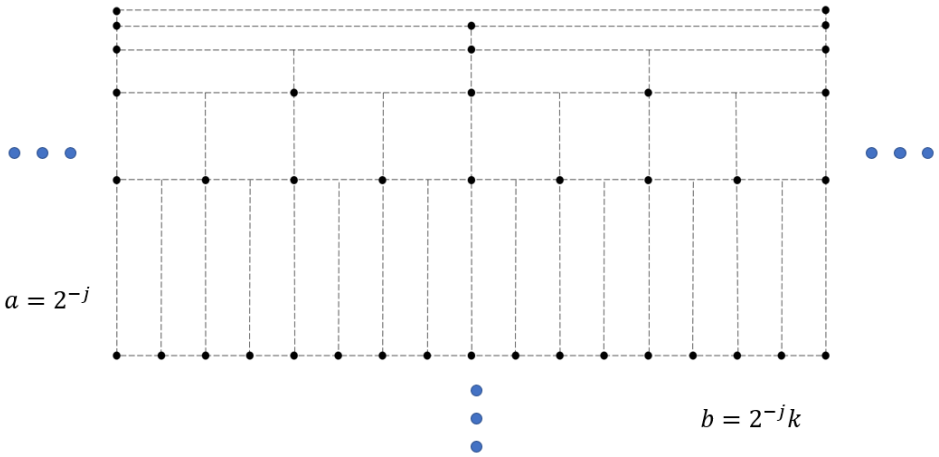


Figura 3.3. Resolución tiempo-frecuencia DWT.
Elaboración propia.

## 3.3. ANÁLISIS MULTIRESOLUCIÓN

En el MRA se asume que el conjunto de todos los niveles de resolución genera una partición del espacio de las señales de energía finita, $\mathcal{L}^2$, i.e., cada nivel es un subespacio de $\mathcal{L}^2$. Al subespacio generado por el nivel de resolución $j$ se denota como $W_j$.

Si la familia *wavelet* es ortogonal, se cumple que

$$W_j \cap W_m = \emptyset, \qquad \forall j \neq m, \tag{3.10}$$

de tal manera que el espacio $\mathcal{L}^2$ se genera a partir de la unión de todos los subespacios (ver Figura 3.4):

$$\bigcup_{j=-\infty}^{\infty} W_j = \mathcal{L}^2. \tag{3.11}$$

El conjunto de coeficientes $\{w_x^m(k), k \,\epsilon\, \mathbb{Z}\}$ resulta de la proyección de $x(t)$ sobre el subespacio $W_m$, y es único.

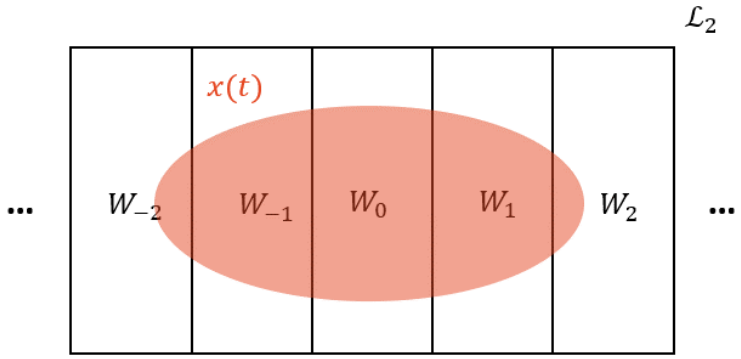


Figura 3.4. Subespacios *wavelet*.
Elaboración propia.

Por otro lado, en MRA se introduce el concepto de función *scaling,* $\varphi(t)$*,* la cual es una función que permite la reconstrucción de una señal; no obstante, difiere de las *wavelets* porque es una función en banda base, la cual tiene un área bajo la curva diferente de cero (Burrus et al., 1998).

El hecho de que $\varphi(t)$ sea una función en banda base tiene dos implicaciones importantes: la primera es que los subespacios *scaling*, $V_j$, del espacio $\mathcal{L}^2$*,* están contenidos dentro de los subespacios asociados a los niveles de resolución superiores, i.e.,

$$\cdots V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \cdots$$
$$V_j \subset V_{j+1}, \forall j \in \mathbb{z}; \tag{3.12}$$

y la segunda es que la completa proyección de la señal sobre el subespacio depende de la resolución, i.e.,

$$x(t) \in V_j \leftrightarrow x(dt) \in V_m, \qquad d \in \mathbb{Z}^+ \text{y } m > j. \tag{3.13}$$

El subespacio *scaling* de resolución $j$*,* $V_j$*,* se genera a partir de las funciones *scaling* asociadas a dicho nivel de resolución, como se muestra en (3.14).

$$\varphi_{j,k}(t) = 2^{\frac{j}{2}}\varphi\left(2^j t - k\right), \qquad j,k \in \mathbb{Z}. \tag{3.14}$$

Las funciones son mutuamente ortogonales para un mismo nivel de resolución, pero no lo son para diferentes niveles. Los coeficientes *scaling*[20], $s_x^j(k)$, se obtienen de la proyección de la señal sobre el subespacio *scaling* del mismo nivel de resolución (3.15), de tal forma que se garantiza su inversión, como se presenta a continuación:

[20] En este trabajo de grado se nombran los grupos de coeficientes según la función utilizada para su obtención, es decir, coeficientes *wavelet* y *scaling*; no obstante, en la literatura es muy utilizada la denominación heredada del procesamiento de imágenes, en la cual a los coeficientes *wavelet* se les conoce como coeficientes de detalles y a los coeficientes *scaling* como coeficientes de aproximación (Stéphane Mallat, 2009).

$$s_x^j(k) = \int_{-\infty}^{\infty} x(t)\varphi_{j,k}(t)dt \tag{3.15}$$

$$x^{(j)}(t) = \sum_k s_x^j(k)\varphi_{j,k}(t) \tag{3.16}$$

La aplicación del MRA consiste en utilizar de manera complementaria los subespacios *wavelet* y *scaling*, como se muestra en la Figura 3.5.

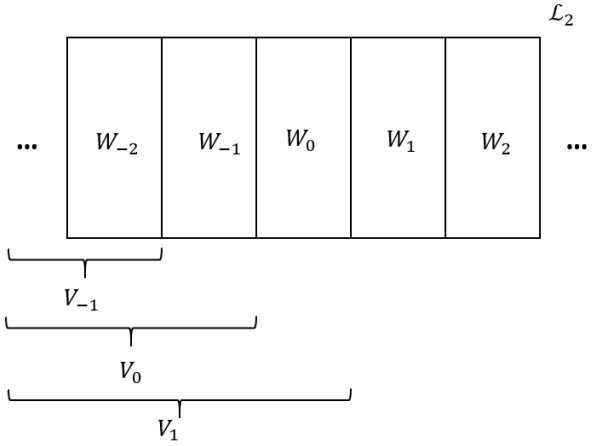


Figura 3.5.subespacios *scaling* y *wavelet*.
Elaboración propia.

De la Figura 3.6 se infiere que, para un mismo nivel de resolución $j$, el producto interno entre una *wavelet* y una *scaling* es igual a cero, i.e., son funciones ortogonales:

$$\langle \psi_{j,k}(t), \varphi_{j,l}(t) \rangle = \int_{-\infty}^{\infty} \psi_{j,k}(t)\varphi_{j,l}(t)dt = 0$$

o de forma equivalente,

$$W_j \cap V_j = \emptyset. \tag{3.17}$$

Adicionalmente,

$$V_j = W_{j-1} \cup V_{j-1},$$

por lo tanto,

$$V_j = \bigcup_{m=-\infty}^{j-1} W_m. \tag{3.18}$$

Lo anterior permite que la señal analizada se represente por medio de una combinación de coeficientes *wavelet* y *scaling*, como se muestra a continuación:

$$x(t) = \sum_{k=-\infty}^{\infty} s_x^0(k)\varphi_{0,k}(t) + \sum_{j=0}^{\infty}\sum_{k=-\infty}^{\infty} w_x^j(k)\psi_{j,k}(t) \tag{3.19}$$

En la DWT, el teorema de Rayleigh, que garantiza la conservación de la energía se cumple cuando la familia *wavelet* es ortogonal:

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = \sum_{k=-\infty}^{\infty} |s_x^0(k)|^2 + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} \left|w_x^j(k)\right|^2 \qquad (3.20)$$

La implementación computacional de MRA se lleva a cabo a través de un conjunto de operadores discretos (filtros digitales, diezmadores y sobremuestreadores) dentro de un proceso conocido como Transformada Rápida *Wavelet* (FWT, *Fast Wavelet Transform*) o algoritmo de Mallat (Stéphane Mallat, 2009).

## 3.4. ALGORITMO DE MALLAT

En la parte izquierda de la Figura 3.6 se muestra la descomposición de la señal de tiempo discreto $x[n]$ en sus respectivos coeficientes *wavelet* y *scaling*, lo cual es equivalente a la DWT, mientras que en la parte derecha de la imagen se sintetizan los coeficientes para realizar el proceso inverso.

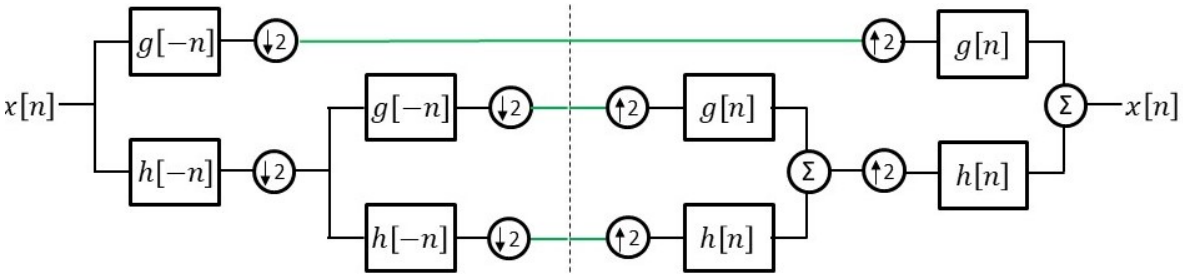


Figura 3.6. Algoritmo de Mallat.
Elaboración propia.

El algoritmo de Mallat utiliza un banco de filtros FIR que permiten crear los filtros *wavelet*, $g[n]$, y los *scaling*, $h[n]$. En la expresión matemática (3.21) se muestra el proceso por medio del cual se obtienen los coeficientes del filtro $h[n]$, el cual corresponde a un filtro pasa bajas, por su parte, en la expresión matemática (3.22) se encuentra el cálculo de $g[n]$, que corresponde a un filtro pasa altas, como se muestra a continuación:

$$h[n] = \sqrt{2} \int_{-\infty}^{\infty} \varphi(t)\varphi(2t-n)dt\,, \forall n \in \mathbb{Z}, \qquad (3.21)$$

y

$$g[n] = \sqrt{2} \int_{-\infty}^{\infty} \psi(t)\varphi(2t-n)dt\,, \forall n \in \mathbb{Z}. \qquad (3.22)$$

Adicionalmente, $h[n]$ y $g[n]$ deben cumplir las condiciones de ortogonalidad detalladas a continuación:

$$\sum_{n=-\infty}^{\infty} h[n]h[n-2m] = \delta[m], \tag{3.23}$$

$$\sum_{n=-\infty}^{\infty} g[n]g[n-2m] = \delta[m], \tag{3.24}$$

y

$$\sum_{n=-\infty}^{\infty} h[n]g[n-2m] = \delta[m]. \tag{3.25}$$

Dado que a partir de funciones finitas en el tiempo, como lo son las *wavelets* y *scaling*, no es posible conseguir una partición ideal del espectro, se busca que en el dominio de la frecuencia los filtros utilizados tengan simetría, de tal forma que se pueda aprovechar el traslape presente en cada una de las subbandas generadas. Los filtros utilizados a fin de obtener una reconstrucción perfecta son los Filtros Espejo en Cuadratura (QMF, *Quadrature Mirror Filter*), con los cuales se obtiene una división del espectro como la mostrada en la Figura 3.7.

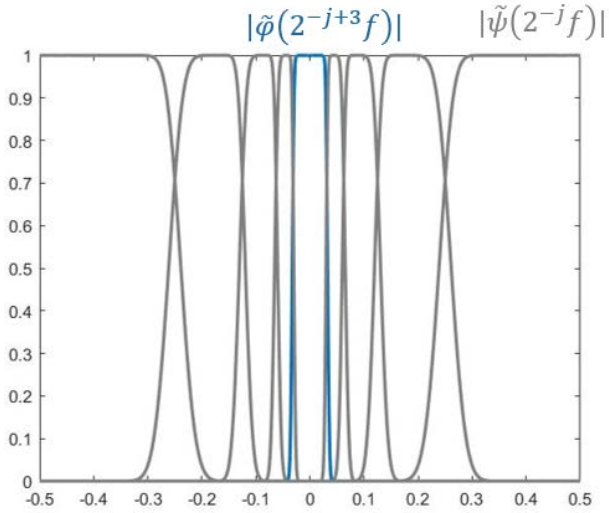


Figura 3.7. Divisiones del espectro con la FWT.
Elaboración propia.

Es importante mencionar que los diezmadores y sobremuestreadores presentes en el algoritmo de Mallat tienen la función de acotar el número de muestras en función del ancho de banda de la señal, según lo establecido en el teorema de muestreo de Nyquist-Shannon.

Dada la importancia del comportamiento en frecuencia de los filtros utilizados en el algoritmo de Mallat, Daubechies (Burrus et al., 1998; Daubechies, 1994a) propone

partir del análisis en la frecuencia de la función *scaling* para la creación del banco de filtros de la FWT, dado que gracias a la relación que existe entre las *wavelets* y las *scaling* (ver ecuación 3.21), es posible crear $g[n]$ a partir de $h[n]$, como se muestra a continuación:

$$g[n] = (-1)^n h\big[L_f - 1 - n\big], \qquad n = 0{:}\, L_f - 1, \tag{3.26}$$

donde $L_f$ es la longitud de $h[n]$.

No obstante, es importante aclarar que no todas las funciones *wavelet* y *scaling* se pueden construir a partir de filtros FIR, siendo la *wavelet* de Ricker o *mexican hat* un ejemplo de esto. En el Apéndice D se detallan diferentes familias *wavelet* ortogonales que se pueden utilizar en el algoritmo de Mallat.

# CAPÍTULO 4: DISEÑO DEL CUANTIFICADOR

El presente documento final de trabajo de grado de maestría sigue una metodología o modelo en 'V', por lo que en este capítulo se detalla todo lo relacionado con la fase de diseño, para lo cual se debe partir de unos requerimientos o especificaciones (Balaji & Murugaiyan, 2012).

## 4.1. ESPECIFICACIONES

El algoritmo de cuantificación de señales de voz utilizando *wavelets* toma una señal en tiempo discreto $u[n]$, aplica la DWT y cuantifica los coeficientes resultantes. La señal reconstruida con los coeficientes cuantificados $v[n]$ es una versión distorsionada de $u[n]$. En la Figura 4.1 se ejemplifica el funcionamiento del algoritmo de cuantificación utilizando dos etapas del algoritmo de Mallat.

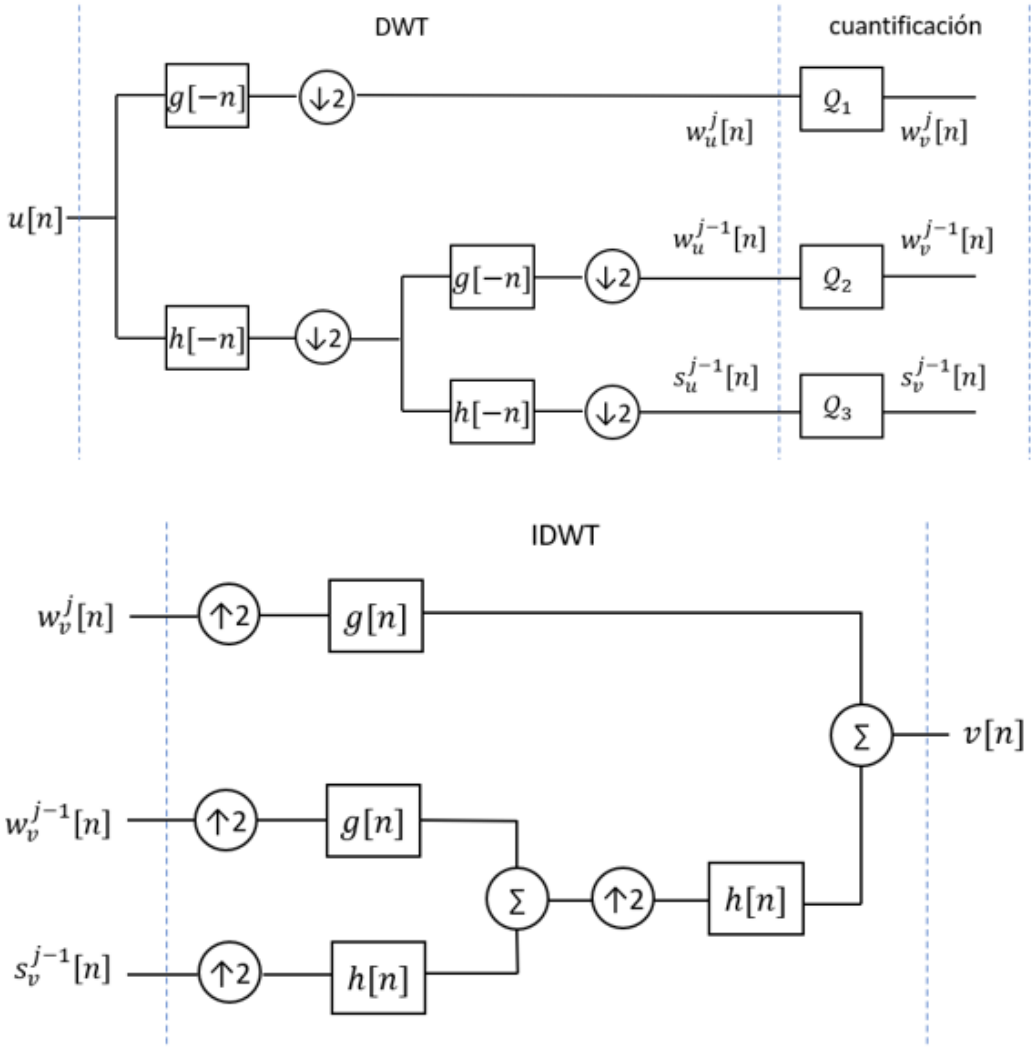


Figura 4.1. Diagrama de bloques del algoritmo de cuantificación utilizando *wavelets*.

Elaboración propia.

La cantidad de etapas que constituyen el algoritmo de Mallat o el MRA determina el número de subbandas en las que se divide el espectro (ver Figura 4.2) y el número de coeficientes que se deben cuantificar.


Figura 4.2. Subbandas del algoritmo de Mallat con 2 etapas.
Elaboración propia.

Si se tiene un número de etapas del algoritmo de Mallat $\xi = 2$, entonces el vector de coeficientes es de la forma:

$$c_u = \left[s_u^{j-1} \ \vdots \ w_u^{j-1} \ \vdots \ w_u^{j}\right]. \tag{4.1}$$

La longitud de longitud del vector de coeficientes $L_c$ está dada por la longitud de la señal $L_u$ y la longitud de los filtros utilizados $L_f$.

$$L_c = \left\lfloor \frac{L_u}{2^{\xi}} + \frac{2^{\xi}-1}{2^{\xi}} L_f \right\rfloor + \sum_{k=1}^{\xi} \left\lfloor \frac{L_u}{2^{k}} + \frac{2^{k}-1}{2^{k}} L_f \right\rfloor \tag{4.2}$$

A partir de (4.2) se observa que si $L_f$ tiene un valor pequeño en comparación con $L_u$, entonces $L_c \approx L_u$. Adicionalmente, entre mayor sea el valor de $\xi$, mayor será el número de muestras adicionales, haciendo que $L_c > L_u$.

Una vez caracterizados los coeficientes resultantes del algoritmo de Mallat y el funcionamiento general de la cuantificación utilizando wavelets, se definen las siguientes premisas

- En general, el proceso de cuantificación puede ser utilizado con fines de digitalizar una señal analógica o de comprimir una señal digital. En este caso, las pruebas para verificar el desempeño del planteamiento teórico tienen lugar en un entorno de simulación, por lo que la señal de entrada es digital, y por tanto corresponde al proceso de compresión; no obstante, esto no impide que pueda ser utilizado en procesos de digitalización de señales analógicas, una vez se muestree la señal.

En la Figura 4.3 se muestra un ejemplo de aplicación del algoritmo de aplicación planteado, en el cual se asume que la señal de voz es analógica y necesita ser digitalizada para poder ser transmitida por un sistema de telecomunicaciones digital.

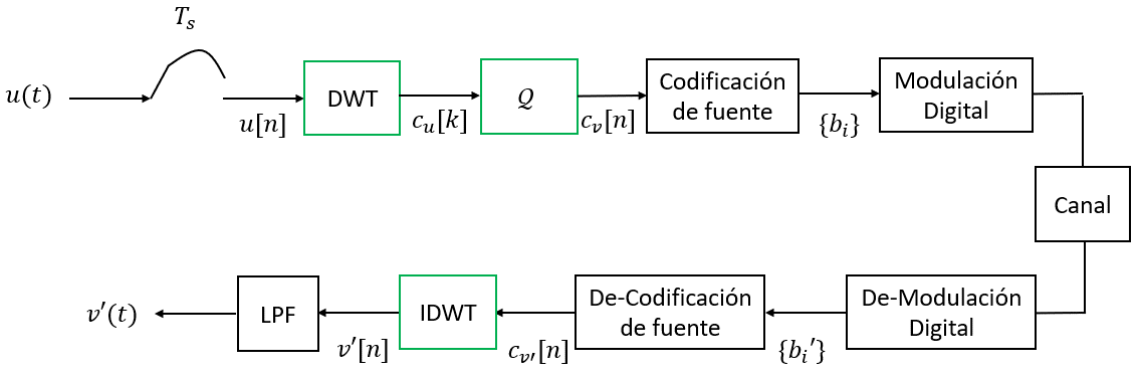


Figura 4.3. Ejemplo de aplicación de la cuantificación utilizando *wavelets*. Elaboración propia.

- El diseño del cuantificador está enfocado en minimizar la distorsión[21] de la señal resultante, por lo que la restricción del cuantificador es el número de niveles de cuantificación, $N = \{8, 16, 32, 64, 128\}$. Con esto, lo que se busca corroborar es si la aplicación de la representación *wavelet* sobre el proceso de cuantificación de la señal de voz permite reducir la distorsión en comparación con la cuantificación de estas señales en el dominio del tiempo o en el dominio de la frecuencia.

- Las señales de voz a utilizar son 33750 señales digitales en formato 'wav' (8 bits por muestra), muestreadas a 16 KHz y con una duración de 64 milisegundos cada una (ver Apéndice E).

## 4.2. ANTECEDENTES

La DWT es ampliamente utilizada en algoritmos de compresión. Una de las razones es que de su aplicación se obtiene un número de coeficientes *wavelet* igual al número de muestras de la señal y que la energía de la señal puede estar concentrada únicamente en unos cuantos coeficientes. Además, es adecuada para el análisis de señales no estacionarias, porque busca relacionar los dominios del tiempo y la frecuencia. Lo anterior hace de la DWT una técnica efectiva para la compresión de señales de voz (Joseph & Babu Anto, 2012; Vig & Chauhan, 2018).

Técnicas basadas en DWT son utilizadas también para mejorar el desempeño de *vocoders* existentes. *Seto y Ogunfunmi* proponen una codificación utilizando la DWT

[21] En el Apéndice B se muestran medidas objetivas y subjetivas para evaluar la distorsión introducida por el cuantificador.

del error de codificación[22] de las señales producidas por el *vocoder* iLBC (Seto & Ogunfunmi, 2019). *Wang et al.* proponen un *vocoder* con una baja tasa de bit (3.3 kbps), mediante la caracterización del espectro de magnitud de la señal de voz utilizando familias *wavelet* biortogonales (Wang et al., 2007).

Finalmente, *Bousselmi y Ouni* comparan la robustez frente al ruido de señales de voz que se descomponen utilizando la DWT y la Transformada de Paquetes de Marco Estricto (TFPT, *Tight Framelet Packet Transform*), obteniendo mejores resultados con la TFPT (Bousselmi & Ouni, 2017b). La TFPT utiliza como funciones base familias *wavelet* que no son ortogonales, es decir, que en los niveles de descomposición hay información redundante, la cual es útil en ciertas aplicaciones (Lu & Fan, 2011). La estabilidad de la reconstrucción con TFPT se evalúa por *Bousselmi y Ouni*, comparando en este caso su desempeño frente a los Paquetes *Wavelet* (WP, *Wavelet Packet*), obteniendo resultados congruentes con sus otros experimentos (Bousselmi & Ouni, 2017a, 2017b).

## 4.3. ALGORITMOS DE CUANTIFICACIÓN

Para evaluar el desempeño de la cuantificación de señales de voz en el dominio *wavelet* se utilizan dos enfoques: cuantificación uniforme y cuantificación no uniforme.

### 4.3.1. Cuantificación Uniforme

Para el diseño del cuantificador uniforme se concatenan los coeficientes resultantes del algoritmo de Mallat para definir el rango dinámico $D$, a partir del cual se distribuyen los niveles de cuantificación.

$$D = \{\min(c_u[n]), \max(c_u[n])\}, \tag{4.3}$$

donde $c_u[n]$ es un vector que contiene los coeficientes resultantes de la aplicación del algoritmo de Mallat.

Así, la longitud del intervalo está dada por:

$$\Delta = \frac{\max(c_u[n]) - \min(c_u[n])}{N}. \tag{4.4}$$

Los valores de los niveles de cuantificación, $a_i$, corresponden al punto medio de cada intervalo, i.e.,

$$a_i = \min(c_u[n]) + \frac{\Delta}{2}(2N - 1). \tag{4.5}$$

[22] El error de codificación no corresponde a la señal de voz, sino que corresponde a información adicional que se utiliza en los *vocoders* más avanzados para mejorar la calidad de la señal de voz digitalizada.

A partir de las regiones y niveles de cuantificación se obtiene la característica de transferencia del cuantificador uniforme, $Q$. El diagrama de bloques de este cuantificador se muestra en la Figura 4.4.

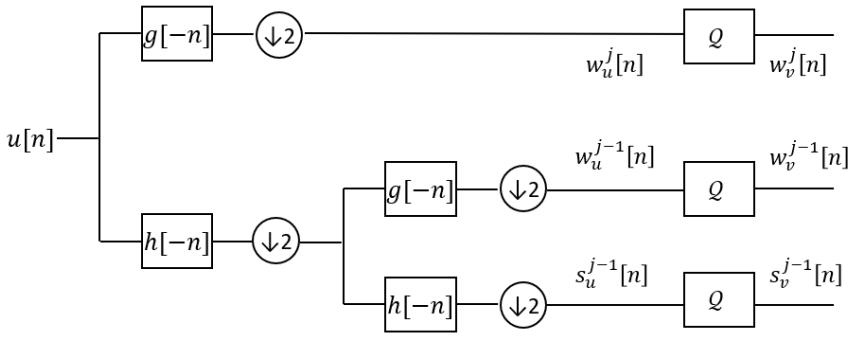


Figura 4.4. Diagrama de bloques del cuantificador uniforme.
Elaboración propia.

Es importante resaltar que cuando se utiliza una secuencia binaria para representar los coeficientes cuantificados se requiere de información adicional. Esta información corresponde a los descriptores del cuantificador, a partir de los cuales se puede reconstruir la característica de transferencia del cuantificador y recuperar los valores de los coeficientes cuantificados. Los descriptores del cuantificador son: $\max(c_u[n])$, $\min(c_u[n])$ y $N$.

En la Figura 4.5 se muestra el diagrama de flujo del cuantificador uniforme, a partir del cual se realiza la implementación computacional de este algoritmo.

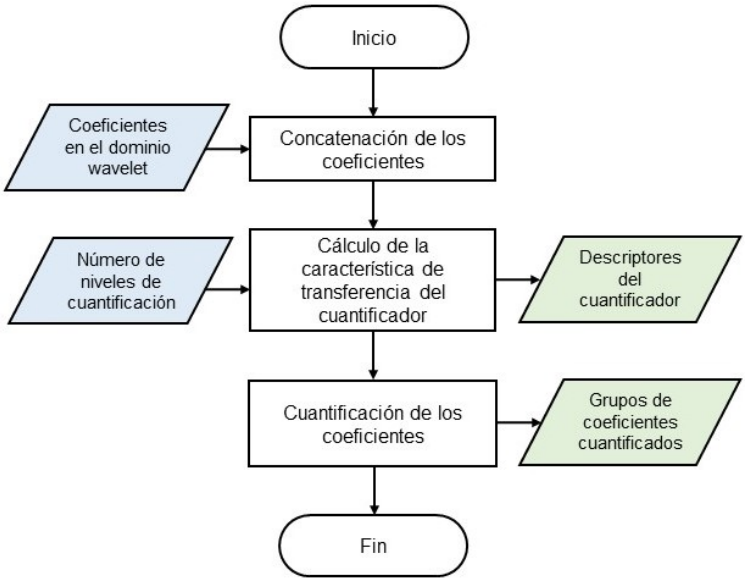


Figura 4.5. Diagrama de flujo cuantificador uniforme.
Elaboración propia.

### 4.3.2. Cuantificación No Uniforme

Para explicar el planteamiento del algoritmo de cuantificación no uniforme es necesario explicar primero el *índice de Gini*.

#### *Índice de Gini*

Las curvas de Lorenz comúnmente son utilizadas para determinar la forma en la que se distribuye la riqueza en una población (equitativa o inequitativamente), cuya información se puede sintetizar por medio del *índice de Gini*. En la expresión matemática (4.6) se detalla el cálculo del *índice de Gini*, el cual puede tomar valores entre 0 y 1, el cual se presenta a continuación:

$$Gini = \frac{A}{A+B}, \tag{4.6}$$

donde, $A$ es el área entre la curva obtenida y la recta que representa una distribución completamente equitativa, mientras que $B$ es el área bajo la curva obtenida (ver Figura 4.6).


Figura 4.6. Curva de Lorenz.
Elaboración propia.

En el caso del conjunto de coeficientes del MRA de una señal, el *índice de Gini* indica la forma cómo se distribuye la energía de la señal original sobre dicho conjunto (Ogden & Vidakovic, 2000). Un *índice de Gini* cercano a 1 indica que pocos coeficientes resultantes del MRA son los que más aportan a la reconstrucción de la señal original.

Para el cálculo de la curva de Lorenz se toman los coeficientes al cuadrado:

$$\varepsilon_c[n] = (c_u[n])^2, \tag{4.7}$$

se organizan de forma ascendente iniciando por cero:

$$\varepsilon_c[n] \leftarrow [0 : \mathrm{asc}(\varepsilon_c[n])], \tag{4.8}$$

donde $asc(\cdot)$ es la función que organiza de forma ascendente los elementos de su argumento.

Posteriormente, estos valores se normalizan con respecto a la energía de la señal $\mathcal{E}_u$, tal como se presenta a continuación:

$$\ell_c[n] = \frac{\varepsilon_c[n]}{\mathcal{E}_u}, \tag{4.9}$$

donde $\ell_c[n]$ son los elementos a partir de los cuales se realiza el cálculo de la curva de Lorenz $\ell_B$, por medio de una suma acumulativa, como se muestra a continuación:

$$\ell_B[n] = \sum \ell_c[n]. \tag{4.10}$$

En la Tabla D.2 del Apéndice D se muestran los resultados promedio del *índice de Gini* para 100 familias *wavelet* ortogonales, variando el número de etapas del algoritmo de Mallat desde uno hasta tres. Para estos resultados se cumple que $0.7999 \leq Gini \leq 0.9521$, lo cual indica que se tiene una alta desigualdad en la distribución de la energía, siendo los coeficientes *scaling* los de mayor concentración de la energía (ver Figura 4.7).

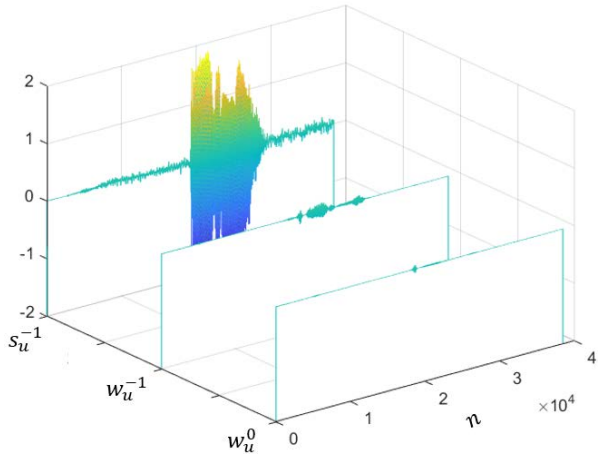


Figura 4.7. Coeficientes wavelet de una señal de voz.
Elaboración propia.

Por lo anterior, se evidencia la conveniencia de realizar una cuantificación independiente para los coeficientes resultantes del MRA, tal como se muestra en la Figura 4.1. Se debe resaltar que emplear un cuantificador diferente para cada conjunto de coeficientes implica que se aumenta la información adicional, puesto que se deben describir todos los cuantificadores utilizados. El número de cuantificadores es igual al número de etapas más uno.

### Asignación del número de niveles de cuantificación

Gracias al *índice de Gini* se comprueba que en los coeficientes resultantes del algoritmo de Mallat se tiene una distribución desigual de la energía. Por lo tanto, es a partir del porcentaje de energía de la señal contenido en cada subbanda que se

realiza la asignación del número de niveles de cuantificación para cada una de éstas.

Para realizar el cálculo del número de niveles de cuantificación correspondiente para cada subbanda, es importante considerar que la longitud de la secuencia binaria resultante de la cuantificación no uniforme no puede superar a la longitud de la cuantificación uniforme, i.e.,

$$L_c \log_2 N \geq \sum_{k=1}^{K} L_c^k \log_2 N_k, \tag{4.11}$$

donde, $L_c$ es el número de coeficientes en $c_u[n]$; $K = \xi + 1$ es el número de subbandas resultantes del MRA (número de etapas más uno); y $L_c^k$ y $N_k$ son el número de coeficientes y niveles de cuantificación de cada subbanda, respectivamente.

El porcentaje de energía contenido en cada subbanda está dado por:

$$\varepsilon_k = \frac{1}{\varepsilon_u} \sum c_u^k[n], \tag{4.12}$$

donde $c_u^k[n]$ son los coeficientes dentro de cada una de las $k$ subbandas. El vector que contiene el porcentaje de energía contenido en cada subbanda se denota como $\varepsilon_K[n]$, en el cual se encuentran organizados los valores de forma descendente.

El cálculo de los niveles de cuantificación $N_k$ se realiza de forma iterativa recorriendo el vector $\varepsilon_K[n]$, para lo cual se parte de que la totalidad de recursos disponibles es $r = L_c \log_2 N$, así:

$$N_k = 2^{\left\lfloor \frac{\varepsilon_k r}{L_c^k} \right\rfloor}, \tag{4.13}$$

donde $\lfloor \cdot \rfloor$ es el operador parte entera inferior.

En la primera iteración se calcula el número de niveles de cuantificación para la subbanda con mayor porcentaje de energía y se reduce la cantidad de recursos disponibles:

$$r \leftarrow r - L_c^k \log_2 N_k. \tag{4.14}$$

En la Figura 4.8 se muestra el diagrama de flujo del algoritmo de cuantificación no uniforme utilizando *wavelets*.

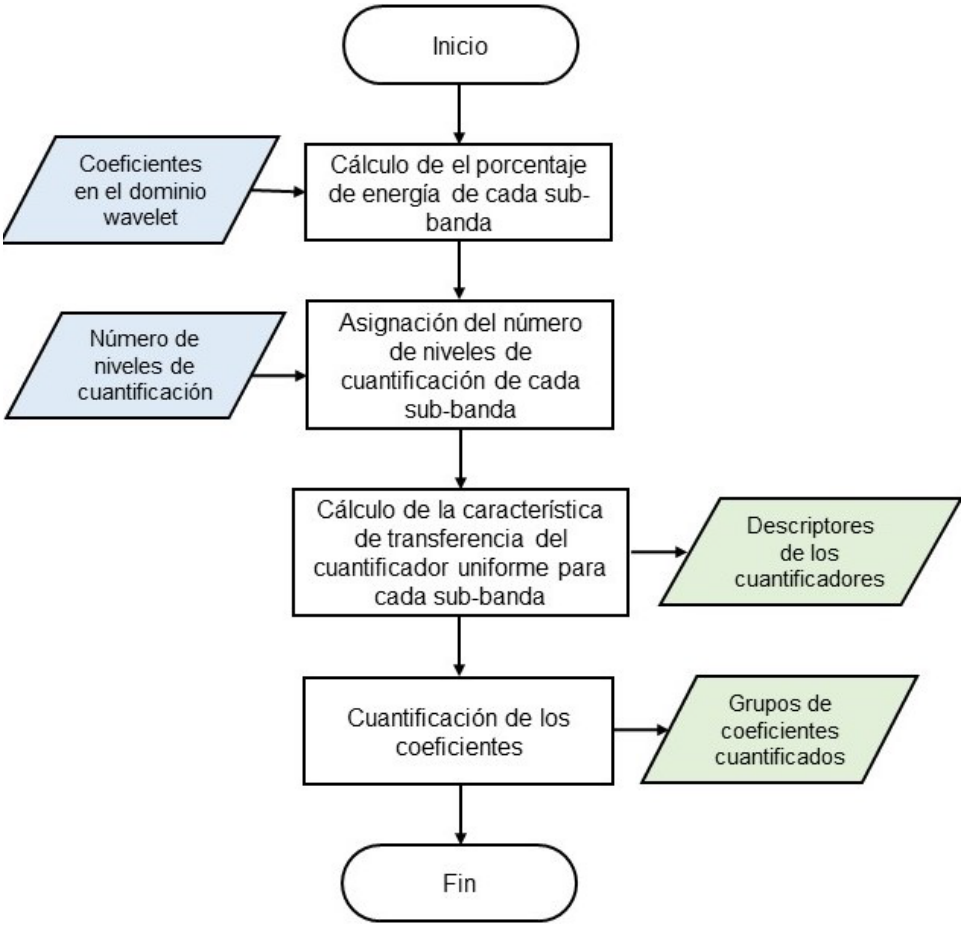


Figura 4. 8. Diagrama de flujo cuantificador no uniforme.
Elaboración propia.

## 4.4. EVALUACIÓN DE LOS ALGORITMOS DE CUANTIFICACIÓN

Para la evaluación de los algoritmos de cuantificación se utilizan medidas objetivas y subjetivas, las cuales se encuentran en mayor detalle en el Apéndice B.

### 4.4.1. Medida Objetiva

Entre las diferentes medidas objetivas utilizadas para medir la distorsión (ver Apéndice F) se destacan tres, dado que cumplen con que su magnitud medida es menor o igual a 1.

La primera de estas medidas es una versión normalizada del MSE denotada como $M - NRMSE$ (ver ecuación 4.14). En este caso un resultado igual a uno indica que las dos señales son idénticas.

$$M - NRMSE = 1 - \frac{1}{2} \cdot \sqrt{\frac{\mathrm{E}[(U-V)^2]}{\mathrm{E}[U^2]}}. \tag{4.14}$$

La siguiente medida corresponde al coeficiente de correlación de Pearson $\rho$, a partir del cual se determina la intensidad de la dependencia lineal entre la entrada y salida del cuantificador, esta medida toma valores entre -1 y 1, i.e.,$-1 \leq \rho \leq 1$.

$$\rho(u,v) = \frac{\mathrm{cov}(U,V)}{\sigma_U \sigma_V}. \tag{4.15}$$

Finalmente, se propone el uso del índice de similitud estructural notado como $SSIM$. Este índice no realiza una comparación directa entre las dos señales, sino que realiza una ponderación de los valores de: luminosidad ($\ell$), contraste ($c$) y estructura ($\mathcal{s}$).

$$\ell(U,V) = \frac{2\mu_U\mu_V + C_1}{\mu_U^2 + \mu_V^2 + C_1}, \tag{4.16}$$

$$c(U,V) = \frac{2\sigma_U\sigma_V + C_2}{\sigma_U^2 + \sigma_V^2 + C_2}, \tag{4.17}$$

y

$$\mathcal{s}(U,V) = \frac{\sigma_{UV} + C_3}{\sigma_U\sigma_V + C_3}, \tag{4.18}$$

donde, $\mu_U$ y $\mu_V$ son las medias de $U$ y $V$, respectivamente; $\sigma_U$ y $\sigma_V$ sus desviaciones estándar; $\sigma_{UV}$ es la covarianza de las dos señales; y $C_1 = 0.0001, C_2 = 0.0009$ y $C_3 = 0.00045$ son valores constantes introducidos para evitar inestabilidad (*Structural Similarity (SSIM) Index for Measuring Image Quality - MATLAB Ssim*, n.d.).

Finalmente, el $SSIM$ se obtiene de una combinación de las tres medidas, donde los valores constantes $\alpha = \beta = \gamma = 1$ determinan la importancia relativa de cada una de estas medidas. El valor máximo del $SSIM$, indicando que las señales comparadas son idénticas es:

$$SSIM = \ell(U,V)^{\alpha} c(U,V)^{\beta} \mathcal{s}(U,V)^{\gamma}. \tag{4.19}$$

No obstante, con el fin de sintetizar los resultados objetivos, se crea una medida objetiva promedio $p$, dada por:

$$p = \frac{M - NRMSE + \rho + SSIM}{3}. \tag{4.20}$$

Cuando $p = 1$ las dos señales comparadas son estadísticamente iguales.

### 4.4.2. Medida Subjetiva

Para la evaluación subjetiva es importante resaltar que las señales de audio originales y cuantificadas se deben escuchar dentro del entorno de simulación MATLAB®, por lo que para implementar la $MOS$ que representa la evaluación subjetiva se utiliza un programa interactivo, en el cual las personas pueden escuchar los audios y asignarles una valoración. Adicionalmente, se utilizaron los mismos audífonos y nivel de volumen para homogeneizar las condiciones las pruebas. En la Figura 4.9 se muestra el programa por medio del cual se pueden evaluar los diferentes audios.


Figura 4.9. Aplicación de la MOS.
Elaboración propia.

Dado que cada categoría de evaluación tiene asignado un valor numérico:

malo (1) - pobre (2) - regular (3) - bueno (4) - excelente (5),

entonces, si se tienen $e$ evaluaciones de un mismo audio se tiene que:

$$MOS = \frac{1}{e}\sum_{l=1}^{e} OS_l, \tag{4.21}$$

donde $OS$ representa las notas de opinión con las que se ha evaluado el audio.

# CAPÍTULO 5: ANÁLISIS DE RESULTADOS Y CONCLUSIONES

En este capítulo se realiza un análisis comparativo entre el algoritmo de cuantificación uniforme y el no uniforme, para lo cual se contrastan los resultados de las medidas objetivas y subjetivas.

## 5.1. RESULTADOS SUBJETIVOS

De forma general, el *índice de Gini* aumenta conforme se incrementa el número de etapas del algoritmo de Mallat, no obstante, en la Figura 5.1 se muestra que el desempeño del cuantificador no mejora con respecto al aumento del número de etapas ($\xi$), donde en color verde se grafican todos los resultados de las diferentes familias *wavelet* para una etapa, en color azul para dos etapas y en color rojo para tres.

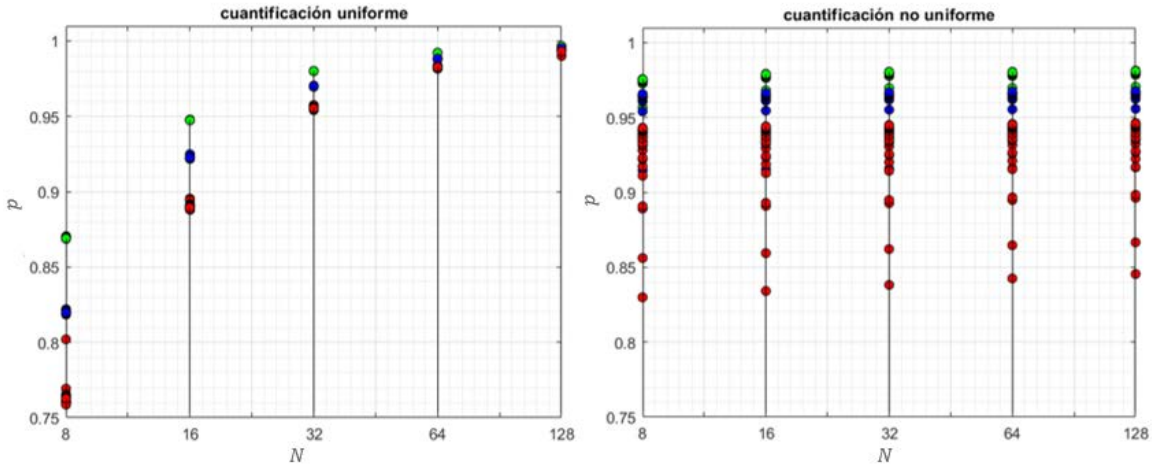


Figura 5.1.Medida objetiva promedio según el nivel de resolución y el cuantificador.
Elaboración propia.

### 4.4.1. Cuantificación Uniforme

En el caso de la cuantificación uniforme, las diferencias obtenidas en los valores de $p$ al variar el número de etapas del MRA se reducen conforme se aumenta el número de niveles de cuantificación, $N$; no obstante, para todos los valores de $N$ el MAR con $\xi = 1$ obtiene mejores resultados, por lo cual, se infiere que el costo adicional relacionado con el aumento en el número de muestras y en el procesamiento asociado con la obtención de los diferentes coeficientes no representa en estos casos ningún beneficio.

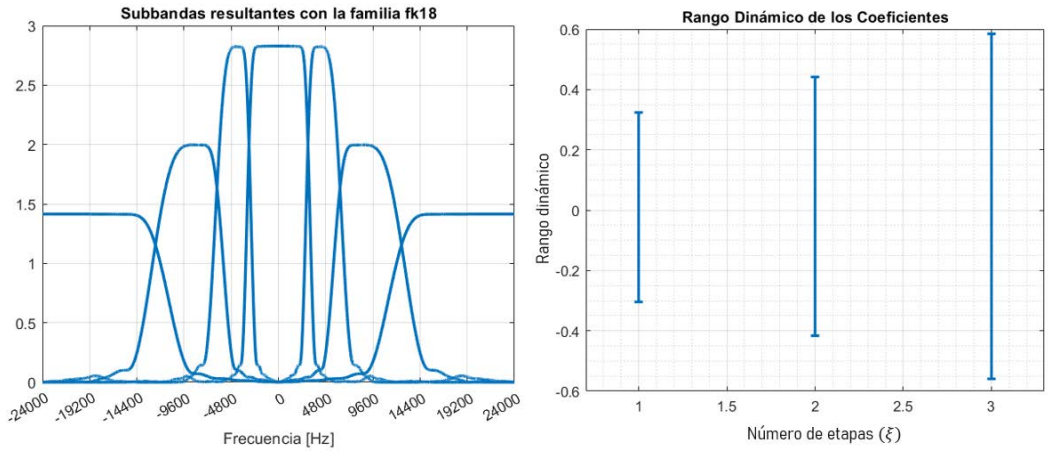


Figura 5.2. Rango dinámico de los coeficientes según el número de etapas.
Elaboración propia.

Más detalladamente, los resultados con respecto a la cuantificación uniforme muestran que aumentar el número de niveles de resolución implica una extensión en el rango dinámico de los coeficientes (ver Figura 5.2 derecha), lo cual significa que los niveles de cuantificación equiespaciados se distribuyen sobre un rango de amplitudes mayor. Lo anterior se corrobora con el aumento en el valor del *índice de Gini*, lo que significa que hay muy pocas muestras que contienen la mayor cantidad de la energía de la señal de voz, dicho de otra forma, que existe una gran desigualdad entre los valores de amplitud de los coeficientes. Adicionalmente, en la parte izquierda de la Figura 5.2 se muestra la diferencia de amplitudes de las subbandas, las cuales aumentan conforme se acercan al origen, i.e., incrementa según el número de etapas necesarias para su obtención.

Para el caso de: $\xi = 3$, $N = 8$ y el cuantificador uniforme, se observa una mayor dispersión en los resultados de las diferentes familias *wavelet* (puntos rojos Figura 5.1). Particularmente, se destaca el punto con una diferencia significativa, el cual tiene un valor de $p = 0.8$ y está asociado a la familia *wavelet* de Haar o Daubechies 1, siendo esta familia la que tiene el menor valor en el *índice de Gini*. Lo anterior corrobora la hipótesis anteriormente planteada, i.e., para el cuantificador diseñado no es beneficioso el aumento en el *índice de Gini.*

En la Figura 5.3 se muestra de manera comparativa las subbandas generadas por las familias *wavelet db1* (menor longitud, $L_f = 2$) y *db45* (mayor longitud, $L_f = 90$), gracias a lo cual se evidencia la relación entre la longitud de los filtros y la selectividad en frecuencia, y se deduce que dicha selectividad favorece el incremento del *índice de Gini*, pues a mayor selectividad, los lóbulos secundarios de las funciones de cada subbanda desaparecen, eliminando el traslape entre éstas, y haciendo que el espectro de la señal de voz se encuentre contenido de forma excluyente en unas pocas subbandas (ver en figura de la derecha).

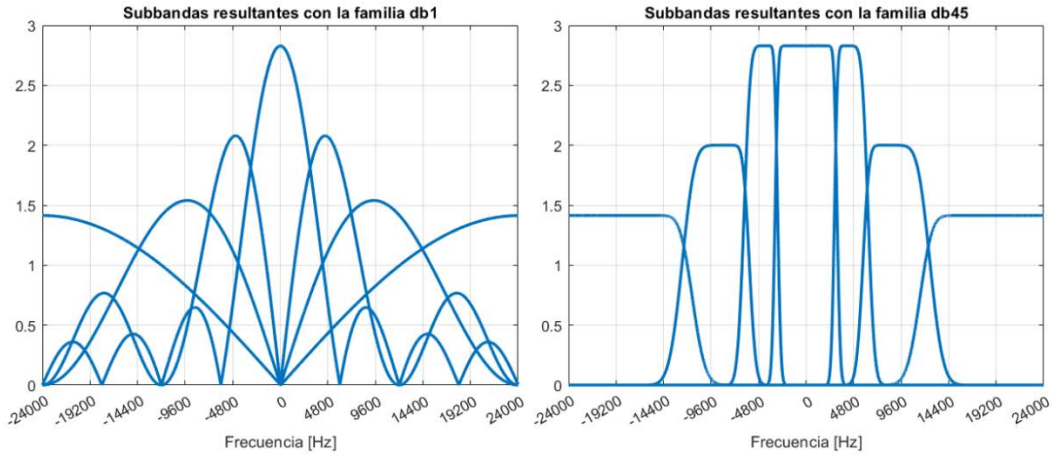


Figura 5.3. Subbandas para familias wavelet con diferente selectividad en frecuencia. Elaboración propia.

Dado que en la DWT las subbandas tienen diferente amplitud, i.e., sus coeficientes tienen diferentes escalas, lo cual aumenta el rango dinámico de los coeficientes concatenados y resulta contraproducente para el cuantificador uniforme, se sugieren dos alternativas:

1. Utilizar el mismo número de niveles de cuantificación en cada uno de los $c_u^k[n]$ conjuntos de coeficientes, con lo cual se tiene información adicional asociada a los descriptores de los diferentes cuantificadores.
2. Utilizar WP (ver Apéndice A.6) para un número de etapas $\xi > 1$, con lo cual las subbandas que dividen el espectro tienen la misma amplitud, sin embargo, su implementación implica un mayor número de muestras a cuantificar, debido a la respuesta transitoria de los filtros.

### 4.4.2. Cuantificación No Uniforme

El algoritmo de cuantificación no uniforme, mostrado en la Figura 4.7 inicia por la subbanda con mayor energía de la señal para calcular el número de niveles de cuantificación que le corresponden. Esta subbanda por lo general es la que abarca las componentes de más baja frecuencia, i.e., es la que está asociada a los coeficientes *scaling*. Dado que el valor mínimo del *índice de Gini* en el dominio *wavelet* es de 0.8, se infiere que la gran desigualdad en la distribución de la energía que muestra este índice implica que un mayor número de niveles de cuantificación se asignan a los coeficientes *scaling*.

En cuanto a la cuantificación no uniforme se observa que, para un bajo número de niveles de cuantificación, como 8 y 16, existe una mejora significativa con respecto a los resultados obtenidos con el cuantificador uniforme, no obstante, no se evidencia una diferencia significativa conforme se incrementa el número de niveles de cuantificación. En la Tabla 5.1 se muestra, a manera de ejemplo, los resultados obtenidos en el M-NRMSE para los diferentes niveles de cuantificación $N = \{8, 16, 32, 64, 128\}$ y los diferentes valores del número de etapas $\xi = \{1, 2, 3\}$. Los

resultados en color lila corresponden a la reconstrucción de la señal únicamente a partir de sus coeficientes *scaling* sin cuantificar.

Tabla 5.1. M-NRMSE para la familia *wavelet coif 3* y cuantificación no uniforme. Elaboración propia.

| $\xi$ | $N$ | | | | |
|---|---|---|---|---|---|
| | 8 | 16 | 32 | 64 | 128 |
| 1 | 0.9588 | 0.9633 | 0.9649 | 0.9648 | 0.9658 |
| | 0.96331 | | | | |
| 2 | 0.9451 | 0.9455 | 0.9459 | 0.9464 | 0.9468 |
| | 0.94389 | | | | |
| 3 | 0.9141 | 0.9149 | 0.9159 | 0.9167 | 0.9173 |
| | 0.91085 | | | | |

En la Figura 5.1 se observa que para el caso de la cuantificación no uniforme existe una mayor varianza según la familia *wavelet*, esto se debe a que en este caso la asignación de los niveles de cuantificación se realiza en función de la distribución de la energía, así, las familias *wavelet* con mayores valores en el *índice de Gini obtienen* también los mejores resultados.

Hasta el momento no se tienen indicios o resultados que justifiquen el uso de familias *wavelet* de alto orden y un número de etapas mayor a uno, dado que, tanto para la cuantificación uniforme como para la cuantificación no uniforme, no se obtiene una mejora significativa asociada a estos cambios, los cuales, por el contrario, sí tienen asociado un incremento en el número de muestras a cuantificar. En la Tabla 5.2 se ejemplifica lo anterior, donde $\xi$ representa el número de etapas, $L_f$ la longitud de los filtros y $n_c$ el número de coeficientes que se deben cuantificar.

En la Tabla 5.2 se encuentran resaltados en color verde los mejores resultados de dos secuencias de coeficientes cuantificados con longitud, $L_c$, de la secuencia binaria equiparable, la cual se calcula utilizando (4.11), y permite evidenciar que, a partir de los resultados objetivos, lo más conveniente es trabajar con un $\xi = 1$ y familias *wavelet* de bajo orden.

El análisis objetivo realizado sobre los dos tipos de cuantificadores es consistente para las diferentes medidas consideradas (ver Apéndice F). No obstante, estos resultados no son suficientes para evaluar y determinar la viabilidad de realizar el proceso de cuantificación de las señales de voz en el dominio *wavelet*. Por lo anterior, en la siguiente sección se realiza una comparación de los resultados objetivos y los subjetivos aplicando la MOS.

Tabla 5.2. Resultados comparativos entre $\boldsymbol{\xi}$ y $\boldsymbol{L_c}$.
Elaboración propia.

| Familia wavelet | $\xi$ | $L_f$ | $L_c$ | $Gini$ | Cuantificador | $p$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | $N=8$ | $N=16$ | $N=32$ | $N=64$ | $N=128$ |
| db3 | 1 | 6 | 1030 | 0.807 | Uniforme | 0.869 | 0.948 | 0.980 | 0.992 | 0.997 |
| | | | | | No uniforme | 0.975 | 0.978 | 0.980 | 0.980 | 0.981 |
| | 3 | | 1041 | 0.936 | Uniforme | 0.764 | 0.890 | 0.956 | 0.983 | 0.993 |
| | | | | | No uniforme | 0.911 | 0.913 | 0.914 | 0.916 | 0.917 |
| db42 | 1 | 84 | 1108 | 0.821 | Uniforme | 0.869 | 0.948 | 0.980 | 0.992 | 0.997 |
| | | | | | No uniforme | 0.976 | 0.979 | 0.981 | 0.981 | 0.981 |
| | 3 | | 1275 | 0.952 | Uniforme | 0.759 | 0.888 | 0.954 | 0.982 | 0.993 |
| | | | | | No uniforme | 0.943 | 0.944 | 0.945 | 0.946 | 0.946 |

En el apéndice F se muestran en detalle los resultados promedio de las diferentes medidas objetivas consideradas, tanto para el cuantificador uniforme, como para el cuantificador no uniforme. En ambos casos se considera el número de niveles de cuantificación $N = \{8, 16, 32, 64, 128\}$.

## 5.2. RESULTADOS SUBJETIVOS

Es importante tener en cuenta que la evaluación que los oyentes realizan sobre un audio puede estar sesgada por factores como el tono de voz o el acento de una persona, por lo que las diferentes pruebas que se realizan dentro de este capítulo utilizan una misma señal de voz para evaluar las variaciones consideradas, i.e., los resultados son relativos de cada sección y no se puede realizar una comparación general.

En las pruebas subjetivas participaron 27 personas, las cuales realizaron su evaluación haciendo uso del entorno de simulación MATLAB®, procurando que no existiera ningún procesamiento adicional sobre los audios a evaluar.

En las pruebas subjetivas no se pueden considerar todos los casos planteados, ya que si son muy extensas se corre el riesgo de que los oyentes se cansen y dejen de evaluar concienzudamente los audios. Por lo anterior, se plantean 6 pruebas, cada una de las cuales está compuesta por 3 variaciones de un mismo audio.

Las dos primeras pruebas buscan determinar si se percibe una diferencia al variar el tipo de familia *wavelet* y el número de etapas, para lo cual se utiliza el caso con mejores resultados objetivos, i.e., cuando se tiene una cuantificación uniforme, 1 etapa del algoritmo de Mallat y 128 niveles de cuantificación.

Las pruebas 3, 4 y 5 buscan evaluar comparativamente los resultados obtenidos con la cuantificación en el dominio *wavelet* uniforme y no uniforme, y la

cuantificación uniforme en el dominio del tiempo. Para el caso de la cuantificación en el dominio *wavelet* se utiliza la familia *wavelet db3* y una etapa del algoritmo de Mallat. En cada una de las pruebas se modifica el número de niveles de cuantificación, considerando los casos de $N = \{8, 32, 128\}$.

Finalmente, en la prueba 6 se comparan los resultados obtenidos con los algoritmos propuestos frente a la señal original, la cual se encuentra digitalizada con 256 niveles. Para este caso se utiliza nuevamente la familia *wavelet db3,* una etapa del algoritmo de Mallat y 128 niveles de cuantificación.

### 5.2.1. Variación del Tipo de Familia *Wavelet*

En la parte izquierda de la Figura 5.4 se muestra el valor de $p$ para las diferentes familias *wavelet*, las cuales se diferencian según la longitud de sus filtros $L_f$. Para las pruebas subjetivas (MOS) se comparan los resultados obtenidos con las familias *wavelet db1* ($L_f = 2$), *fk18* ($L_f = 18$) y *db42* ($L_f = 84$), los cuales se muestran en la parte derecha de la Figura 5.4.

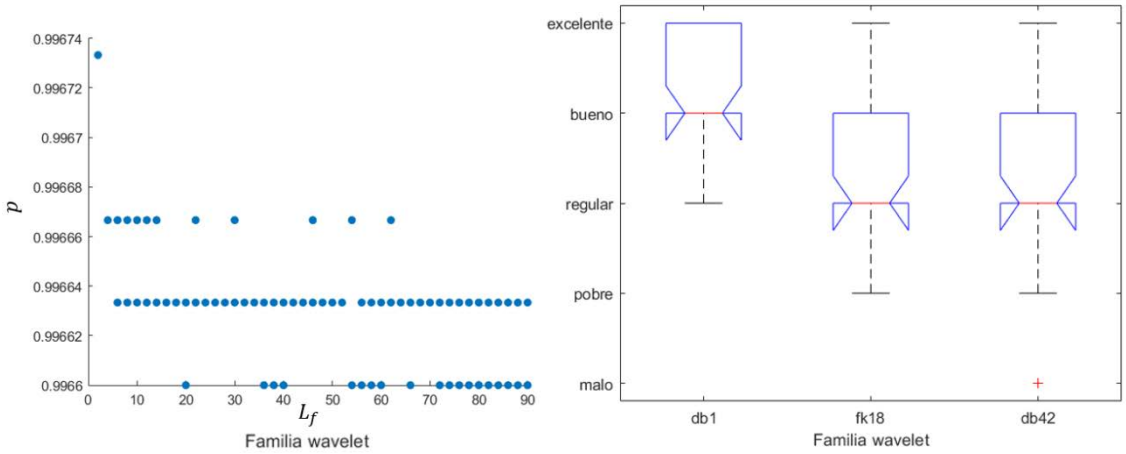


Figura 5.4. Variación del tipo de familia *wavelet*.
Elaboración propia.

Las variaciones en los resultados objetivos mostrados en la Figura 5.4 son del orden de las diezmilésimas, no obstante, el mejor resultado corresponde a la familia *wavelet db1* y los resultados más pobres se concentran en las familias con los filtros de mayor longitud. Las familias *wavelet* con un mayor número de coeficientes tienen buenas valoraciones, no obstante, existe un menor consenso sobre la calidad de estos audios, lo que reduce su valor medio.

### 5.2.2. Variación del Número de Etapas

Para el análisis del efecto de variar el número de etapas se utiliza la familia *wavelet db3.* En la Figura 5.5 se muestran los resultados obtenidos en las MOS para cada caso. Los resultados subjetivos presentan una diferencia significativa entre sus

valores medios para los tres casos, puesto que las ranuras de los diagramas de cajas no se traslapan entre sí. Por otro lado, se observa que existe congruencia entre los resultados subjetivos y objetivos, reafirmando así la deducción de que aumentar el número de etapas del algoritmo de Mallat no conduce a una mejora en la calidad de la señal de voz reconstruida.

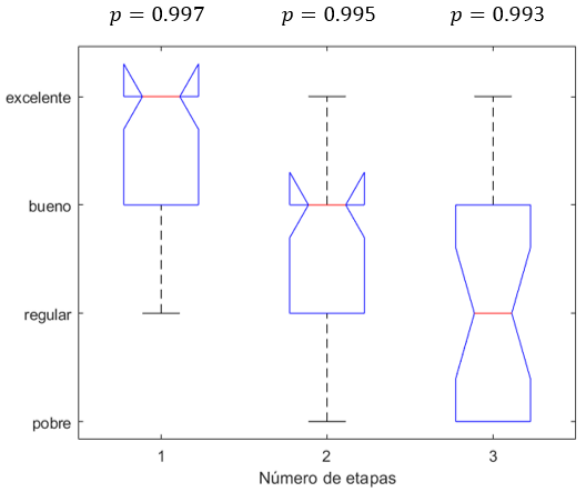


Figura 5.5. Variación del número de etapas.
Elaboración propia.

### 5.2.3. Variación del Número de Niveles de Cuantificación

Dado que para el análisis de la variación en el número de niveles de cuantificación se considera una cuantificación uniforme en el dominio del tiempo, en la Tabla 5.3 se consignan los valores de $p$ para la cuantificación temporal al variar el número de niveles de cuantificación. En el Apéndice F se encuentran detallados los resultados obtenidos para este cuantificador con las diferentes medidas.

Los resultados de la Tabla 5.3 son similares a los obtenidos con la familia *wavelet db3* y el algoritmo de Mallat con una etapa de profundidad, aunque los de la cuantificación temporal se encuentran ligeramente por encima para todos los valores de $N$.

Tabla 5.3. Resultados objetivos cuantificación temporal.
Elaboración propia.

| | $N$ | | | | |
|---|---|---|---|---|---|
| | 8 | 16 | 32 | 64 | 128 |
| $p$ | 0.915 | 0.966 | 0.987 | 0.995 | 0.998 |

**$N = 8$**

En la Figura 5.6 se observa que para $N = 8$, la cuantificación no uniforme en el dominio *wavelet* (NU-DWT) permite obtener resultados de gran calidad a pesar del reducido número de niveles de cuantificación, mientras que con la cuantificación

uniforme, tanto en el dominio del tiempo (U-t) como en el dominio *wavelet* (U-DWT), se obtienen resultados cuyos valores medios corresponden a la categoría de 'malo'.

En el caso de la cuantificación temporal se tiene un menor consenso al momento de evaluar el audio y las medidas objetivas lo sitúan aproximádamente en el punto medio de los dos cuantificadores en el dominio *wavelet*, no obstante, al analizar los valores de media y niveles de significancia (ranuras de los diagramas de cajas) se tiene que no existe una diferencia significativa entre los dos cuantificadores uniformes, mientras que sí existe una diferencia significativa con respecto al cuantificador no uniforme.

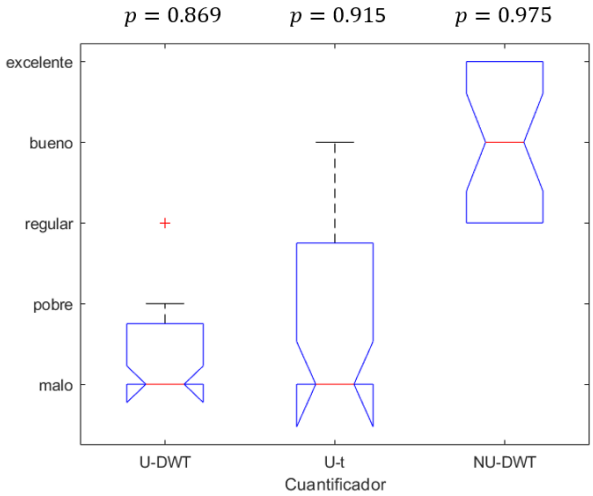


Figura 5.6. Comparación con 8 niveles de cuantificación.
Elaboración propia.

### $N = 32$

Los resultados obtenidos por los tres cuantificadores para $N = 32$ se sintetizan en la Figura 5.8, en la cual se observa que los resultados objetivos no coinciden con los subjetivos, puesto que la cuantificación uniforme en el dominio del tiempo es la que obtiene el resultado objetivo más alto y las dos cuantificaciones en el dominio *wavelet* tienen el mismo valor, no obstante, la valoración subjetiva muestra que la cuantificación no uniforme en el dominio *wavelet* es la que mejores resultados arroja, enfatizando la superioridad de este algoritmo de cuantificación.

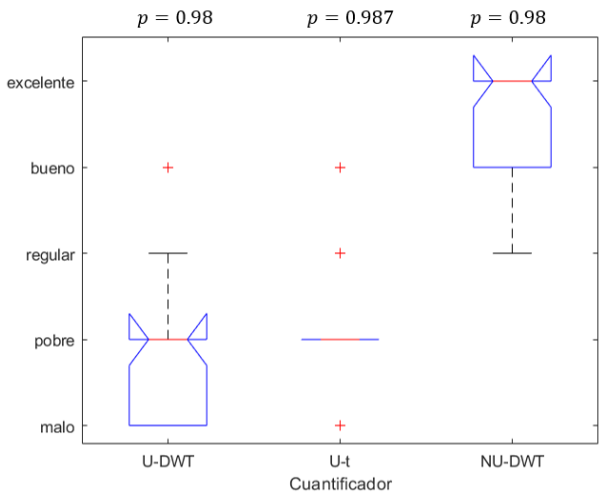


Figura 5.7. Comparación con 32 niveles de cuantificación.
Elaboración propia.

### $N = 128$

Para el caso de $N = 128$, según los resultados objetivos, la cuantificación no uniforme en el dominio *wavelet* es la que realiza una reconstrucción menos fiel de las señales de voz, sin embargo, continúa con la tendencia de ser el algoritmo con mejores resultados según la evaluación subjetiva (ver Figura 5.9).

La cuantificación uniforme en el dominio del tiempo es la que presenta mayor variabilidad en sus datos, pero su desempeño sigue sin tener una diferencia significativa con la cuantificación uniforme en el dominio *wavelet*.

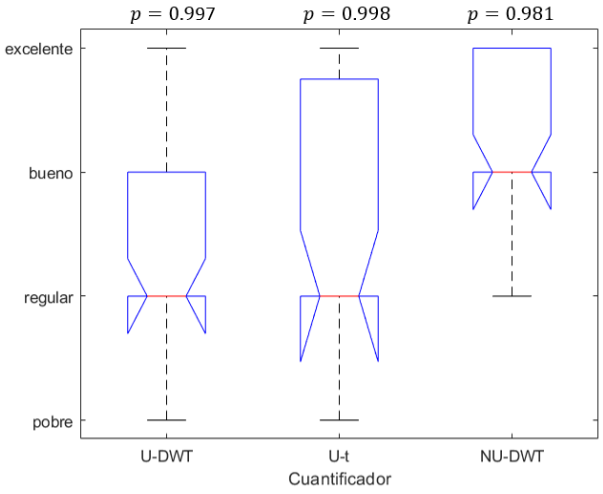


Figura 5.8. Comparación con 128 niveles de cuantificación.
Elaboración propia.

### 5.2.4. Comparación con la Señal Original

Finalmente, se comparan los dos algoritmos de cuantificación propuestos con respecto a la señal de voz original ($t$). Los resultados obtenidos, mostrados en la Figura 5.10, indican que, no existe una diferencia significativa entre los valores medios de las evaluaciones realizada sobre la señal de voz original y la que ha sido cuantificada con un cuantificador no uniforme en el dominio *wavelet*, aunque en este último caso se tiene un mayor consenso con respecto a dicha evaluación. Es importante recalcar que la señal de voz original, sobre la cual se realiza el proceso de cuantificación, es una señal digitalizada con 256 posibles niveles de amplitud, por lo cual es posible que exista un mayor acuerdo al evaluar la calidad de la señal cuantificada con el algoritmo UN-DWT.

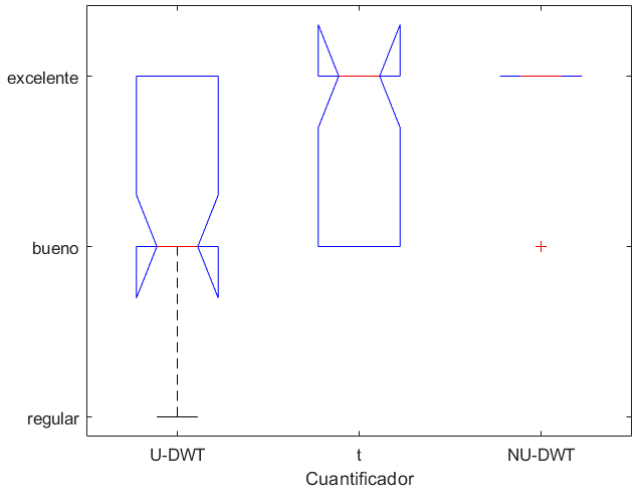


Figura 5.9. Comparación con la señal de voz original.
Elaboración propia.

A partir de las comparaciones realizadas entre los diferentes algoritmos, se evidencia una tendencia que remarca la superioridad del algoritmo de cuantificación no uniforme en el dominio *wavelet*, a pesar de no concordar con los resultados objetivos.

La cuantificación no uniforme prioriza la información contenida en las subbandas con mayor porcentaje de energía. Este enfoque de distribución de los recursos disponibles tiene afinidad con la forma en la que el oído humano capta los sonidos, dado que en éste se definen también bandas críticas (subbandas) y se presentan efectos de enmascaramiento, los cuales pueden ayudar a disimular la distorsión introducida en las subbandas con menores niveles de cuantificación.

## 5.3. CONCLUSIONES

Las familias *wavelet* consideradas en este trabajo de grado son ortogonales, pero su variación no tiene un impacto significativo sobre los resultados objetivos de los algoritmos de cuantificación propuestos, sin embargo; los resultados muestran que no es conveniente usar familias con filtros de longitud elevada, más aún, sí tienen asociado un incremento en el número de coeficientes que se deben cuantificar.

El aumento en profundidad dado por el número de etapas del algoritmo de Mallat permite tener una mayor división del rango de frecuencias determinado por la frecuencia de muestreo. Dado que las subbandas resultantes tienen diferentes valores de ganancia, se obtiene una relación directamente proporcional entre el número de etapas y el *índice de Gini*. No obstante, para ninguno de los algoritmos propuestos resulta favorable el incremento del número de etapas, puesto que en el caso de la cuantificación uniforme implica un aumento en el rango dinámico de los valores que se deben cuantificar, y en el caso de la cuantificación no uniforme se traduce en una mayor distorsión sobre los coeficientes *wavelet* de cada etapa.

La cuantificación uniforme en el dominio del tiempo tiene un comportamiento muy similar a la cuantificación uniforme en el dominio *wavelet*, por lo que muestra que esta última no es una forma eficiente de realizar la cuantificación de los coeficientes resultantes del MRA. Con el fin de aprovechar correctamente las características de la señal en el dominio transformado, buscando reducir la distorsión percibida en la señal reconstruida, es necesario explotar la capacidad de compresión de la DWT, tal como lo hace la cuantificación no uniforme, al centrar la mayor parte de los niveles de cuantificación sobre los coeficientes *scaling*.

## 5.4. TRABAJOS FUTUROS

El presente trabajo de grado de maestría evaluó la conveniencia de implementar un cuantificador de señales de voz basado en la DWT, para lo cual se diseñaron diferentes algoritmos de cuantificación. No obstante, es necesario evaluar comparativamente los algoritmos propuestos con alternativas que se encuentran planteadas en la literatura, por lo anterior se proponen los siguientes trabajos futuros:

- Analizar comparativamente el algoritmo de cuantificación no uniforme utilizando *wavelets* con respecto a la cuantificación basada en la transformada discreta de coseno modificada.
- Analizar comparativamente los algoritmos de cuantificación propuestos con respecto a la cuantificación utilizando WP.

Adicionalmente, es importante validar el correcto funcionamiento de los algoritmos propuestos en escenarios más amplios o al variar los parámetros iniciales, por lo cual se proponen los siguientes trabajos futuros:

- Evaluar el desempeño de los algoritmos propuestos utilizando bases de datos con señales de voz en diferentes idiomas.
- Analizar el efecto de variar la frecuencia de muestreo sobre el desempeño de los algoritmos propuestos.
- Analizar el efecto de variar la duración de las tramas de las señales de voz sobre el desempeño de los algoritmos propuestos.
- Analizar el efecto al considerar un traslape entre las tramas de las señales de voz a cuantificar con los algoritmos propuestos.
- Analizar la conveniencia de utilizar familias *wavelet* biortogonales con los algoritmos propuestos.

# REFERENCIAS


Akhaee, M. A., Kalantari, N. K., & Marvasti, F. (2009). Robust Multiplicative Audio and Speech Watermarking Using Statistical Modeling. *IEEE International Conference on Communications*. https://doi.org/10.1109/ICC.2009.5199424

Alberta Education. (2009). *Module 1*. http://moodle2.rockyview.ab.ca/mod/book/tool/print/index.php?id=52012&chapterid=25630

Alcántara, J. I., Holube, I., & Moore, B. C. J. (2005). Effects of phase and level on vowel identification: Data and predictions based on a nonlinear basilar-membrane model. *The Journal of the Acoustical Society of America*, *100*(4), 2382–2392. https://doi.org/10.1121/1.417948

Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology*, *37*(5), 715–727. https://doi.org/10.1037/0022-3514.37.5.715

*Articulatory Phonetics Speech Sound Form*. (n.d.). Retrieved August 22, 2019, from http://www.ablongman.com/html/productinfo/bauman3e/020554925X_ch02.pdf

Balaji, S., & Murugaiyan, S. M. (2012). Waterfall vs V-Model vs Agile : A Comparative Study on SDLC. *International Journal of Information Technology and Business Management*, *2*(1), 26–30.

Bennett, W. R. (1948). Spectra of Quantized Signals. *Bell System Technical Journal*, *27*(3), 446–472. https://doi.org/10.1002/j.1538-7305.1948.tb01340.x

Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(1), 97–103. https://doi.org/10.1037/0096-1523.32.1.97

Berg, R. E. (1982). The Physics of Sound. *American Journal of Physics*, *50*(10), 953. https://doi.org/10.1119/1.12960

Bousselmi, S., & Ouni, K. (2017a). Study on Speech Reconstruction Stability Using Tight Framelet Packet Transform. *2017 14th International Multi-Conference on Systems, Signals and Devices, SSD 2017*, *2017-Janua*(3), 601–605. https://doi.org/10.1109/SSD.2017.8167007

Bousselmi, S., & Ouni, K. (2017b). The Comparison of Time-Frequency Analysis Methods for Speech Coding Application. *2016 7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications, SETIT 2016*, *0*(5), 427–431. https://doi.org/10.1109/SETIT.2016.7939908

Bracewell, R. N. (2000). The Discrete Hartley Transform. In *The Fourier Transform ans Its Applications* (3rd ed., pp. 293–322). McGraw-Hill. http://www.sciencedirect.com/science/article/pii/S0376736110057092

Breitenstein, C., Van Lancker, D., & Daum, I. (2001). The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample. *Cognition and Emotion*, *15*(1), 57–79. https://doi.org/10.1080/02699930126095

Broesch, T., & Bryant, G. A. (2018). Fathers' Infant-Directed Speech in a Small-Scale Society. *Child Development*, *89*(2), e29–e41. https://doi.org/10.1111/cdev.12768

Burkhardt, F., & Sendlmeier, W. F. (n.d.). Verification of Acoustical Correlates of Emotional Speech using Formant- Synthesis. *Analysis*.

Burrus, C. S., Gopinath, R. A., Guo, H., Odegard, J. E., & Selesnick, I. W. (1998). *Introduction to Wavelets and Wavelet Transforms*. Prentice Hall. https://books.google.com.co/books/about/Introduction_to_Wavelets_and_Wavelet_Tra.html?id=4DgZAQAAIAAJ&redir_esc=y

Cai, H., Sun, J., & Ou, S. (2007). Blind Speech Separation Employing Laplacian Normal Mixture Distribution Model. *2007 International Conference on Mechatronics and Automation*, 3185–3189. https://doi.org/10.1109/ICMA.2007.4304071

Chen, L., Lin, Y. C., Hustad, K. C., & Kent, R. D. (2016). Perceptual speech intelligibility and speech production variability in Mandarin-speaking children with cerebral palsy. *The Journal of the Acoustical Society of America*, *139*(4), 2045–2045. https://doi.org/10.1121/1.4950051

Cover, T. M., & Thomas, J. A. (1991). *Elements of Information Theory* (D. Schilling (Ed.); City Colle). john wiley & sons, inc. https://doi.org/10.1177/0022219410375001

Cutler, C. (1950). *Differential Quantization of Communication Signals* (Patent No. 2605361).

Daubechies, I. (1994a). *Ten Lectures on Wavelets* . the society for industrial and applied mathematics. https://books.google.com.co/books/about/Ten_Lectures_on_Wavelets.html?id=9t5SG06AiT0C&redir_esc=y

Daubechies, I. (1994b). *Where do wavelets come from? - A personal point of view*. http://perso.ens-lyon.fr/paulo.goncalves/pub/WaveletsHistoryByDaubechies.pdf

Deza, M. M., & Deza, E. (2009). Encyclopedia of distances. In *Encyclopedia of Distances*. https://doi.org/10.1007/978-3-642-00234-2

Dudley, H. (1940). the carrier nature of speech. *Bell System Technical Journal*, *XIX*, 495–515.

Eldar, Y. C. (2014). *Sampling Theory*. Cambridge University Press. https://doi.org/10.1017/CBO9780511762321

Fant, G. (n.d.). *Acoustic theory of speech production : with calculations based on X-ray studies of Russian articulations*.

Fraccaro, P. J., O'Connor, J. J. M., Re, D. E., Jones, B. C., DeBruine, L. M., & Feinberg, D. R. (2013). Faking it: Deliberately altered voice pitch and vocal attractiveness. *Animal Behaviour*, *85*(1), 127–136. https://doi.org/10.1016/j.anbehav.2012.10.016

Gallager, R. (2008). Chapter 3. Quantization. In *Principles of Digital Communication*.

Gazor, S., & Zhang, W. (2003). Speech probability distribution. *IEEE Signal Processing Letters*, *10*(7), 204–207. https://doi.org/10.1109/LSP.2003.813679

Gick, B., Schellenberg, M., Stavness, I., & Taylor, R. C. (n.d.). *Articulatory Phonetics*. Retrieved August 23, 2019, from https://linguistics.sites.olt.ubc.ca/files/2019/01/GickalRoutledgeHbk19preprintArticulatoryPhonetics.pdf

Gonzáles, G. (2019). *Partes del aparato fonador*. Educaplay. https://www.educaplay.com/en/learningresources/1704959/print/partes_del_ap

arato_fonador.htm
Gray, R. M. (1995). *Quantization Noise in A / D Converters*. 1–36.
Hanzo, L., A. Somerville, F. C., & Woodward, J. P. (2001). *Voice Compression and Communications: Principles and Applications for Fixed and Wireless Channels* . Wiley. https://www.wiley.com/en-us/Voice+Compression+and+Communications%3A+Principles+and+Applications+for+Fixed+and+Wireless+Channels-p-9780471150398
Hanzo, L., Somerville, F. C., & Woodard, J. (2007). *Voice and Audio Compression for Wireless Communications* (2nd ed.). john wiley & sons, inc. https://doi.org/10.1002/9780470516034
Heiberger, R. M., & Holland, B. (20015). Statistical Analysis and Data Display. In *Journal of the Royal Statistical Society: Series A (Statistics in Society)* (2nd ed., Vol. 168, Issue 2). Springer. https://doi.org/10.1111/j.1467-985x.2005.358_6.x
Hu, Y., & Loizou, P. C. (2008). Evaluation of objective quality measures for speech enhancement. *IEEE Transactions on Audio, Speech and Language Processing*, *16*(1), 229–238. https://doi.org/10.1109/TASL.2007.911054
Huffman, M. K. (2016). *Articulatory Phonetics* (Vol. 1). Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.55
ITU. (1990). Recommendation G.726. In *ccitt*.
ITU. (1996). Recommendation ITU-T P.800 Methods for Subjective Determination of Transmission Quality. In *International Telecommunication Union* (Vol. 800).
ITU. (2017). Recommendation ITU-T P,10/G.100 Vocabulary for Performance, Quality of Service and Quality of Experience. In *International Telecommunication Union* (Issue P.10/G.100).
Janska, A. C., & Clark, R. A. J. (2010). Native and non-native speaker judgements on the quality of synthesized speech. *Proceedings of the 11th Annual Conference of the International Speech Communication Association, INTERSPEECH 2010*, *july 2009*, 1121–1124. https://doi.org/10.21437/interspeech.2010-356
Jensen, J. H., Christensen, M. G., Ellis, D. P. W., Member, S., & Jensen, S. H. (2009). Quantitative Analysis of a Common Audio Similarity Measure. *IEEE Transactions on Audio, Speech and Language Processing*, *17*(4), 693–703.
Jiang, B., & Yang, J. (2011). Preferred frame length for the short-time magnitude spectrum on speech intelligibility and speech quality. *ICICS 2011 - 8th International Conference on Information, Communications and Signal Processing*. https://doi.org/10.1109/ICICS.2011.6174266
Jiang, W., Wang, J., Zhao, Y., Liu, B., & Ji, X. (2013). Multi-Channel Audio Compression Method Based on ITU-T G.719 Codec. *Proceedings - 2013 9th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IIH-MSP 2013*, 293–297. https://doi.org/10.1109/IIH-MSP.2013.81
Joseph, S. M., & Babu Anto, P. (2012). Speech Coding Based on Orthogonal and Biorthogonal Wavelet. *Procedia Technology*, *6*, 397–404. https://doi.org/10.1016/j.protcy.2012.10.047
Jukic, A., & Doclo, S. (2014). Speech dereverberation using weighted prediction error with Laplacian model of the desired signal. *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5172–

5176. https://doi.org/10.1109/ICASSP.2014.6854589

Kadambe, S. (1992). On the window selection and the cross terms that exist in the magnitude squared distribution of the short time Fourier transform. *1992 IEEE 6th SP Workshop on Statistical Signal and Array Processing, SSAP 1992 - Conference Proceedings*, 22–25. https://doi.org/10.1109/SSAP.1992.246839

Kandadai, S., Hardin, J., & Creusere, C. D. (2008). Audio Quality Assessment Using The Mean Structural Similarity Measure. *ICASSP IEEE*, *1*, 221–224.

Kaya, M., & Arioz, U. (2015). Feature weighting with Laplacian score. *2015 23nd Signal Processing and Communications Applications Conference (SIU)*, 280–283. https://doi.org/10.1109/SIU.2015.7129814

Khetrapal, A. (2019, February 26). *How Does the Ear Work?* News Medical Life Sciences. https://www.news-medical.net/health/How-Does-the-Ear-Work.aspx

Kipnis, A., Eldar, Y. C., & Goldsmith, A. J. (2016). Optimal trade-off between sampling rate and quantization precision in A/D conversion. *2015 53rd Annual Allerton Conference on Communication, Control, and Computing, Allerton 2015*, 1083–1090. https://doi.org/10.1109/ALLERTON.2015.7447129

Kohlrausch, A., & Sander, A. (2005). Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets. *The Journal of the Acoustical Society of America*, *97*(3), 1817–1829. https://doi.org/10.1121/1.413097

Kokkinakis, K., & Nandi, A. K. (2005). Speech modelling based on generalized Gaussian probability density functions. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *I*(3), 381–384. https://doi.org/10.1109/ICASSP.2005.1415130

Kreyszig, E. (2011). *ADVANCED ENGINEERING MATHEMATICS*. www.ieee.org.

Lagrange, M., Badeau, R., & Richard, G. (2010). Robust Similarity Metrics Between Audio Signals Based on Asymmetrical Spectral Envelope Matching. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 405–408. https://doi.org/10.1109/ICASSP.2010.5495783

Lai, W. S., Tseng, C. J., & Ding, J. J. (2013). Improved structural similarity measurement for vocal signals. *Proceedings - IEEE International Symposium on Circuits and Systems*, 301–304. https://doi.org/10.1109/ISCAS.2013.6571842

Laitinen, M. V., Disch, S., & Pulkki, V. (2013). Sensitivity of human hearing to changes in phase spectrum. *AES: Journal of the Audio Engineering Society*, *61*(11), 860–877.

Lee, S., Yildirim, S., Kazemzadeh, A., & Narayanan, S. (2005). An articulatory study of emotional speech production. *Erospeech*.

Lenhardt, M. L. (2003). Ultrasonic Hearing in Humans: Applications for Tinnitus Treatment. *International Tinnitus Journal*, *9*(2), 69–75. https://pdfs.semanticscholar.org/b510/d94c76859e21aebb11d5e561b42d3cebc5fb.pdf

Lenhardt, M., Skellett, R., Wang, P., & Clarke, A. (1991). Human ultrasonic speech perception. *Science*, *253*(5015), 82–85. https://doi.org/https://doi.org/10.1126/science.2063208

Letowski, T. R., & Scharine, A. A. (2017). Correlational Analysis of Speech Intelligibility Tests and Metrics for Speech Transmission. In *US Army Research*

*Laboratory* (Issue December). https://doi.org/10.13140/RG.2.2.26581.93921

Liu, H., Wang, E. Q., Chen, Z., Liu, P., Larson, C. R., & Huang, D. (2010). Effect of tonal native language on voice fundamental frequency responses to pitch feedback perturbations during sustained vocalizations. *The Journal of the Acoustical Society of America*, *128*(6), 3739–3746. https://doi.org/10.1121/1.3500675

Liu, W. M., Jellyman, K. A., Mason, J. S. D., & Evans, N. W. D. (2006). Assessment of objective quality measures for speech intelligibility estimation. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *1*, 1225–1228. https://doi.org/10.1109/icassp.2006.1660248

Lloyd, S. P. (1982). Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137. https://doi.org/10.1109/TIT.1982.1056489

Loizou, P. C. (2011). Speech Quality Assessment. In *Multimedia Analysis, Processing and Communications* (pp. 623–654). Springer. https://doi.org/10.1007/978-3-319-02732-6_5

Loret, F. (2012). *Le fonctionnement de la voix*. https://doi.org/10.7202/1001666ar

Lu, D. Y., & Fan, Q. Bin. (2011). A Class of Tight Framelet Packets. *Czechoslovak Mathematical Journal*, *61*(3), 623–639. https://doi.org/10.1007/s10587-011-0035-9

Mahmmod, B. M., Ramli, A. R., Abdulhussian, S. H., Al-Haddad, S. A. R., & Jassim, W. A. (2017). Low-Distortion MMSE Speech Enhancement Estimator Based on Laplacian Prior. *IEEE Access*, *5*, 9866–9881. https://doi.org/10.1109/ACCESS.2017.2699782

Mallat, Stéphane. (2009). *A Wavelet Tour of Signal Processing* (Vol. 59, Issue 3). Elseiver.

Mallat, Stephane, & Zhong, S. (1992). Characterization of Signals from Multiscale Edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *14*(7), 710–732. https://doi.org/10.1109/34.142909

Marsh, C. (2013). Introduction to Continuous Entropy. In *Princeton University* (p. 17). Princeton University. https://www.crmarsh.com/static/pdf/Charles_Marsh_Continuous_Entropy.pdf

MATLAB. (n.d.). *Choose a Wavelet - MATLAB & Simulink*. Retrieved May 10, 2021, from https://www.mathworks.com/help/wavelet/gs/choose-a-wavelet.html#mw_2e48c894-f8ff-47b9-95f6-b4a08eb95c11

Merino de la Fuente, J. M., & Muñoz-Repiso, L. (2013). La percepción acústica: tono y timbre. *Revista de Ciencias*, *3*, 21–32. https://dialnet.unirioja.es/servlet/articulo?codigo=4458407

Merino, M. J., & Muñoz-Repiso, L. (2013). La percepción acústica : Física de la audición. *Revista de Ciencias*, *2*, 19–26. https://doi.org/10.3892/mmr.2014.2054

Metze, F., Sheikh, Z. A. W., Waibel, A., Gehring, J., Kilgour, K., Nguyen, Q. B., & Nguyen, V. H. (2013). Models of tone for tonal and non-tonal languages. *2013 IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU 2013 - Proceedings*, 261–266. https://doi.org/10.1109/ASRU.2013.6707740

Mowlaee, P., Kulmer, J., Stahl, J., Mayer, F., & Mowlaee, P. (2016). Introduction: Phase Processing, History. *Single Channel Phase-Aware Signal Processing in Speech Communication: Theory and Practice*, 1–32.

https://doi.org/10.1002/9781119238805.ch1

Nyquist, H. (1928). Certain Topics in Telegraph Transmission Theory. *Transactions of the American Institute of Electrical Engineers*, *47*(2), 617–644. https://doi.org/10.1109/5.989875

Ogden, R. T., & Vidakovic, B. (2000). Statistical Modeling by Wavelets. In *Journal of the American Statistical Association* (Vol. 95, Issue 451). https://doi.org/10.2307/2669487

Ogunfunmi, T., Togneri, R., & Narasimha, M. (Sim). (2015). Speech and Audio Processing for Coding, Enhancement and Recognition. In *Edinburgh Journal of Botany* (Issue 1). Springer.

Pal, M., Paul, D., & Saha, G. (2018). Synthetic speech detection using fundamental frequency variation and spectral features. *Computer Speech and Language*, *48*(October), 31–50. https://doi.org/10.1016/j.csl.2017.10.001

Polkosky, M. D., & Lewis, J. R. (2003). Expanding the MOS: Development and psychometric evaluation of the MOS-R and MOS-X. *International Journal of Speech Technology*, *6*(2), 161–182. https://doi.org/10.1023/A:1022390615396

Pumphrey, R. J. (1950). Upper Limit of Frequency for Human Hearing. *Nature*, *166*(4222), 571–571. https://doi.org/10.1038/166571b0

R Nave, M. O. (n.d.). *Vocal Sound Production*. Retrieved January 30, 2022, from http://hyperphysics.phy-astr.gsu.edu/hbasees/Music/voice.html

RAE. (n.d.). *rarefacer | Definición | Diccionario de la lengua española*. Retrieved May 19, 2021, from https://dle.rae.es/rarefacer

Raitio, T., Juvela, L., Suni, A., Vainio, M., & Alku, P. (2015). Phase perception of the glottal excitation and its relevance in statistical parametric speech synthesis. *Speech Communication*, *81*, 104–119. https://doi.org/10.1016/j.specom.2016.01.007

Ramírez Viáfara, J. M., Romo Romero, H. A., & Silva Zambrano, M. M. (2020). *Telecomunicaciones digitales* (U. del Cauca (Ed.); Vol. 1).

Rashidi-nejad, M., & Abutalebi, H. R. (2012). Speech Enhancement Using Adaptive MMSE Estimator Under Signal Presence Uncertainty and Laplacian Prior. *6th International Symposium on Telecommunications (IST)*, 843–847. https://doi.org/10.1109/ISTEL.2012.6483103

Rice, S. O. (1945). Mathematical Analysis of Random Noise. *Bell System Technical Journal*, *24*(1), 46–156. https://doi.org/10.1002/j.1538-7305.1945.tb00453.x

Rich, D. (2015). *Hearing Anatomy of the auditory pathway Hair cells and transduction of sound waves Regional specialization of the cochlea to respond to different frequencies. - ppt download*. SlidePlayer. https://slideplayer.com/slide/4279877/

Rohde&Schwarz. (2012). *Next-Generation (3G/4G) Voice Quality Testing with POLQA ® White Paper*.

Sayood, K. (2006). *Introduction To Data Compression*. Elseiver.

scienceABC. (2017). *Basilar Membrane: What Is It? What Are Its Functions? » Science ABC*. https://www.scienceabc.com/humans/basilar-membrane-what-is-it-and-what-does-it-do.html

Seto, K., & Ogunfunmi, T. (2019). A Scalable Wideband Speech Codec Using the Wavelet Packet Transform Based on the Internet Low Bitrate Codec. *Computer Speech and Language*, *54*, 61–70. https://doi.org/10.1016/j.csl.2018.09.001

Shannon, C. (1948a). A Mathematical Theory of Communication. *The Bell System Technical Journal*, *27*, 379–423. http://www.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf

Shannon, C. (1948b). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*, 379–423.

*sonido | Definición de sonido - Diccionario de la lengua española - Edición del Tricentenario*. (n.d.).

Souza, D. M., Costa, I. A., & Nobrega, R. A. (2017). A study of distance/similarity measurements in the context of signal processing (density estimation). *INSCIT 2017 - 2nd International Symposium on Instrumentation Systems, Circuits and Transducers: Chip on the Sands, Proceedings*. https://doi.org/10.1109/INSCIT.2017.8103517

Spanias, A., Painter, T., & Atti, V. (2007). *Audio Signal Processing and Coding*. Wiley-Interscience.

Streijl, R. C., Winkler, S., & Hands, D. S. (2014). Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives. *Multimedia Systems 2014 22:2*, *22*(2), 213–227. https://doi.org/10.1007/S00530-014-0446-1

*Structural similarity (SSIM) index for measuring image quality - MATLAB ssim*. (n.d.). Retrieved November 9, 2021, from https://www.mathworks.com/help/images/ref/ssim.html

Szczegielniak, A. (n.d.). *Phonetics: The Sounds of Language Introduction to Linguistic Theory*. Retrieved August 21, 2019, from https://scholar.harvard.edu/files/adam/files/phonetics.ppt.pdf

Tinati, M., & Mozaffary, B. (2006). Laplacian Mixture Modeling for Overcomplete Mixture Matrix Estimation in Wavelet Packet Domain by Adaptive EM-type Algorithm. *2006 IEEE Conference on Cybernetics and Intelligent Systems*, 1–5. https://doi.org/10.1109/ICCIS.2006.252352

Vig, R., & Chauhan, S. S. (2018). Speech Compression Using Multi-Resolution Hybrid Wavelet Using DCT and Walsh Transforms. *Procedia Computer Science*, *132*, 1404–1411. https://doi.org/10.1016/j.procs.2018.05.070

Virtanen, T., & Helén, M. (2007). Probabilistic Model Based Similarity Measures For Audio Query-by-Example. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 82–85.

voice foundation. (n.d.). *Understanding Voice Production*. Retrieved December 13, 2018, from https://voicefoundation.org/health-science/voice-disorders/anatomy-physiology-of-voice-production/understanding-voice-production/

Voiers, W. D. (1977). Diagnostic Acceptability Measure for Speech Communication Systems. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *1977-May*, 204–207. https://doi.org/10.1109/ICASSP.1977.1170198

Vošvrda, M., & Schürrer, J. (2015). Wavelet Coefficients Energy Redistribution and Heisenberg Principle of Uncertainty. In *Institute of Information Theory and Automation* (Vol. 2, Issue 1). http://library.utia.cas.cz/separaty/2015/E/vosvrda-0449775.pdf

Wang, J., Zhao, S., & Kuang, J. (2007). *A 3.3 Kbps CWI Speech Coding Algorithm Based On Biorthogonal Wavelet Transform*. 2–4.

Woods, D. L., Yund, E. W., Herron, T. J., & Cruadhlaoich, M. A. I. U. (n.d.). *Consonant identification in consonant-vowel-consonant syllables in speech-spectrum noise*. https://doi.org/10.1121/1.3293005

Zhu, X., Suk, H. Il, Wang, L., Lee, S. W., & Shen, D. (2017). A novel relational regularization feature selection method for joint regression and classification in AD diagnosis. *Medical Image Analysis*, *38*, 205–214. https://doi.org/10.1016/j.media.2015.10.008

Zieliński, S., Hardisty, P., Hummersone, C., & Rumsey, F. (2007). Potential Biases in MUSHRA Listening Tests. *Audio Engineering Society - 123rd Audio Engineering Society Convention 2007*, *2*(August), 652–661.

# APÉNDICE A - COMPLEMENTOS

En el apéndice A se muestran algunas demostraciones matemáticas o aclaraciones complementarias, con el fin de clarificar algunos de los enunciados del documento.

## A.1. Transformación de una Variable Aleatoria Continua a una Variable Aleatoria Uniforme

Sea $X$ una variable aleatoria continua y $Y$ una variable aleatoria igual a una relación funcional de $X$. Dado que $X$ es una variable aleatoria cuyo recorrido puede variar entre menos infinito e infinito, es necesario definir una función que transforme ese recorrido a un recorrido limitado entre 0 y 1.

La CDF de una variable aleatoria acumula las probabilidades de ocurrencia de los valores a lo largo de su recorrido, de esta forma la CDF inicia en $-\infty$ (la acumulación es nula) y toma un valor máximo de 1 en $\infty$ (se ha acumulado el 100%). Adicionalmente, estas funciones son monótonamente crecientes y biyectivas.

$$y = g(x) = F_X(x), \qquad 0 \leq y \leq 1\,. \tag{A1}$$

$$\begin{aligned} F_Y(y) &= \Pr\{Y \leq y\} = \Pr\{F_X(x) \leq y\}, \\ &= \Pr\left\{{F_X}^{-1}\left(F_X(x)\right) \leq {F_X}^{-1}(y)\right\}, \\ &= \Pr\left\{X \leq {F_X}^{-1}(y)\right\}, \\ &= F_X\left({F_X}^{-1}(y)\right), \qquad (A2) \\ &= y, \qquad (A3) \end{aligned}$$

donde $0 \leq y \leq 1$.

La pdf de una variable aleatoria se puede calcular como la derivada de la CDF, para este caso se tiene:

$$f_Y(y) = \frac{d}{dy}[F_Y(y)] = \frac{d}{dy}(y) = 1. \tag{A4}$$

donde $0 \leq y \leq 1$.

Así, $Y$ es uniforme entre 0 y 1.

## A.2. Ruido de Cuantificación DPCM

DPCM no cuantifica directamente las muestras de la señal, sino que realiza el proceso de cuantificación sobre la diferencia entre las muestras. No obstante, el esquema mostrado en la Figura 2.12 se muestra que la diferencia a cuantificar corresponde a la diferencia entre la muestra de la señal y la muestra cuantificada inmediatamente anterior (Sayood, 2006).

Si la diferencia a cuantificar es $d_n = x_n - x_{n-1}$ se tiene entonces que la salida del cuantificador es:

$$d'_n = Q\{d_n\} = d_n - q_n\,, \tag{A5}$$

donde $q_n$ corresponde al ruido de cuantificación, así la señal reconstruida está dada por:

$$\begin{aligned} x'_n &= x'_{n-1} + d'_n, \\ &= x_{n-1} + \sum_{k=1}^{n-1} q_{k-1} + d_n + q_n, \\ &= x_n + \sum_{k=1}^{n} q_{k-1}. \end{aligned} \tag{A6}$$

Con lo que el ruido de cuantificación se acumula conforme se avanza en el proceso de cuantificación. La solución a este problema consiste en que la diferencia a cuantificar sea de la forma $d_n = x_n - x'_{n-1}$, así:

$$x'_n = x_n + q_n. \tag{A7}$$

Para este caso la muestra recuperada $x'_n$ únicamente tiene asociado el ruido de cuantificación correspondiente a su instante de tiempo.

## A.3. Frecuencia de Muestreo y Cuantificación

En los cuantificadores diferenciales, aumentar la frecuencia de muestreo equivale a aumentar la correlación existente entre las mismas, lo cual implica una reducción en la diferencia que existe entre dos muestras consecutivas. Lo anterior lleva a disminuir el rango dinámico de la información a cuantificar, por lo que, para un número fijo de niveles de cuantificación, se reducen las regiones de cuantificación y por lo tanto el error de cuantificación asociado al proceso.

En los cuantificadores que no utilizan una aproximación diferencial, el aumento de la frecuencia de muestreo no representa una mejora evidente. Teóricamente, si la pérdida de información intrínseca al proceso de cuantificación se modela como

ruido, se tiene una relación directamente proporcional entre la frecuencia de muestreo y el desempeño del cuantificador.

El cuantificador no es un sistema lineal, por lo que no es posible caracterizarlo en el dominio de la frecuencia. Algunos trabajos han mostrado que el efecto del cuantificador se puede modelar como un ruido blanco que se adiciona a la señal (Bennett, 1948; Gray, 1995; Rice, 1945). No obstante, otros enfoques muestran que el ruido se puede considerar como blanco únicamente en casos específicos y que una mayor frecuencia de muestreo permite ver que el ruido de cuantificación no es blanco (Kipnis et al., 2016).

## A.4. Desplazamiento Diádico en la Transformada Wavelet Discreta

Una familia *wavelet*, $\psi_{a,b}(t)$, se crea a partir de su *wavelet* madre, $\psi(t)$. En el caso de la WT el proceso se logra realizando variaciones en los parámetros de escala, $a$, y traslación, $b$. La familia *wavelet* se presenta a continuación:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right). \qquad (A8)$$

En la DWT se busca limitar los posibles valores de $a$ y $b$, de tal forma que las variaciones en escala y traslación sean discretas. Por lo anterior, se asume que los desplazamientos de estos dos parámetros se dan en función de una base arbitraria $n$, i.e.,

$$a = n^{-j}$$
$$b = n^{-j}k.$$

Al reemplazar los valores de $a$ y $b$ en definición de la familia *wavelet* (A9) se tiene que:

$$\psi_{j,k}(t) = n^{\frac{j}{2}}\psi\left(\frac{t-n^{-j}k}{n^{-j}}\right). \qquad (A9)$$

Vista desde el dominio de la frecuencia, la familia *wavelet* notada como $\tilde{\psi}_{j,k}(f)$, es de la forma:

$$\tilde{\psi}_{j,k}(f) = n^{-\frac{j}{2}}\tilde{\psi}\left(\frac{f}{n^j}\right)e^{-j2\pi n^{-j}kf}. \qquad (A10)$$

Asumiendo que la *wavelet* madre tiene un ancho de banda arbitrario $w$, y dado que las *wavelets* son funciones pasa banda, al representar gráficamente los espectros ideales (pulsos rectangulares) de dos *wavelets* con etapas consecutivas, se obtiene la Figura A.1.

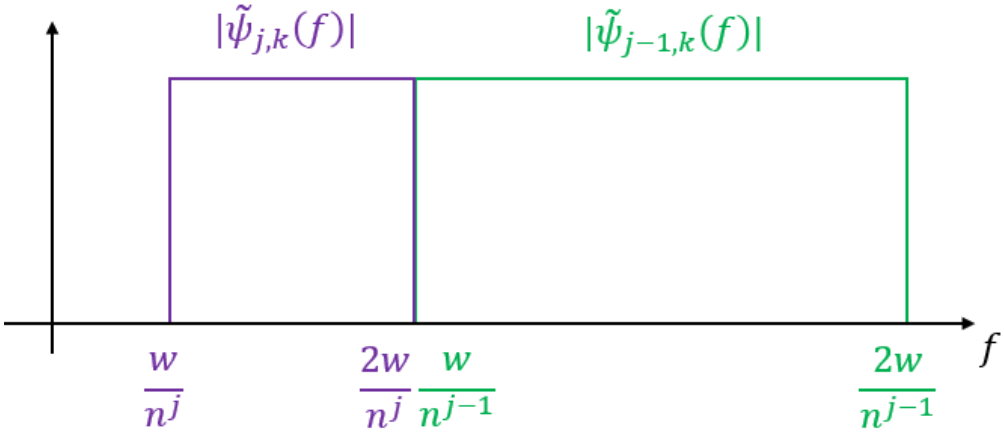


Figura A.1. Subbandas *wavelet* para dos etapas.
Elaboración propia.

Con el fin de que no exista redundancia en la información obtenida a partir de la DWT; es necesario que no exista traslape entre cada una de las subbandas asociadas a las *wavelets* de la familia $\psi_{j,k}(t)$, por lo cual se plantea la siguiente desigualdad:

$$\frac{2w}{n^j} \leq \frac{w}{n^{j-1}},$$

a partir de la cual se obtiene que:

$$n \geq 2,$$

de tal manera que 2 es el desplazamiento mínimo para que no exista traslape entre las subbandas. No obstante, es importante aclarar que un número $n > 2$ implica que las subbandas se encuentran separadas, dejando intervalos de frecuencia sin cubrir (ver Figura A.2), lo cual hace inviables estos valores.

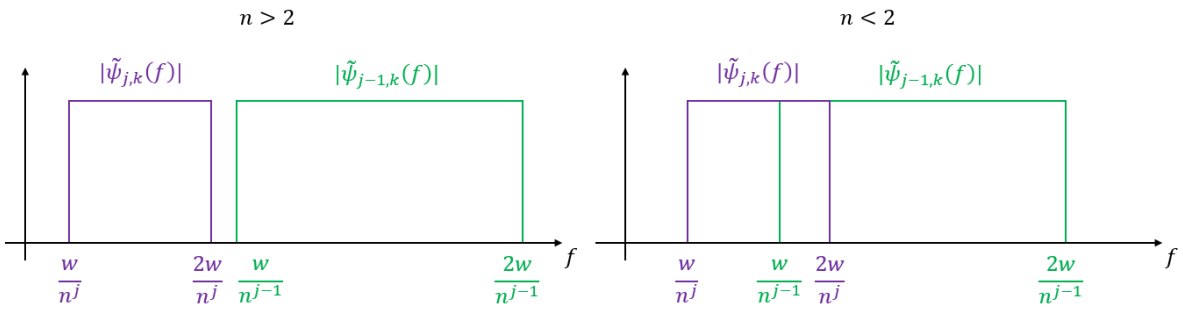


Figura A.2. Subbandas según el desplazamiento en la DWT.
Elaboración propia.

## A.5. Algoritmo de Mallat

El algoritmo de Mallat permite implementar computacionalmente el enfoque del MRA (ver Figura A.3), en el cual los subespacios *scaling* y *wavelet* se complementan para poder representar correctamente a las señales que se encuentren dentro del espacio $\mathcal{L}_2$.

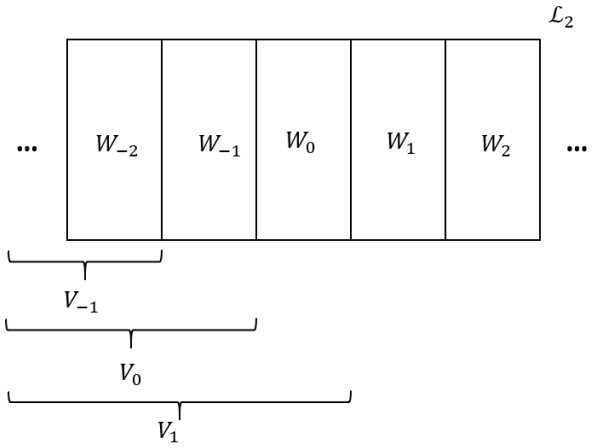


Figura A.3. Subespacios del MRA.
Elaboración propia.

Matemáticamente, se tiene que los subespacios *scaling* y *wavelet* cumplen dos condiciones:

1. El subespacio *scaling* de resolución $j$ está contenido dentro del subespacio *scaling* de resolución $j+1$, i.e.,
$$V_j \subset V_{j+1}.$$

2. El subespacio *scaling* de resolución $j$ está compuesto por la unión de los subespacios *scaling* y *wavelet* de resolución $j-1$, i.e.,
$$V_j = W_{j-1} \cup V_{j-1}.$$

La primer condición implica que $\varphi_{j-1}[n]$ se puede generar como una combinación lineal de $\varphi_j[n]$, esto es:

$$\varphi_{j-1}[n] = \sum_k h[n] 2^{\frac{j}{2}} \varphi\left[2^j n - k\right]. \qquad (A11)$$

Del mismo modo, la segunda condición muestra que $\psi_{j-1}[n]$ también se puede crear como una combinación lineal de $\varphi_j[n]$, esto es:

$$\psi_{j-1}[n] = \sum_k g[n] 2^{\frac{j}{2}} \varphi\left[2^j n - k\right] \qquad (A12)$$

La Figura A.4 muestra el diagrama de bloques de una rama del algoritmo de Mallat, el cual parte de las condiciones anteriores para obtener computacionalmente $\varphi_{j-1}[n]$ a partir de $\varphi_j[n]$.

$$\begin{aligned} y[n] &= h[2n] * \varphi_j[2n] * h[n], \\ &= h[2n] * \varphi_j[n], \\ &= h[n'] * \varphi_j\left[\frac{n'}{2}\right], \\ &= \varphi_{j-1}[n]. \end{aligned} \tag{A13}$$

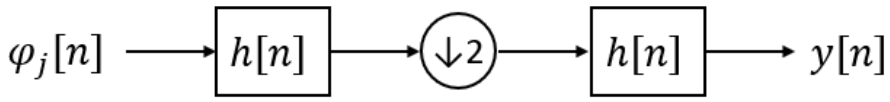


Figura A.4. Sección algoritmo de Mallat.
Elaboración propia.

## A.6. Paquetes *Wavelet* (WP)

Los WP son una variación del algoritmo de Mallat, en la cual los coeficientes resultantes de cada etapa se convolucionan con los filtros asociados a las funciones *wavelet* y *scaling*, y no únicamente los coeficientes *scaling* como en la DWT (ver Figura A.5). Lo anterior implica que para un nivel de resolución $j$ se obtienen $2^j$ conjuntos de coeficientes, en lugar de los $2j - (j-1)$ resultantes con la DWT; por otro lado, las muestras adicionales inherentes a las convoluciones son el contrapeso de tener un mayor número de subbandas[23] con la misma amplitud (ver Figura A.6).

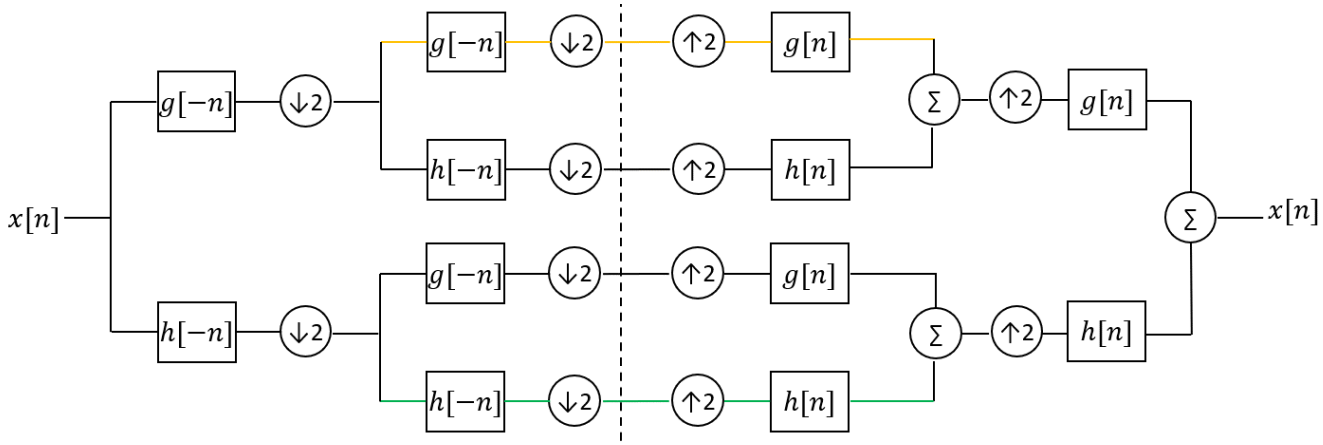


Figura A.5. Diagrama de bloques WP.
Elaboración propia.

[23] Los valores de frecuencia de la Figura A.6 se encuentran normalizados con respecto a la frecuencia de muestreo.

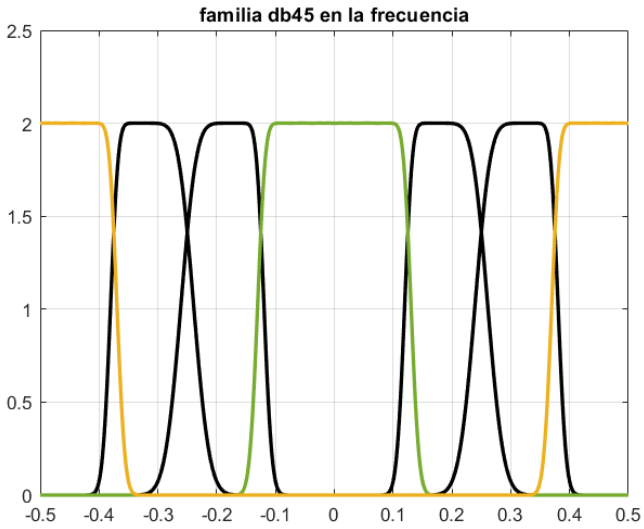


Figura A.6. Subbandas WP.
Elaboración propia.

# APÉNDICE B - MEDIDAS DE DISTORSIÓN

En el apéndice B se muestran algunas medidas para calcular la distorsión sufrida por la señal cuantificada, debida a la pérdida de información inherente al proceso de cuantificación.

## B.1. Desigualdad del Procesamiento de Señales

La desigualdad del procesamiento de señales plantea que, dadas tres variables aleatorias, conectadas por una cadena de Markov de la forma:

$$T \to U \to V, \tag{B1}$$

donde $T$ corresponde a la variable aleatoria original, $U$ es la observación de $T$ y $V$ es una manipulación de $U$, es decir, $V = \mathrm{g}(U)$; no existe ningún tipo de procesamiento que se pueda realizar sobre una variable aleatoria $U$ tal que el resultado $V$ tenga más información sobre $T$ que la observación inicial, esto es:

$$I(T;U) \geq I(T;\mathrm{V}). \tag{B2}$$

(B2) es un resultado teórico de la teoría de la información, el cual muestra que al manipular la variable aleatoria $U$, la variable resultante $V$ contiene la misma o menor información sobre $T$ que $U$ (Cover & Thomas, 1991).

El Error Cuadrático Medio (MSE, *Mean Square Error*) es ampliamente utilizado para medir la distorsión, sin embargo, se debe tener presente que sus valores dependen de la escala y la traslación de $V$ con respecto a $U$ (Cover & Thomas, 1991), (Gallager, 2008).

En el caso del proceso de cuantificación, la relación funcional $\mathrm{g}$ entre $U$ y $V$ es la característica de transferencia del cuantificador, por lo que en esta sección se analizan diferentes métodos para medir la distorsión derivada del procesamiento.

El interrogante radica en determinar una medida adecuada para medir la distorsión que el cuantificador introduce sobre la señal de voz, para lo cual se pueden utilizar medidas objetivas o subjetivas. Las medidas objetivas resultan de operaciones matemáticas que busca determinar la similitud entre dos señales, mientras que las subjetivas utilizan la opinión de diferentes personas para determinar la distorsión percibida.

La respuesta del sentido del oído no se puede modelar matemáticamente, por lo que para determinar la percepción de las personas se hace necesario recurrir a encuestas de opinión, no obstante, éstas también tienen sesgos y variaciones. Es

por esto por lo que la validez del resultado depende del número de personas consultadas.

Dentro de las medidas subjetivas más utilizadas se encuentra la Nota Media de Opinión (MOS, *Mean Opinion Score*) que se incluye en recomendaciones para medir la calidad de la voz y el audio como: la ITU-T P.10 (ITU, 2017) y el ITU-T P.800 (ITU, 1996). La Medida Diagnóstica de Aceptabilidad (DAM, *Diagnostic Acceptability Measure*) considera un mayor número de parámetros y permite esperar resultados más exactos, además, analiza la calidad de los periodos de silencio (Voiers, 1977). El método de Múltiples Estímulos con Referencias Ocultas y Ancladas (MUSHRA, *MUltiple Stimuli with Hidden Reference and Anchor*) tiene una escala de valoración superior a la MOS (0-100) y busca reducir el número de personas, aunque se ha demostrado que se puede presentar un sesgo dependiendo de la distribución de los estímulos (Zieliński et al., 2007).

En cuanto a las medidas objetivas, se tiene un mayor abanico de posibilidades, puesto que existen medidas en el dominio del tiempo tales como: el Índice de Similitud Estructural Promedio (MSSIM, *Mean Structural SIMilarity*) (Kandadai et al., 2008); variaciones de la Relación Señal a Ruido (SNR, *Signal to Noise Ratio*); y modelos probabilísticos (Virtanen & Helén, 2007). Existen otros enfoques que se sustentan en el dominio de la frecuencia como son los Coeficientes Cepstrales en las Frecuencias de Mel (MFCC, *Mel Frequency Cepstral Coefficients*), los cuales buscan determinar similitud de las envolventes espectrales (Lagrange et al., 2010), (Jensen et al., 2009).

*Deza & Deza* presentan diferentes alternativas para el cálculo de la distorsión, dentro de las cuales se enlista una serie de alternativas para el caso particular de señales de audio (Deza & Deza, 2009); no obstante, se debe tener en cuenta que las medidas cuantitativas no son adecuadas para evaluar todos los efectos de las modificaciones introducidas a una señal de voz o audio a partir de los diferentes procesamientos que sobre éstas se apliquen (W. M. Liu et al., 2006), por lo que se utilizan las medidas cualitativas con el fin de corroborar los resultados obtenidos.

La Evaluación Perceptual de la Calidad del Habla (PESQ, *Perceptual Evaluation of Speech Quality*) es una medida objetiva con una alta correlación con el MOS, la cual se considera en la recomendación de la ITU-T P.862 para medir la pérdida de calidad provocada por la distorsión introducida por los *vocoders* (Hu & Loizou, 2008; Loizou, 2011). No obstante, es importante tener en cuenta que este tipo de medidas, así como sus versiones posteriores[24], se desarrollan pensando en medir los efectos adversos de las redes de telecomunicaciones y no como tal para medir la pérdida de información asociada a un procesamiento.

[24] POLQA (*Perceptual Objective Listening Quality Analysis*) se puede considerar como una evolución de PESQ, el cual permite analizar señales de voz en un mayor rango de frecuencias, por lo que incluye el análisis de señales de voz de alta calidad o banda ancha (Rohde&Schwarz, 2012).

## B.2. Medidas Objetivas de Comparación Directa

El MSE busca determinar la diferencia entre dos señales por medio del cálculo del valor esperado de su diferencia al cuadrado, i.e.,

$$MSE = \mathrm{E}[(\boldsymbol{X} - \boldsymbol{Y})^2]. \tag{B3}$$

El MSE es susceptible a los cambios de escala y a los desplazamientos; no obstante, existen algoritmos que buscan subsanar algunas de estas deficiencias, entre los cuales se resalta la Deformación de Tiempo Dinámica (DTW, *Dynamic Time Warping*). DTW es un algoritmo que permite comparar series de tiempo que no necesariamente tienen la misma escala temporal o están desplazadas.

Suponiendo que no existe desplazamiento entre las señales que se van a comparar y que éstas se encuentran normalizadas, se define una variación para normalizar los resultados (Lai et al., 2013):

$$M - NRMSE = 1 - \frac{1}{2} \cdot \sqrt{\frac{\mathrm{E}[(\boldsymbol{X} - \boldsymbol{Y})^2]}{\mathrm{E}[\boldsymbol{X}^2]}}. \tag{B4}$$

Otras alternativas para medir la diferencia entre dos señales se basan en calcular el valor esperado del valor absoluto de su resta, en lugar de elevarla al cuadrado, como en (B3) y (B4) (Souza et al., 2017):

$$MAE = \mathrm{E}[|\boldsymbol{X} - \boldsymbol{Y}|]. \tag{B5}$$

Al considerar la resta entre las dos señales como ruido, se utiliza (B6) para determinar la Relación Señal a Ruido ($SNR$) como una medida de distorsión (Deza & Deza, 2009), i.e.,

$$SNR = \frac{\mathrm{E}[\boldsymbol{X}^2]}{\mathrm{E}[(\boldsymbol{X} - \boldsymbol{Y})^2]}. \tag{B6}$$

## B.3. Medidas Objetivas Basadas en Parámetros Estadísticos

El coeficiente de correlación de Pearson permite determinar la intensidad en la dependencia lineal entre dos señales (Heiberger & Holland, 20015):

$$\rho(\boldsymbol{X}, \boldsymbol{Y}) = \frac{\mathrm{cov}(\boldsymbol{X}, \boldsymbol{Y})}{\sigma_X \sigma_Y}. \tag{B7}$$

El coeficiente de correlación toma valores entre -1 y 1, i.e., $-1 \leq \rho \leq 1$, donde el signo indica si la dependencia entre las señales es directa o inversamente proporcional y el caso de la igualdad a cero, $\rho = 0$, implica que las dos señales no están correlacionadas.

El Índice de Similitud Estructural ($SSIM$) se sustenta en la idea de medir los cambios sufridos por la información estructural de la señal $\boldsymbol{Y}$ en comparación con $\boldsymbol{X}$, por lo que no realiza una comparación directa entre éstas. El cálculo de este índice se realiza una ponderación entre tres medidas diferentes: luminosidad ($\ell$), contraste ($c$) y estructura ($\mathcal{s}$) (Kandadai et al., 2008; Lai et al., 2013):

$$\ell(\boldsymbol{X},\boldsymbol{Y}) = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1}, \tag{B8}$$

$$c(\boldsymbol{X},\boldsymbol{Y}) = \frac{2\sigma_X\sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2}, \tag{B9}$$

$$\mathcal{s}(\boldsymbol{X},\boldsymbol{Y}) = \frac{\sigma_{XY} + C_3}{\sigma_X\sigma_Y + C_3}, \tag{B10}$$

donde $\mu_X$ y $\mu_Y$ son las medias de $\boldsymbol{X}$ y $\boldsymbol{Y}$ respectivamente; $\sigma_X$ y $\sigma_Y$ sus desviaciones estándar; $\sigma_{XY}$ es la covarianza de las dos señales; y $C_1, C_2$ y $C_3$ son valores constantes introducidos para evitar inestabilidad.

Finalmente, el $SSIM$ se obtiene de una combinación de las tres medidas, donde los valores constantes $\alpha > 0$, $\beta > 0$ y $\gamma > 0$ determinan la importancia relativa de cada una de estas medidas, como se presenta a continuación:

$$SSIM = \ell(\boldsymbol{X},\boldsymbol{Y})^{\alpha} c(\boldsymbol{X},\boldsymbol{Y})^{\beta} \mathcal{s}(\boldsymbol{X},\boldsymbol{Y})^{\gamma}. \tag{B11}$$

### B.3. Medidas Subjetivas

En esta prueba los oyentes califican una serie de archivos de audio utilizando una escala de cinco niveles:

malo (1) - pobre (2) - regular (3) - bueno (4) - excelente (5).

Después de escuchar cada muestra, los oyentes expresan una opinión, fundamentada únicamente en la muestra escuchada. El promedio de todos los puntajes obtenidos para un mismo audio representa su MOS (Streijl et al., 2014).

Una de las principales ventajas de la MOS es que aporta una eficiente retroalimentación sobre la calidad de las señales de voz cuantificadas, a partir de la evaluación de los oyentes. Además, la aplicación de MOS no requiere procedimientos de estandarización y calibración del oyente, estímulos de habla preespecificados, entornos de prueba y otros requisitos de procedimiento rígidos, como lo hacen otros tipos de pruebas, haciendo de la MOS una herramienta flexible (Polkosky & Lewis, 2003).

# APÉNDICE C – ANÁLISIS DE FOURIER

## C.1. Series de Fourier

En la primera mitad del siglo XIX J. Fourier publicó su ´*Teoría analítica del calor*', a partir de la cual desarrolló lo que hoy se conoce como serie de Fourier, la cual es una herramienta matemática que permite representar cualquier señal periódica de potencia finita, de forma biunívoca, como una combinación lineal de los elementos de un conjunto ortogonal de funciones $\{\varphi_n(t)\}$. El conjunto ortogonal utilizado por Fourier es $\varphi_n(t) = \{\sin(2\pi n f_o t), \cos(2\pi n f_o t)\}$, dando origen a la serie trigonométrica de Fourier. Así, para una señal $x(t)$ con periodo $T_o$ y potencia finita $P_x$, se tiene que su expresión mediante la serie de Fourier está dada por:

$$x(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos(2\pi n f_o t) + b_n \sin(2\pi n f_o t), \qquad (C1)$$

donde, $a_0$ corresponde al nivel DC de la señal; $f_o$ es la frecuencia de la señal; y los coeficientes $a_n$ y $b_n$ equivalen a la correlación de la señal con las sinusoidales de diferentes frecuencias, i.e., el cálculo del producto interno de la señal con funciones coseno para el caso de $a_n$ y con funciones seno para el caso de $b_n$:

$$a_n = \frac{2}{T_o} \int_{-T_o/2}^{T_o/2} x(t) \cos(2\pi n f_o t) dt, \qquad (C2)$$

y

$$b_n = \frac{2}{T_o} \int_{-T_o/2}^{T_o/2} x(t) \sin(2\pi n f_o t) dt, \qquad (C3)$$

donde $\frac{2}{T_o}$ corresponde al factor de escala dado que el producto interno de una sinusoide consigo misma es igual a $\frac{T_o}{2}$.

Toda señal se puede representar mediante la suma de su parte par, $x_p(t)$, y su parte impar, $x_i(t)$, dicha representación se consigue de la siguiente manera:

$$x_p(t) = \frac{x(t) + x(-t)}{2};$$

y

$$x_i(t) = \frac{x(t) - x(-t)}{2},$$

así, la representación de las componentes $x_p(t)$ y $x_i(t)$ mediante la serie de Fourier está dada por (C4) y (C5), de donde se infiere que, si la señal periódica a analizar es de naturaleza par, entonces los componentes $b_n$ son nulos y que en caso de que la señal sea impar los $a_n$ son iguales a cero.

$$x_p(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos(2\pi n f_o\, t). \qquad (C4)$$

$$x_i(t) = \sum_{n=1}^{\infty} b_n \sin(2\pi n f_o\, t). \qquad (C5)$$

De forma generalizada las series de Fourier plantean que para la reconstrucción de la señal se necesita un conjunto de funciones base ortogonales $\{\varphi_n(t)\}$. La representación de señales sinusoidales por medio de exponenciales complejas da origen a la serie compleja de Fourier (ver ecuación C6), donde el conjunto base es $\varphi_n(t) = \{e^{j2\pi n f_o t}\}$. En este caso las exponenciales complejas corresponden a fasores que giran, en sentidos opuestos, a cada una de las frecuencias consideradas.

$$x(t) = \sum_{n=-\infty}^{\infty} c_n\, e^{-j2\pi n f_o\, t}, \qquad (C6)$$

donde, $c_n$ corresponde al conjunto de coeficientes resultantes del producto interno de la señal con las exponenciales complejas, el cual se presenta a continuación:

$$c_n = \frac{1}{T_o} \int_{-T_o/2}^{T_o/2} x(t)\, \mathrm{e}^{-\mathrm{j}2\pi n f_o\, t}\, dt. \qquad (C7)$$

Para cualquiera de las formas de la serie de Fourier planteadas se tiene que $n$ siempre toma valores enteros, esto implica que las señales periódicas tienen una naturaleza discreta en el dominio de la frecuencia y que la distancia de las componentes en este dominio corresponde al inverso del periodo de la señal.

### C.2. Transformada de Fourier

La extensión del análisis de Fourier para señales no periódicas de energía finita, que se conoce como FT. La FT está dada por:

$$\tilde{x}(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi f t} dt. \qquad (C8)$$

Como la FT es una expansión de la serie de Fourier sigue cumpliendo que la representación de la señal en el dominio de la frecuencia corresponde a una relación biunívoca y, por lo tanto, invertible, por medio de la siguiente función:

$$x(t) = \int_{-\infty}^{\infty} \tilde{x}(f) e^{j2\pi f t} df. \qquad (C9)$$

El hecho de que esta transformación sea invertible indica que toda la información de la señal, que se encuentra en el dominio del tiempo, se conserva en el dominio de la frecuencia. La conservación de la información se puede demostrar comprobando que la energía de la señal, $\mathcal{E}_x$, es invariante sin importar el dominio en el que ésta esté representada, esto se conoce como el teorema de la energía de Rayleigh:

$$\mathcal{E}_x = \int_{-\infty}^{\infty} |x(t)|^2 \, dt = \int_{-\infty}^{\infty} |\tilde{x}(f)|^2 \, df. \qquad (C10)$$

Generalmente la FT se trabaja con el conjunto ortogonal de funciones base representado mediante exponenciales complejas por facilidad matemática.

La definición de la FT utilizando las funciones sinusoidales se expresa en (C11). Al calcular la FT por medio de su forma trigonométrica se tiene como resultado una representación bidimensional con coeficientes reales, dicha representación se puede ver como una componente de la señal en fase, $\tilde{x}_I(f)$, y una en cuadratura, $\tilde{x}_Q(f)$, como se presenta a continuación:

$$\tilde{x}_I(f) = \sqrt{2} \int_{-\infty}^{\infty} x_p(t) \cos(2\pi f t) \, dt,$$

$$\tilde{x}_Q(f) = \sqrt{2} \int_{-\infty}^{\infty} x_i(t) \sin(2\pi f t) \, dt,$$

$$\tilde{x}(f) = \tilde{x}_p(f) - j\tilde{x}_i(f). \qquad (C11)$$

La transformada inversa de la forma trigonométrica se presenta a continuación:

$$x(t) = \sqrt{2} \int_{-\infty}^{\infty} \tilde{x}_I(f) \cos(2\pi f t) \, df + \sqrt{2} \int_{-\infty}^{\infty} \tilde{x}_Q(f) \sin(2\pi f t) \, df. \qquad (C12)$$

El análisis de Fourier, compuesto por la serie y la transformada, consiste en la forma más simple de asociar un comportamiento en el dominio del tiempo con ciertas componentes en frecuencia; sin embargo, se debe tener en cuenta que no es la única forma de realizar una transformación lineal, ni de utilizar funciones sinusoidales como funciones base. Dentro de las alternativas existentes se

encuentra la transformada de Hartley (Bracewell, 2000) o variantes como la Transformada Coseno (CT, *Cosine Transform*) y la Transformada Seno (ST, *Sine Transform*).

La representación de la señal en el dominio de la frecuencia por medio de la CT se nota por medio de $\hat{x}(f)$ y en el caso de la ST se utiliza $\check{x}(f)$ para realizar su distinción, como se presenta a continuación:

$$\hat{x}(f) = 2\int_0^{\infty} x(t)\cos(2\pi ft)\,dt. \qquad (C13)$$

$$\check{x}(f) = 2\int_0^{\infty} x(t)\sin(2\pi ft)\,dt. \qquad (C14)$$

La constante que acompaña a la integral para realizar las transformaciones tiene como una función normalizar el aporte de cada una de las componentes de frecuencia, puesto que el producto interno de una señal sinusoidal consigo misma es ½, en este caso se toma dicha constante como 2; no obstante, otros autores toman como constante $\sqrt{\frac{2}{\pi}}$ (Kreyszig, 2011).

Es importante resaltar que, dentro del análisis de Fourier, el espacio de señales $\mathcal{L}_2$ solo puede generarse a partir de la combinación de espacios generados por las funciones base $\varphi_n(t) = \{\sin(2\pi nft), \cos(2\pi nft)\}$ (ver Figura C.1), por lo que la aplicación de las CT y ST conlleva una pérdida de información en señales compuestas tanto por una componente impar como una componente par.

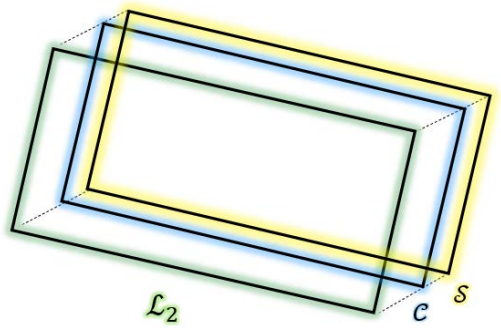


Figura C.1. Espacio de señales $\mathcal{L}_2$ representado a partir de las funciones base del análisis de Fourier. Elaboración propia.

# APÉNDICE D – WAVELETS

En las transformaciones lineales que plantean una relación biunívoca entre los dos dominios, i.e., garantizan la conservación de la energía o la potencia de la señal sin importar su representación, se tiene que cualquier cambio que se aplique sobre la señal en uno de los dominios tiene repercusiones sobre la imagen en el dominio restante. Lo anterior motiva el estudio del uso de dominios transformados para procesos en los que la distorsión de la señal es inherente, como la cuantificación, dado que se espera que al regresar la señal a su dominio original la distorsión percibida sea menor.

La utilización de la WT para crear una representación alternativa de la información de la señal implica que se tomen decisiones con respecto al número de etapas del algoritmo de Mallat y el tipo de familia *wavelet* a utilizar en el MRA, por lo que en este apéndice se caracterizan las familias *wavelet* consideradas en este trabajo de grado.

Existen familias *wavelet* que, por compartir el método por medio del cual se crearon o la persona autora de las mismas, tienen nombres similares, los cuales se diferencian por un número referente a los momentos de desvanecimiento de la función (Burrus et al., 1998). En la Tabla D.1 se muestran las familias *wavelet* ortogonales del MRA, cuyos filtros FIR se encuentran disponibles en el entorno de simulación MATLAB®, por lo cual se especifica el indicador que permite su generación. En total, se tienen 100 familias *wavelet* ortogonales diferentes, puesto que la *wavelet* de *Haar* es igual a la *Daubechies-1*.

Tabla D.1. Familias *wavelet* ortogonales en MATLAB®.
Adaptado de (MATLAB, n.d.).

| Familia | Identificador |
|---|---|
| Coiflet | 'coif$N$', $N = 1{:}5$; |
| Daubechies | 'db$N$', $N = 1{:}45$; |
| Fejér-Korovkin | 'fk$N$', $N = 4{,}6{,}8{,}14{,}18{,}22$; |
| Haar | 'haar' 'db1' |
| Symlet | 'sym$N$', $N = 2{:}45$; |

## D.1. Índice de Gini de las Familias *Wavelet*

En la Tabla D.2 se consignan los valores del índice de Gini para todas las familias *wavelet* y para un MRA que varía entre 1 y 3 etapas de profundidad, y para cada familia *wavelet* está consignada la longitud de los filtros con los cuales se generan ($L_f$). En la Figura D.2[25] se muestra en color azul las diferentes subbandas y en color verde el espectro de la señal de voz, lo cual corrobora que los valores del *índice de*

[25] Los valores de frecuencia de la Figura D.2 se encuentran normalizados con respecto a la frecuencia de muestreo.

*Gini* se encuentran entre 0.7999 y 0.9521, que, en ambos casos, indican una gran desigualdad en la concentración de la energía. De forma comparativa, el *índice de Gini* de las señales en el dominio del tiempo es de 0.6288 (significativamente menor), mientras que sobre el espectro de las señales se tienen *índices de Gini* de 0.9655 con su magnitud y de 0.9713 con su parte real (en ambos casos por encima del mejor índice de Gini obtenido a partir de los coeficientes del MRA).


Figura D.1. Subbandas del espectro de una señal de voz.
Elaboración propia.

Tabla D.2. Índice de Gini.
Elaboración propia.

| Familia Wavelet | $L_f$ | Número de Etapas | | | Familia Wavelet | $L_f$ | Número de Etapas | | | Familia Wavelet | $L_f$ | Número de Etapas | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | | | 1 | 2 | 3 | | | 1 | 2 | 3 |
| coif1 | 6 | 0.807 | 0.894 | 0.932 | db30 | 60 | 0.817 | 0.907 | 0.949 | sym13 | 26 | 0.811 | 0.901 | 0.943 |
| coif2 | 12 | 0.808 | 0.897 | 0.938 | db31 | 62 | 0.817 | 0.907 | 0.949 | sym14 | 28 | 0.812 | 0.901 | 0.944 |
| coif3 | 18 | 0.810 | 0.899 | 0.941 | db32 | 64 | 0.818 | 0.907 | 0.949 | sym15 | 30 | 0.812 | 0.901 | 0.944 |
| coif4 | 24 | 0.811 | 0.900 | 0.942 | db33 | 66 | 0.818 | 0.908 | 0.949 | sym16 | 32 | 0.812 | 0.902 | 0.944 |
| coif5 | 30 | 0.812 | 0.901 | 0.944 | db34 | 68 | 0.818 | 0.908 | 0.950 | sym17 | 34 | 0.812 | 0.902 | 0.944 |
| **Promedio** | | 0.809 | 0.898 | 0.940 | db35 | 70 | 0.819 | 0.908 | 0.950 | sym18 | 36 | 0.813 | 0.903 | 0.945 |
| db1 | 2 | 0.800 | 0.880 | 0.910 | db36 | 72 | 0.819 | 0.909 | 0.950 | sym19 | 38 | 0.813 | 0.903 | 0.945 |
| db2 | 4 | 0.806 | 0.894 | 0.931 | db37 | 74 | 0.819 | 0.909 | 0.950 | sym20 | 40 | 0.814 | 0.903 | 0.946 |
| db3 | 6 | 0.807 | 0.896 | 0.936 | db38 | 76 | 0.820 | 0.909 | 0.951 | sym21 | 42 | 0.814 | 0.903 | 0.946 |
| db4 | 8 | 0.808 | 0.897 | 0.938 | db39 | 78 | 0.820 | 0.909 | 0.951 | sym22 | 44 | 0.814 | 0.904 | 0.946 |
| db5 | 10 | 0.808 | 0.897 | 0.939 | db40 | 80 | 0.820 | 0.910 | 0.951 | sym23 | 46 | 0.815 | 0.905 | 0.947 |
| db6 | 12 | 0.809 | 0.898 | 0.940 | db41 | 82 | 0.821 | 0.910 | 0.951 | sym24 | 48 | 0.815 | 0.905 | 0.947 |
| db7 | 14 | 0.809 | 0.900 | 0.941 | db42 | 84 | 0.821 | 0.910 | 0.952 | sym25 | 50 | 0.815 | 0.905 | 0.947 |
| db8 | 16 | 0.809 | 0.899 | 0.941 | db43 | 86 | 0.821 | 0.911 | 0.952 | sym26 | 52 | 0.816 | 0.906 | 0.948 |
| db9 | 18 | 0.810 | 0.899 | 0.942 | db44 | 88 | 0.822 | 0.911 | 0.952 | sym27 | 54 | 0.816 | 0.905 | 0.948 |
| db10 | 20 | 0.810 | 0.899 | 0.942 | db45 | 90 | 0.822 | 0.911 | 0.952 | sym28 | 56 | 0.816 | 0.906 | 0.948 |
| db11 | 22 | 0.810 | 0.900 | 0.943 | **Promedio** | | 0.811 | 0.901 | 0.942 | sym29 | 58 | 0.817 | 0.906 | 0.948 |
| db12 | 24 | 0.811 | 0.900 | 0.943 | fk4 | 4 | 0.803 | 0.887 | 0.920 | sym30 | 60 | 0.817 | 0.907 | 0.849 |
| db13 | 26 | 0.811 | 0.901 | 0.943 | fk6 | 6 | 0.807 | 0.896 | 0.937 | sym31 | 62 | 0.818 | 0.907 | 0.949 |
| db14 | 28 | 0.812 | 0.901 | 0.944 | fk8 | 8 | 0.808 | 0.897 | 0.939 | sym32 | 64 | 0.818 | 0.907 | 0.949 |
| db15 | 30 | 0.812 | 0.901 | 0.944 | fk14 | 14 | 0.809 | 0.898 | 0.942 | sym33 | 66 | 0.818 | 0.908 | 0.949 |
| db16 | 32 | 0.812 | 0.902 | 0.945 | fk18 | 18 | 0.810 | 0.899 | 0.943 | sym34 | 68 | 0.819 | 0.908 | 0.950 |
| db17 | 34 | 0.813 | 0.902 | 0.945 | fk22 | 22 | 0.811 | 0.900 | 0.943 | sym35 | 70 | 0.819 | 0.908 | 0.950 |
| db18 | 36 | 0.813 | 0.903 | 0.945 | **Promedio** | | 0.808 | 0.896 | 0.937 | sym36 | 72 | 0.819 | 0.909 | 0.950 |
| db19 | 38 | 0.813 | 0.903 | 0.946 | sym2 | 4 | 0.806 | 0.894 | 0.931 | sym37 | 74 | 0.819 | 0.909 | 0.950 |
| db20 | 40 | 0.814 | 0.903 | 0.946 | sym3 | 6 | 0.807 | 0.896 | 0.936 | sym38 | 76 | 0.820 | 0.909 | 0.951 |
| db21 | 42 | 0.814 | 0.904 | 0.946 | sym4 | 8 | 0.808 | 0.897 | 0.938 | sym39 | 78 | 0.820 | 0.909 | 0.951 |
| db22 | 44 | 0.814 | 0.904 | 0.946 | sym5 | 10 | 0.808 | 0.897 | 0.939 | sym40 | 80 | 0.820 | 0.910 | 0.951 |
| db23 | 46 | 0.815 | 0.904 | 0.947 | sym6 | 12 | 0.809 | 0.898 | 0.940 | sym41 | 82 | 0.821 | 0.910 | 0.951 |
| db24 | 48 | 0.815 | 0.905 | 0.947 | sym7 | 14 | 0.809 | 0.898 | 0.940 | sym42 | 84 | 0.821 | 0.911 | 0.951 |
| db25 | 50 | 0.815 | 0.905 | 0.947 | sym8 | 16 | 0.809 | 0.899 | 0.941 | sym43 | 86 | 0.822 | 0.911 | 0.952 |
| db26 | 52 | 0.816 | 0.905 | 0.948 | sym9 | 18 | 0.810 | 0.899 | 0.941 | sym44 | 88 | 0.822 | 0.911 | 0.952 |
| db27 | 54 | 0.816 | 0.906 | 0.948 | sym10 | 20 | 0.810 | 0.900 | 0.942 | sym45 | 90 | 0.822 | 0.911 | 0.952 |
| db28 | 56 | 0.816 | 0.906 | 0.948 | sym11 | 22 | 0.810 | 0.900 | 0.942 | **Promedio** | | 0.811 | 0.900 | 0.942 |
| db29 | 58 | 0.817 | 0.906 | 0.948 | sym12 | 24 | 0.811 | 0.900 | 0.943 | | | | | |

## D.2. Implementación del Algoritmo de Mallat

Dado que el algoritmo de Mallat utiliza un banco de filtros y convoluciones en tiempo discreto para su implementación, se debe tener en cuenta el número de muestras adicionales que se derivan de su implementación, el cual se presenta a continuación:

$$L_y = L_f + L_x - 1, \tag{D2}$$

Donde, $L_x$ es el número de muestras de la señal; $L_h$ es el número de muestras del filtro; y $L_y$ es el número de coeficientes a la salida del filtro. Los filtros para generar las diferentes familias wavelet en MATLAB® cumplen la siguiente condición:

$$2 \leq L_f \leq 90. \tag{D3}$$

Con el algoritmo de Mallat es posible, eliminando la respuesta transitoria de los filtros, garantizar que el número de coeficientes *wavelet* y *scaling* es igual al número de muestras de la señal en el tiempo, no obstante, si los valores de $L_x$ y $L_f$ son comparables, esta aproximación puede generar errores en la señal reconstruida. En la Figura D.3 se muestra el inicio de una trama de 64 ms de una señal de voz muestreada a 16 KHz ($L_x = 1024$), con la cual se está utilizando una wavelet *db45* ($L_f = 90$) y se muestra la comparación de las señales reconstruidas con una y tres etapas de profundidad, gracias a lo cual se evidencia que la distorsión es proporcional al número de muestras adicionales, es decir a $L_f$ y el número de etapas.

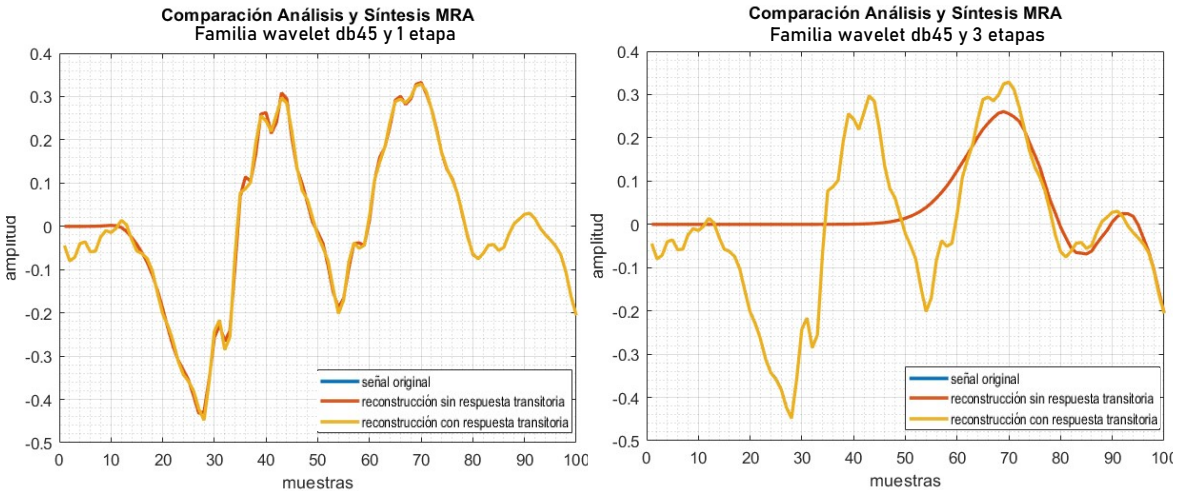


Figura D.2. Efecto de la respuesta transitoria de los filtros en el algoritmo de Mallat.
Elaboración propia.

Para $n$ etapas, la longitud de los coeficientes n-ésimos está dada por:

$$L_y = \left\lfloor \frac{L_x}{2^n} + \frac{2^n - 1}{2^n} L_f \right\rfloor. \tag{D4}$$

Del mismo modo, la señal reconstruida a partir de las $n$ etapas tiene una longitud aproximada[26] de:

$$L_{x'} = L_x + 2^n L_f - 1. \qquad (D5)$$

Por lo anterior, para este trabajo de grado se crean funciones personalizadas para hacer la descomposición de la señal (DWT) y su posterior reconstrucción por medio de la Transformada Inversa Discreta *Wavelet* (IDWT, *Inverse Discrete Wavelet Transform*). En la Figura D.3 la señal amarilla, la cual se superpone con la señal original en azul, es obtenida utilizando estas funciones.

[26] Es posible que se presente una diferencia de una muestra por el redondeo que se hace en submuestreo de la descomposición, dado que el vector de coeficientes resultante siempre debe tener un número entero como longitud.

```
function [s,w] = p_dwt(x, ha, ga)
   %p_dwt es una función personalizada para aplicar la DWT sobre un vector
   %x es el vector que equivale a los coeficientes scaling de más alto orden
   %ha son los coeficientes para representar la función scaling
   %ga son los coeficientes para representar la función wavlet
   %s es el vector de coeficientes scaling de un orden inferior
   %w es el vector de coeficientes wavelet de un orden inferieor

   %dado que la DWT es un algoritmo iterativo recursivo, esta función
   %deberá llamarse por cada nivel de descomposición que se desee
   %implementar
   w=conv(ga,x); w=downsample(w,2);
   s=conv(ha,x); s=downsample(s,2);
end

function recx=p_idwt(s, w, hs, gs)
   %p_idwt es una función personalizada para aplicar la IDWT sobre un vector
   %hs son los coeficientes para representar la función scaling
   %gs son los coeficientes para representar la función wavlet
   %s es el vector de coeficientes scaling de más bajo orden
   %w es el vector de coeficientes wavelet de más bajo orden
   %recx es el vector de coeficientes scaling de un orden superior

   %dado que la IDWT es un algoritmo iterativo recursivo, esta función
   %deberá llamarse por cada nivel de descomposición implementado

        L=length(hs);
        fix=length(w)-length(s); %en caso de que existan diferencias en las longitudes
        de los coeficientes
         if fix<0
                  w=[w zeros(1,fix)];
         elseif fix>0
                 s=[s zeros(1,fix)];
        end
        recx=conv(gs,upsample(w,2))+conv(hs,upsample(s,2));
        recx=recx(L:end-L); %se recordan las 2L-1 muestras que agrega el proceso
        de convolución
end
```

# APÉNDICE E – BASE DE DATOS

Existen muchas bases de datos gratuitas, de las cuales es posible obtener muestras de señales de voz para probar, como en este caso, algoritmos de procesamiento digital de señales. Dentro de las bases de datos disponibles se destacan

- **Emo-DB**. Es una base de datos libre, con muestras en alemán, diferenciadas según la emoción, las cuales se encuentran disponibles en formato 'wav' a 48 KHz y 16 KHz.
  https://www.kaggle.com/piyushagni5/berlin-database-of-emotional-speech-emodb.
- **FestVox**. Es una base de datos libre, con muestras en inglés, las cuales se encuentran disponibles en formato 'wav' a 16 KHz.
  https://github.com/festvox/
- **LibriSpeech**. Es una base de datos libre de *open speech and language resources*, con muestras en inglés, las cuales se encuentran disponibles en formato 'flac' a 16 KHz.
  https://www.openslr.org/12
- **SIWIS**. Es una base de datos libre de la *University of Edinburgh,* con muestras en francés, las cuales se encuentran disponibles en formato 'wav' a 44.1 KHz.
  https://datashare.ed.ac.uk/handle/10283/2353
- **VoxForge**. Es una base de datos libre, con muestras en inglés, las cuales se encuentran en formato 'wav' a 8 KHz.
  http://www.voxforge.org/

Dentro de la búsqueda realizada no se encontraron bases de datos abiertas con señales de voz en español. Dado que existen diferencias entre la evaluación subjetiva realizada por hablantes nativos y no nativos (Janska & Clark, 2010), se hace necesaria la construcción de una base de datos con señales de voz en español, para lo cual se siguen los siguientes criterios

- Las señales de voz se construyen a partir de la lectura de diferentes frases escritas.
- Las muestras de voz se graban sin ruido de fondo.
- La frecuencia de muestro de las señales debe ser 16 KHz, con el fin de representar las señales en el área alta y de esta forma garantizar una mayor naturalidad y entendimiento de las éstas (ver Figura 1.17).
- El formato debe ser 'wav' con el fin de tener un formato de compresión sin pérdidas de amplio uso compatible con MATLAB®.

*Jiang & Yang* definen que las tramas de las señales de voz deben tener una duración entre 25 y 64 ms (B. Jiang & Yang, 2011). Para este trabajo de grado se utiliza una duración de 64 ms con el fin de reducir la información adicional asociada

a la respuesta transitoria de los filtros del MRA y los datos del cuantificador aplicado a cada trama. De la unión del conjunto de señales seleccionadas se crean tramas de 64 ms, generando 33,750 tramas, sobre las cuales se aplica el proceso de cuantificación y se evalúa el desempeño.

# APÉNDICE F – CUANTIFICADORES

## F.1. Medidas Objetivas Cuantificación Uniforme Versus No Uniforme

En el apéndice B se describen 6 medidas objetivas diferentes, no obstante, en el caso particular del MSE no se consideran sus resultados, puesto que los resultados obtenidos con esta medida son aproximadamente cero, impidiendo ver diferencias entre las familias *wavelet* y los diferentes algoritmos de cuantificación.

La primera medida considerada es el M-NRMSE, el cual se define en (B4) y corresponde a una medida normalizada en la que el valor de 1 indica que las dos señales comparadas son idénticas. En la Tabla F.1 se consignan los valores promediados de las señales reconstruidas con las 100 familias *wavelet* utilizando un MRA con diferente número de etapas ($\xi$); adicionalmente, se consideran variaciones en el número de niveles de cuantificación $N = \{8, 16, 32, 64, 128\}$ y el tipo de algoritmo de cuantificación (uniforme y no uniforme).

Tabla F.1. M-NRMSE promedio según el tipo de cuantificación, $\boldsymbol{N}$ y $\boldsymbol{\xi}$.
Elaboración propia.

| ξ | | Uniforme | | | | | No uniforme | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | N | | | | | N | | | | |
| | | 8 | 16 | 32 | 64 | 128 | 8 | 16 | 32 | 64 | 128 |
| 1 | media | 0.868886 | 0.93411 | 0.966838 | 0.983354 | 0.991683 | 0.958759 | 0.963253 | 0.964927 | 0.964719 | 0.965696 |
| | varianza | 7.031E-08 | 1.04E-08 | 2.582E-09 | 3.115E-09 | 2.031E-09 | 1.061E-05 | 1.082E-05 | 1.084E-05 | 1.076E-05 | 1.077E-05 |
| 2 | media | 0.826188 | 0.912333 | 0.955499 | 0.977513 | 0.988676 | 0.945063 | 0.945459 | 0.94602 | 0.94642 | 0.946857 |
| | varianza | 1.528E-07 | 5.031E-08 | 1.242E-08 | 1.65E-08 | 5.154E-08 | 5.102E-05 | 5.131E-05 | 5.147E-05 | 5.158E-05 | 5.197E-05 |
| 3 | media | 0.773352 | 0.884815 | 0.941263 | 0.970179 | 0.984903 | 0.916787 | 0.917531 | 0.918471 | 0.919191 | 0.9199 |
| | varianza | 9.714E-05 | 7.019E-07 | 1.704E-07 | 1.934E-07 | 5.286E-07 | 0.0002322 | 0.000222 | 0.0002118 | 0.0002015 | 0.0001947 |

En la Figura F.1 se sintetizan los resultados consignados en la Tabla F.1: en el lado izquierdo se muestran los resultados de la cuantificación uniforme y en el derecho de la cuantificación no uniforme. En general, con $\xi = 1$ se obtienen los mejores resultados. Con relación al incremento del valor de $N$, en el cuantificador uniforme se evidencia una mejora, mientras que en el no uniforme no se evidencia una mejora significativa. Al comparar los dos algoritmos de cuantificación se muestra que ante un número pequeño de niveles de cuantificación (8,16), es más conveniente utilizar la cuantificación no uniforme.

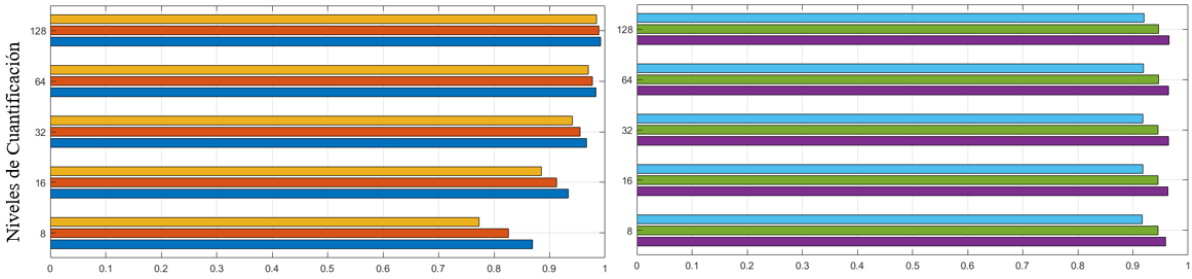


Figura F.1. M-NRMSE promedio.
Elaboración propia.

El MAE, definido según (B5), corresponde a una medida absoluta de error. Si las señales comparadas son idénticas, entonces se tiene que $MAE = 0$. En la Tabla F.2 se consignan los valores promediados de la distorsión obtenida al comparar las señales reconstruidas al usar el MRA con las 100 familias *wavelet* y variar el número de etapas ($\xi$), estos resultados se obtienen para $N = \{8, 16, 32, 64, 128\}$ y la cuantificación uniforme y la cuantificación no uniforme.

Tabla F.2. MAE promedio según el tipo de cuantificación, $\boldsymbol{N}$ y $\boldsymbol{\xi}$.
Elaboración propia.

| ξ | | Uniforme N | | | | | No uniforme N | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 8 | 16 | 32 | 64 | 128 | 8 | 16 | 32 | 64 | 128 |
| 1 | media | 0.018444 | 0.009207 | 0.004613 | 0.0023 | 0.001174 | 0.004038 | 0.00323 | 0.003068 | 0.003103 | 0.003035 |
| | varianza | 2.6909E-09 | 3.4859E-09 | 1.1424E-09 | 9.309E-36 | 1.9434E-09 | 3.2723E-07 | 3.4515E-07 | 3.6119E-07 | 3.5262E-07 | 3.5139E-07 |
| 2 | media | 0.024385 | 0.012219 | 0.006181 | 0.003062 | 0.001595 | 0.005709 | 0.00571 | 0.005694 | 0.005666 | 0.005627 |
| | varianza | 2.4924E-08 | 3.9788E-09 | 2.5797E-08 | 3.1067E-08 | 2.5E-09 | 1.3245E-06 | 1.2787E-06 | 1.2735E-06 | 1.2748E-06 | 1.2909E-06 |
| 3 | media | 0.032003 | 0.016079 | 0.008186 | 0.00414 | 0.002099 | 0.010209 | 0.01009 | 0.009972 | 0.009876 | 0.009786 |
| | varianza | 2.8474E-07 | 3.3595E-08 | 6.8727E-09 | 2.8283E-09 | 1E-10 | 5.9342E-06 | 5.5633E-06 | 5.2849E-06 | 5.016E-06 | 4.8073E-06 |

El comportamiento de los algoritmos de cuantificación frente a esta medida es igual que en el M-NRMSE, aunque en este caso, la mejora en la cuantificación uniforme con respecto al aumento en el número de niveles de cuantificación es más evidente (ver Figura F.2).

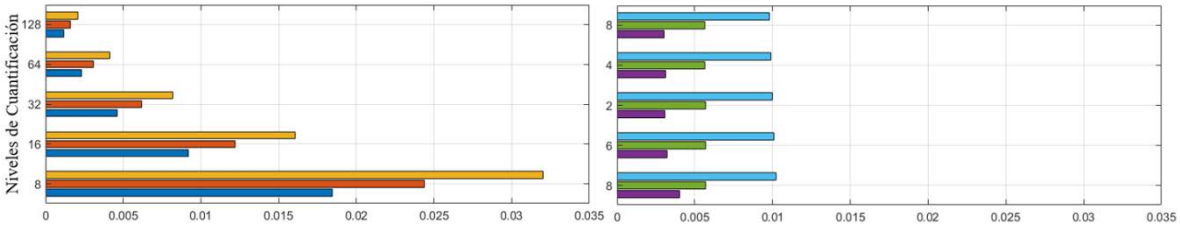


Figura F.2. MAE promedio.
Elaboración propia.

La SNR o $S_QNR$, definida en (B6), considera como ruido el valore esperado del erroral cuadrado, el cual está dado por la diferencia de las dos señales comparadas, no obstante, como los valores de error son tan pequeños, los resultados con esta medida tienden a ser muy grandes (ver Tabla F.3), por lo cual, la representación gráfica de dichos resultados se realiza en escala logarítmica.

Tabla F.3. SNR promedio según el tipo de cuantificación, $\boldsymbol{N}$ y $\boldsymbol{\xi}$.
Elaboración propia.

| ξ | | Uniforme N | | | | | No uniforme N | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 8 | 16 | 32 | 64 | 128 | 8 | 16 | 32 | 64 | 128 |
| 1 | media | 19.210492 | 1.5269E+25 | 2.8493E+26 | 2.5696E+27 | 1.0297E+28 | 682.834229 | 1613.51734 | 1939.34257 | 1939.38253 | 1939.52615 |
| | varianza | 0.02845611 | 2.3316E+52 | 8.1185E+54 | 6.4414E+56 | 9.3689E+57 | 8280.08269 | 85045.0587 | 143501.049 | 143499.133 | 143499.667 |
| 2 | media | 12.875907 | 47.148749 | 1.0721E+25 | 4.326E+25 | 5.5075E+26 | 492.131189 | 491.543135 | 492.208885 | 492.254004 | 492.312817 |
| | varianza | 0.06621553 | 1.66118305 | 1.1493E+52 | 1.8715E+53 | 1.4922E+55 | 10509.5018 | 11090.2557 | 10511.7954 | 10513.4248 | 10515.5225 |
| 3 | media | 8.764246 | 30.140425 | 102.958562 | 4.013E+24 | 9.1559E+25 | 143.189728 | 143.340348 | 143.485409 | 143.622541 | 143.758429 |
| | varianza | 0.07194706 | 1.13619758 | 339.422851 | 1.6104E+51 | 4.8667E+53 | 1305.17723 | 1305.40358 | 1305.07272 | 1304.15984 | 1303.40295 |

A partir de la Figura F.3 se corroboran los comportamientos anteriormente descritos, sin embargo, en este caso se evidencia con mayor claridad o llaneza que en la cuantificación no uniforme sí existe una mejora al aumentar el número de niveles de cuantificación, aunque ésta es más ostensible para $\xi = 1$ y cuando se pasa de $N = \{8,16\}$. Es importante resaltar que, la Figura F.3 es la que plantea la mayor diferencia entre la cuantificación uniforme y la cuantificación no uniforme.

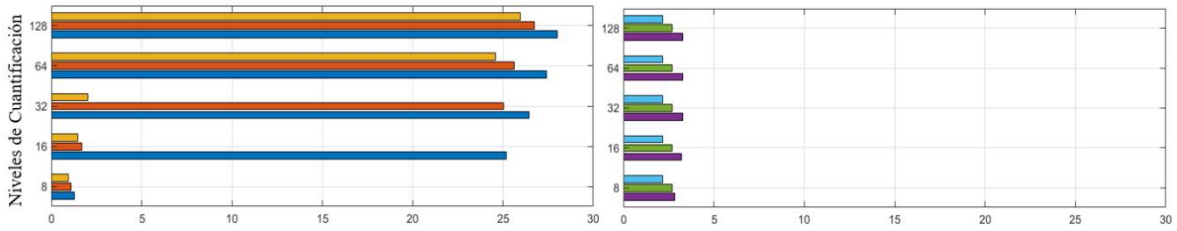


Figura F.3. SNR promedio en escala logarítmica.
Elaboración propia.

El coeficiente de correlación de Pearson -$\rho$ -, definido en (B7), arroja valores normalizados dentro de los cuales el 1 representa una total dependencia lineal. En la Tabla F.4 se consignan los valores promedio de la correlación entre las señales de voz originales y las reconstruidas con el MRA al utilizar 100 familias *wavelet*. A partir los resultados se observa que todos los valores obtenidos son positivos y mayores a 0.9, por lo que, aún en la peor de las reconstrucciones de las señales de voz a partir de los coeficientes cuantificados, se obtiene una fuerte correlación.

Tabla F.4. Coeficiente de correlación promedio según el tipo de cuantificación, $\boldsymbol{N}$ y $\boldsymbol{\xi}$.
Elaboración propia.

| ξ | | Uniforme N | | | | | No uniforme N | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 8 | 16 | 32 | 64 | 128 | 8 | 16 | 32 | 64 | 128 |
| 1 | media | 0.964186 | 0.990513 | 0.9976 | 0.9994 | 0.999816 | 0.99134 | 0.992089 | 0.992995 | 0.993035 | 0.993591 |
| | varianza | 2.1822E-08 | 1.1424E-09 | 1.7929E-30 | 2.4403E-30 | 1.3576E-09 | 3.0586E-07 | 2.4159E-07 | 2.3846E-07 | 2.2917E-07 | 2.3153E-07 |
| 2 | media | 0.938506 | 0.983023 | 0.995619 | 0.998899 | 0.999699 | 0.987342 | 0.987839 | 0.988309 | 0.988642 | 0.988979 |
| | varianza | 1.2703E-07 | 5.8138E-07 | 4.0342E-08 | 1E-10 | 1E-10 | 4.5128E-06 | 4.4646E-06 | 4.4689E-06 | 4.4713E-06 | 4.475E-06 |
| 3 | media | 0.901495 | 0.970742 | 0.992257 | 0.998026 | 0.999492 | 0.977759 | 0.978538 | 0.979295 | 0.97988 | 0.980415 |
| | varianza | 2.8397E-06 | 2.7741E-07 | 1.7223E-08 | 6.3879E-09 | 6.4E-09 | 5.0732E-05 | 4.5558E-05 | 4.125E-05 | 3.7284E-05 | 3.494E-05 |

De las medidas analizadas hasta el momento, el M-NRMSE y el coeficiente de correlación son las más similares, no solo por el comportamiento general, sino también por las dimensiones de los resultados (ver Figura F.4).

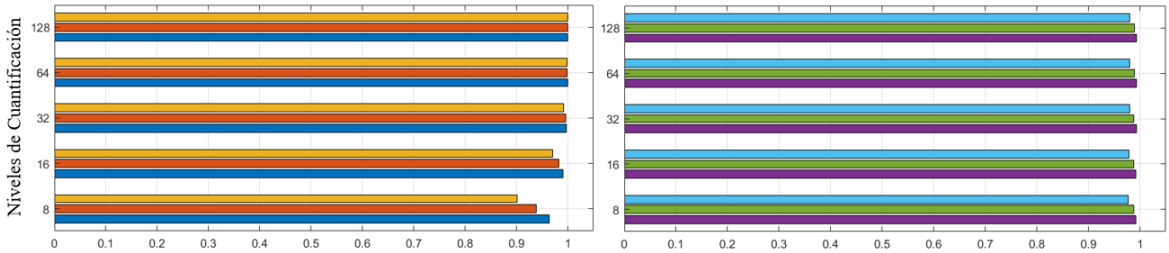


Figura F.4. coeficiente de correlación promedio.
Elaboración propia.

Finalmente, en la Tabla F.5 se consignan los resultados promedio del SSIM para las 100 familias *wavelet* analizadas. Este índice, definido en (B11), arroja resultados normalizados, en donde 1 indica la mayor similitud estructural.

Tabla F.5. SSIM promedio según el tipo de cuantificación, $N$ y $\xi$.
Elaboración propia.

| ξ | | Uniforme N | | | | | No uniforme N | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 8 | 16 | 32 | 64 | 128 | 8 | 16 | 32 | 64 | 128 |
| 1 | media | 0.774327 | 0.917906 | 0.975857 | 0.993696 | 0.9984 | 0.975056 | 0.98048 | 0.982086 | 0.982134 | 0.983032 |
| | varianza | 2.9654E-07 | 3.9762E-08 | 6.9202E-09 | 7.9192E-10 | 4.9802E-32 | 9.4768E-06 | 5.4263E-06 | 5.1808E-06 | 5.1301E-06 | 5.101E-06 |
| 2 | media | 0.693723 | 0.872004 | 0.95827 | 0.988536 | 0.997051 | 0.959576 | 0.960279 | 0.960996 | 0.961509 | 0.96202 |
| | varianza | 2.0757E-06 | 3.9695E-07 | 4.0909E-08 | 4.5495E-09 | 4.7465E-09 | 9.0512E-05 | 9.0418E-05 | 9.0758E-05 | 9.1326E-05 | 9.143E-05 |
| 3 | media | 0.612316 | 0.813518 | 0.931993 | 0.980098 | 0.994753 | 0.913036 | 0.91454 | 0.915928 | 0.9171 | 0.918219 |
| | varianza | 6.0638E-06 | 4.3059E-06 | 1.317E-06 | 1.7151E-07 | 2.0496E-08 | 0.00085557 | 0.00081785 | 0.00078596 | 0.00074898 | 0.00072248 |

Los resultados obtenidos al calcular el SSIM corroboran la hipótesis general del desempeño o comportamiento de los diferentes cuantificadores analizados, dichos resultados se sintetizan en la Figura F.5, gracias a la cual se evidencia que, aunque se obtienen resultados de dimensiones comparables a los de M-NRMSE y el coeficiente de correlación de Pearson, en este caso se pueden distinguir de forma más manifiesta las diferencias, particularmente frente a menores niveles de cuantificación.

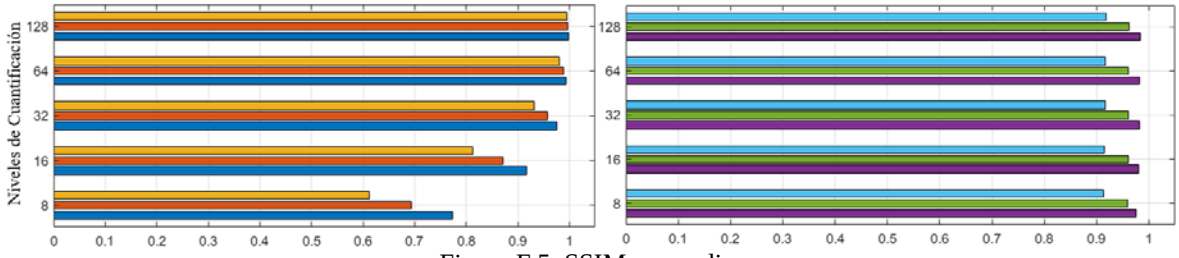


Figura F.5. SSIM promedio.
Elaboración propia.

Las 5 medidas objetivas presentadas anteriormente, muestran también los datos de la varianza de los resultados entre las 100 familias *wavelet* analizadas, dichos resultados, excluyendo los de la SNR que arroja valores de gran dimensión, muestran que no existe mayor diferencia frente a las diferentes familias.

Con el fin de sintetizar los resultados, de tal manera que el análisis de los resultados sea factible, se decide promediar los resultados del M-NRMSE, el coeficiente de correlación de Pearson y el SSIM. Los resultados de esta medida promedio se utilizan como insumo del análisis de resultados objetivos.

En la Tabla F.6 se agrupan las diferentes familias wavelet según la longitud de los filtros, $L_f$, utilizados para su implementación; indicando el *índice de Gini*, $G$; y la medida promedio que se obtiene al variar el número de etapas $\xi = \{1, 2, 3\}$ del MRA y los dos tipos de cuantificación.

Tabla F.6. Resultados medida objetiva promedio para las diferentes familias wavelet.
Elaboración propia.

| | | Número de etapas ($\xi$) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | | | | | | | | | | | 2 | | | | | | | | | | | 3 | | | | | | | | | | |
| Familia wavelet | Lf | G | Uniforme | | | | | No Uniforme | | | | | G | Uniforme | | | | | No Uniforme | | | | | G | Uniforme | | | | | No Uniforme | | | | |
| | | | 8 | 16 | 32 | 64 | 128 | 8 | 16 | 32 | 64 | 128 | | 8 | 16 | 32 | 64 | 128 | 8 | 16 | 32 | 64 | 128 | | 8 | 16 | 32 | 64 | 128 | 8 | 16 | 32 | 64 | 128 |
| db1 | 2 | 0,8 | 0,87 | 0,948 | 0,98 | 0,992 | 0,997 | 0,957 | 0,962 | 0,964 | 0,964 | 0,965 | 0,88 | 0,822 | 0,923 | 0,97 | 0,988 | 0,994 | 0,915 | 0,915 | 0,916 | 0,916 | 0,917 | 0,91 | 0,802 | 0,895 | 0,957 | 0,982 | 0,99 | 0,83 | 0,834 | 0,838 | 0,843 | 0,846 |
| db2 | 4 | 0,806 | 0,87 | 0,948 | 0,98 | 0,992 | 0,997 | 0,973 | 0,976 | 0,978 | 0,978 | 0,979 | 0,894 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,954 | 0,955 | 0,955 | 0,956 | 0,956 | 0,931 | 0,765 | 0,892 | 0,956 | 0,983 | 0,993 | 0,889 | 0,891 | 0,893 | 0,895 | 0,896 |
| fk4 | | 0,803 | 0,87 | 0,948 | 0,98 | 0,992 | 0,997 | 0,963 | 0,968 | 0,97 | 0,97 | 0,971 | 0,887 | 0,821 | 0,923 | 0,97 | 0,989 | 0,995 | 0,933 | 0,934 | 0,934 | 0,935 | 0,935 | 0,92 | 0,769 | 0,894 | 0,958 | 0,984 | 0,993 | 0,856 | 0,86 | 0,862 | 0,865 | 0,867 |
| sym2 | | 0,806 | 0,87 | 0,948 | 0,98 | 0,992 | 0,997 | 0,973 | 0,976 | 0,978 | 0,978 | 0,979 | 0,894 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,954 | 0,955 | 0,955 | 0,956 | 0,956 | 0,931 | 0,765 | 0,892 | 0,956 | 0,983 | 0,993 | 0,889 | 0,891 | 0,893 | 0,895 | 0,896 |
| coif1 | 6 | 0,807 | 0,87 | 0,948 | 0,98 | 0,992 | 0,997 | 0,973 | 0,977 | 0,978 | 0,978 | 0,979 | 0,894 | 0,82 | 0,922 | 0,97 | 0,988 | 0,995 | 0,954 | 0,955 | 0,955 | 0,956 | 0,956 | 0,932 | 0,765 | 0,892 | 0,957 | 0,983 | 0,993 | 0,891 | 0,893 | 0,895 | 0,897 | 0,898 |
| db3 | | 0,807 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,975 | 0,978 | 0,98 | 0,98 | 0,981 | 0,896 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,961 | 0,962 | 0,962 | 0,963 | 0,963 | 0,936 | 0,764 | 0,89 | 0,956 | 0,983 | 0,993 | 0,911 | 0,913 | 0,914 | 0,916 | 0,917 |
| fk6 | | 0,807 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,974 | 0,978 | 0,979 | 0,979 | 0,98 | 0,896 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,961 | 0,961 | 0,962 | 0,962 | 0,962 | 0,837 | 0,763 | 0,89 | 0,955 | 0,983 | 0,993 | 0,917 | 0,919 | 0,92 | 0,921 | 0,922 |
| sym3 | | 0,807 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,975 | 0,978 | 0,98 | 0,98 | 0,981 | 0,896 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,961 | 0,962 | 0,962 | 0,963 | 0,963 | 0,936 | 0,764 | 0,89 | 0,956 | 0,983 | 0,993 | 0,911 | 0,913 | 0,914 | 0,916 | 0,917 |
| db4 | 8 | 0,808 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,897 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,963 | 0,964 | 0,964 | 0,965 | 0,965 | 0,938 | 0,763 | 0,89 | 0,955 | 0,983 | 0,993 | 0,922 | 0,924 | 0,925 | 0,926 | 0,927 |
| fk8 | | 0,808 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,897 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,963 | 0,963 | 0,964 | 0,964 | 0,965 | 0,939 | 0,763 | 0,89 | 0,955 | 0,983 | 0,993 | 0,928 | 0,93 | 0,931 | 0,932 | 0,932 |
| sym4 | | 0,808 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,975 | 0,978 | 0,98 | 0,98 | 0,98 | 0,897 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,963 | 0,964 | 0,964 | 0,964 | 0,965 | 0,938 | 0,763 | 0,891 | 0,956 | 0,983 | 0,993 | 0,922 | 0,923 | 0,925 | 0,926 | 0,927 |
| db5 | 10 | 0,808 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,897 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,964 | 0,965 | 0,965 | 0,966 | 0,966 | 0,939 | 0,762 | 0,89 | 0,955 | 0,983 | 0,993 | 0,928 | 0,93 | 0,931 | 0,932 | 0,933 |
| sym5 | | 0,808 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,975 | 0,978 | 0,98 | 0,98 | 0,98 | 0,897 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,964 | 0,964 | 0,965 | 0,965 | 0,966 | 0,939 | 0,765 | 0,891 | 0,956 | 0,983 | 0,993 | 0,928 | 0,93 | 0,931 | 0,932 | 0,933 |
| coif2 | 12 | 0,808 | 0,87 | 0,948 | 0,98 | 0,992 | 0,997 | 0,975 | 0,978 | 0,98 | 0,98 | 0,981 | 0,897 | 0,82 | 0,922 | 0,97 | 0,988 | 0,995 | 0,963 | 0,964 | 0,964 | 0,965 | 0,965 | 0,938 | 0,763 | 0,891 | 0,956 | 0,983 | 0,993 | 0,923 | 0,924 | 0,926 | 0,927 | 0,928 |
| db6 | | 0,809 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,898 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,964 | 0,965 | 0,966 | 0,966 | 0,966 | 0,94 | 0,762 | 0,889 | 0,955 | 0,983 | 0,993 | 0,931 | 0,933 | 0,934 | 0,934 | 0,935 |
| sym6 | | 0,809 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,898 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,965 | 0,965 | 0,966 | 0,966 | 0,967 | 0,94 | 0,763 | 0,89 | 0,955 | 0,983 | 0,993 | 0,932 | 0,933 | 0,934 | 0,935 | 0,936 |
| db7 | 14 | 0,809 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,9 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,964 | 0,965 | 0,966 | 0,966 | 0,966 | 0,941 | 0,761 | 0,889 | 0,955 | 0,983 | 0,993 | 0,934 | 0,936 | 0,937 | 0,937 | 0,938 |
| fk14 | | 0,809 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,98 | 0,98 | 0,981 | 0,898 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,965 | 0,965 | 0,966 | 0,966 | 0,967 | 0,942 | 0,762 | 0,89 | 0,955 | 0,983 | 0,993 | 0,939 | 0,94 | 0,941 | 0,942 | 0,942 |
| sym7 | | 0,809 | 0,87 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,98 | 0,98 | 0,981 | 0,898 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,965 | 0,965 | 0,966 | 0,966 | 0,967 | 0,94 | 0,765 | 0,891 | 0,956 | 0,983 | 0,993 | 0,935 | 0,936 | 0,937 | 0,937 | 0,938 |
| db8 | 16 | 0,809 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,899 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,965 | 0,965 | 0,966 | 0,967 | 0,967 | 0,941 | 0,762 | 0,89 | 0,955 | 0,983 | 0,993 | 0,936 | 0,937 | 0,938 | 0,939 | 0,939 |
| sym8 | | 0,809 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,899 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,965 | 0,965 | 0,966 | 0,966 | 0,966 | 0,941 | 0,762 | 0,89 | 0,955 | 0,983 | 0,993 | 0,936 | 0,937 | 0,938 | 0,938 | 0,939 |
| coif3 | 18 | 0,81 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,899 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,964 | 0,965 | 0,965 | 0,966 | 0,966 | 0,941 | 0,763 | 0,89 | 0,956 | 0,983 | 0,993 | 0,933 | 0,934 | 0,935 | 0,936 | 0,937 |
| db9 | | 0,81 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,899 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,965 | 0,966 | 0,966 | 0,967 | 0,967 | 0,942 | 0,761 | 0,889 | 0,955 | 0,983 | 0,993 | 0,938 | 0,938 | 0,94 | 0,94 | 0,941 |
| fk18 | | 0,81 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,98 | 0,98 | 0,981 | 0,899 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,965 | 0,966 | 0,966 | 0,967 | 0,967 | 0,943 | 0,762 | 0,889 | 0,955 | 0,983 | 0,993 | 0,941 | 0,942 | 0,943 | 0,944 | 0,944 |
| sym9 | | 0,81 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,899 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,965 | 0,965 | 0,966 | 0,966 | 0,967 | 0,941 | 0,763 | 0,89 | 0,956 | 0,983 | 0,993 | 0,937 | 0,938 | 0,939 | 0,94 | 0,941 |
| db10 | 20 | 0,81 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,981 | 0,981 | 0,899 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,965 | 0,965 | 0,966 | 0,966 | 0,967 | 0,942 | 0,762 | 0,89 | 0,955 | 0,983 | 0,993 | 0,938 | 0,939 | 0,94 | 0,941 | 0,941 |
| sym10 | | 0,81 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,9 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,965 | 0,966 | 0,966 | 0,967 | 0,967 | 0,942 | 0,761 | 0,89 | 0,955 | 0,983 | 0,993 | 0,938 | 0,939 | 0,94 | 0,941 | 0,942 |
| db11 | 22 | 0,81 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,981 | 0,981 | 0,9 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,965 | 0,966 | 0,966 | 0,967 | 0,967 | 0,943 | 0,761 | 0,889 | 0,955 | 0,983 | 0,993 | 0,939 | 0,94 | 0,941 | 0,942 | 0,943 |
| fk22 | | 0,811 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,98 | 0,98 | 0,981 | 0,9 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,965 | 0,966 | 0,967 | 0,967 | 0,967 | 0,943 | 0,762 | 0,889 | 0,955 | 0,983 | 0,993 | 0,942 | 0,943 | 0,944 | 0,945 | 0,945 |
| sym11 | | 0,81 | 0,87 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,98 | 0,981 | 0,9 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,965 | 0,966 | 0,966 | 0,967 | 0,967 | 0,942 | 0,765 | 0,89 | 0,956 | 0,983 | 0,993 | 0,939 | 0,94 | 0,941 | 0,941 | 0,942 |
| coif4 | 24 | 0,811 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,9 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,965 | 0,965 | 0,966 | 0,966 | 0,967 | 0,942 | 0,763 | 0,89 | 0,955 | 0,983 | 0,993 | 0,936 | 0,937 | 0,938 | 0,939 | 0,94 |
| db12 | | 0,811 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,981 | 0,981 | 0,9 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,965 | 0,966 | 0,967 | 0,967 | 0,967 | 0,943 | 0,761 | 0,889 | 0,955 | 0,983 | 0,993 | 0,939 | 0,94 | 0,941 | 0,942 | 0,943 |
| sym12 | | 0,811 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,9 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,965 | 0,966 | 0,966 | 0,967 | 0,967 | 0,943 | 0,764 | 0,89 | 0,956 | 0,983 | 0,993 | 0,94 | 0,941 | 0,942 | 0,942 | 0,943 |
| db13 | 26 | 0,811 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,901 | 0,819 | 0,925 | 0,97 | 0,988 | 0,995 | 0,965 | 0,966 | 0,966 | 0,967 | 0,967 | 0,943 | 0,762 | 0,889 | 0,955 | 0,983 | 0,993 | 0,94 | 0,941 | 0,942 | 0,943 | 0,943 |
| sym13 | | 0,811 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,901 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,965 | 0,966 | 0,966 | 0,967 | 0,967 | 0,943 | 0,76 | 0,889 | 0,955 | 0,983 | 0,993 | 0,94 | 0,941 | 0,942 | 0,942 | 0,943 |
| db14 | 28 | 0,812 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,901 | 0,82 | 0,922 | 0,97 | 0,988 | 0,995 | 0,965 | 0,966 | 0,966 | 0,967 | 0,967 | 0,944 | 0,761 | 0,89 | 0,955 | 0,983 | 0,993 | 0,941 | 0,941 | 0,942 | 0,943 | 0,944 |
| sym14 | | 0,812 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,901 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,967 | 0,944 | 0,763 | 0,89 | 0,955 | 0,983 | 0,993 | 0,94 | 0,941 | 0,942 | 0,943 | 0,944 |
| coif5 | 30 | 0,812 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,901 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,965 | 0,965 | 0,966 | 0,966 | 0,967 | 0,944 | 0,763 | 0,89 | 0,955 | 0,983 | 0,993 | 0,939 | 0,94 | 0,941 | 0,942 | 0,942 |
| db15 | | 0,812 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,901 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,965 | 0,966 | 0,967 | 0,967 | 0,967 | 0,944 | 0,762 | 0,889 | 0,955 | 0,983 | 0,993 | 0,941 | 0,942 | 0,943 | 0,943 | 0,944 |
| sym15 | | 0,812 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,981 | 0,981 | 0,901 | 0,819 | 0,923 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,967 | 0,944 | 0,764 | 0,891 | 0,956 | 0,983 | 0,993 | 0,941 | 0,942 | 0,943 | 0,943 | 0,944 |
| db16 | 32 | 0,812 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,902 | 0,819 | 0,923 | 0,97 | 0,988 | 0,995 | 0,965 | 0,966 | 0,966 | 0,967 | 0,967 | 0,945 | 0,761 | 0,889 | 0,955 | 0,983 | 0,993 | 0,941 | 0,942 | 0,943 | 0,944 | 0,945 |
| sym16 | | 0,812 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,98 | 0,98 | 0,981 | 0,902 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,965 | 0,966 | 0,966 | 0,967 | 0,967 | 0,944 | 0,762 | 0,89 | 0,955 | 0,983 | 0,993 | 0,941 | 0,942 | 0,943 | 0,943 | 0,944 |
| db17 | 34 | 0,813 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,98 | 0,981 | 0,981 | 0,902 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,965 | 0,966 | 0,966 | 0,967 | 0,967 | 0,945 | 0,762 | 0,89 | 0,955 | 0,983 | 0,993 | 0,941 | 0,942 | 0,943 | 0,944 | 0,944 |
| sym17 | | 0,812 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,902 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,965 | 0,966 | 0,967 | 0,967 | 0,967 | 0,944 | 0,765 | 0,89 | 0,955 | 0,983 | 0,993 | 0,941 | 0,942 | 0,943 | 0,944 | 0,944 |

| | | Número de etapas ($\xi$) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | | | | | | | | | | | 2 | | | | | | | | | | | 3 | | | | | | | | | | |
| Familia wavelet | Lf | G | Uniforme | | | | | No Uniforme | | | | | G | Uniforme | | | | | No Uniforme | | | | | G | Uniforme | | | | | No Uniforme | | | | |
| | | | 8 | 16 | 32 | 64 | 128 | 8 | 16 | 32 | 64 | 128 | | 8 | 16 | 32 | 64 | 128 | 8 | 16 | 32 | 64 | 128 | | 8 | 16 | 32 | 64 | 128 | 8 | 16 | 32 | 64 | 128 |
| db18 | 36 | 0,813 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,981 | 0,981 | 0,903 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,945 | 0,761 | 0,889 | 0,955 | 0,983 | 0,993 | 0,942 | 0,943 | 0,944 | 0,944 | 0,945 |
| sym18 | | 0,813 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,903 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,945 | 0,763 | 0,89 | 0,955 | 0,983 | 0,993 | 0,942 | 0,942 | 0,943 | 0,944 | 0,945 |
| db19 | 38 | 0,813 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,981 | 0,981 | 0,903 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,965 | 0,966 | 0,966 | 0,967 | 0,967 | 0,946 | 0,761 | 0,889 | 0,955 | 0,983 | 0,993 | 0,942 | 0,942 | 0,943 | 0,944 | 0,945 |
| sym19 | | 0,813 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,98 | 0,981 | 0,903 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,945 | 0,761 | 0,89 | 0,955 | 0,983 | 0,993 | 0,942 | 0,942 | 0,943 | 0,944 | 0,945 |
| db20 | 40 | 0,814 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,981 | 0,981 | 0,903 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,965 | 0,966 | 0,967 | 0,967 | 0,967 | 0,946 | 0,761 | 0,889 | 0,955 | 0,983 | 0,993 | 0,942 | 0,943 | 0,944 | 0,945 | 0,945 |
| sym20 | | 0,814 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,98 | 0,98 | 0,981 | 0,903 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,965 | 0,966 | 0,966 | 0,967 | 0,967 | 0,946 | 0,763 | 0,89 | 0,955 | 0,983 | 0,993 | 0,942 | 0,943 | 0,944 | 0,944 | 0,945 |
| db21 | 42 | 0,814 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,904 | 0,82 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,946 | 0,761 | 0,89 | 0,955 | 0,983 | 0,993 | 0,942 | 0,943 | 0,944 | 0,944 | 0,945 |
| sym21 | | 0,814 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,903 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,946 | 0,764 | 0,89 | 0,956 | 0,983 | 0,993 | 0,942 | 0,943 | 0,944 | 0,944 | 0,945 |
| db22 | 44 | 0,814 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,904 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,946 | 0,761 | 0,889 | 0,955 | 0,983 | 0,993 | 0,942 | 0,943 | 0,944 | 0,945 | 0,945 |
| sym22 | | 0,814 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,98 | 0,98 | 0,981 | 0,904 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,946 | 0,762 | 0,89 | 0,955 | 0,983 | 0,993 | 0,942 | 0,943 | 0,944 | 0,944 | 0,945 |
| db23 | 46 | 0,815 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,904 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,965 | 0,966 | 0,967 | 0,967 | 0,967 | 0,947 | 0,761 | 0,889 | 0,955 | 0,983 | 0,993 | 0,942 | 0,943 | 0,944 | 0,945 | 0,945 |
| sym23 | | 0,815 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,98 | 0,981 | 0,905 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,947 | 0,761 | 0,89 | 0,955 | 0,983 | 0,993 | 0,942 | 0,943 | 0,944 | 0,945 | 0,945 |
| db24 | 48 | 0,815 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,98 | 0,98 | 0,981 | 0,905 | 0,82 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,947 | 0,761 | 0,889 | 0,955 | 0,983 | 0,993 | 0,942 | 0,943 | 0,944 | 0,945 | 0,945 |
| sym24 | | 0,815 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,98 | 0,98 | 0,981 | 0,905 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,965 | 0,966 | 0,966 | 0,967 | 0,967 | 0,947 | 0,763 | 0,89 | 0,955 | 0,983 | 0,993 | 0,942 | 0,943 | 0,944 | 0,944 | 0,945 |
| db25 | 50 | 0,815 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,981 | 0,981 | 0,905 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,947 | 0,761 | 0,889 | 0,955 | 0,983 | 0,993 | 0,943 | 0,943 | 0,944 | 0,945 | 0,946 |
| sym25 | | 0,815 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,905 | 0,819 | 0,923 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,947 | 0,764 | 0,89 | 0,956 | 0,983 | 0,993 | 0,942 | 0,943 | 0,944 | 0,945 | 0,945 |
| db26 | 52 | 0,816 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,981 | 0,981 | 0,905 | 0,82 | 0,922 | 0,97 | 0,988 | 0,995 | 0,965 | 0,966 | 0,967 | 0,967 | 0,968 | 0,948 | 0,76 | 0,889 | 0,955 | 0,983 | 0,993 | 0,942 | 0,943 | 0,944 | 0,945 | 0,945 |
| sym26 | | 0,816 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,906 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,948 | 0,762 | 0,89 | 0,955 | 0,983 | 0,993 | 0,942 | 0,943 | 0,944 | 0,945 | 0,945 |
| db27 | 54 | 0,816 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,981 | 0,981 | 0,906 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,948 | 0,761 | 0,888 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,944 | 0,945 | 0,946 |
| sym27 | | 0,816 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,981 | 0,981 | 0,905 | 0,82 | 0,922 | 0,97 | 0,988 | 0,995 | 0,965 | 0,966 | 0,967 | 0,967 | 0,968 | 0,948 | 0,763 | 0,89 | 0,955 | 0,983 | 0,993 | 0,942 | 0,943 | 0,944 | 0,945 | 0,945 |
| db28 | 56 | 0,816 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,981 | 0,981 | 0,906 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,948 | 0,76 | 0,889 | 0,955 | 0,983 | 0,993 | 0,943 | 0,943 | 0,944 | 0,945 | 0,946 |
| sym28 | | 0,816 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,98 | 0,98 | 0,981 | 0,906 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,967 | 0,948 | 0,763 | 0,89 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,944 | 0,945 | 0,946 |
| db29 | 58 | 0,817 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,981 | 0,98 | 0,981 | 0,906 | 0,82 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,948 | 0,761 | 0,889 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,945 | 0,945 | 0,946 |
| sym29 | | 0,817 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,906 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,948 | 0,763 | 0,89 | 0,955 | 0,983 | 0,993 | 0,943 | 0,943 | 0,944 | 0,945 | 0,945 |
| db30 | 60 | 0,817 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,907 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,949 | 0,76 | 0,889 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,944 | 0,945 | 0,946 |
| sym30 | | 0,817 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,98 | 0,98 | 0,981 | 0,907 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,949 | 0,763 | 0,89 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,944 | 0,945 | 0,946 |
| db31 | 62 | 0,817 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,98 | 0,98 | 0,981 | 0,907 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,949 | 0,76 | 0,889 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,945 | 0,945 | 0,946 |
| sym31 | | 0,818 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,981 | 0,981 | 0,907 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,949 | 0,763 | 0,89 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,944 | 0,945 | 0,946 |
| db32 | 64 | 0,818 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,98 | 0,981 | 0,907 | 0,82 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,949 | 0,76 | 0,888 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,945 | 0,945 | 0,946 |
| sym32 | | 0,818 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,98 | 0,98 | 0,981 | 0,907 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,949 | 0,762 | 0,89 | 0,955 | 0,983 | 0,993 | 0,943 | 0,943 | 0,944 | 0,945 | 0,946 |
| db33 | 66 | 0,818 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,981 | 0,981 | 0,908 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,949 | 0,76 | 0,888 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,945 | 0,945 | 0,946 |
| sym33 | | 0,818 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,908 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,949 | 0,762 | 0,891 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,945 | 0,945 | 0,946 |
| db34 | 68 | 0,819 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,981 | 0,981 | 0,908 | 0,82 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,95 | 0,76 | 0,888 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,945 | 0,945 | 0,946 |
| sym34 | | 0,819 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,98 | 0,98 | 0,981 | 0,908 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,95 | 0,762 | 0,89 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,945 | 0,945 | 0,946 |
| db35 | 70 | 0,819 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,981 | 0,981 | 0,909 | 0,82 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,95 | 0,76 | 0,889 | 0,954 | 0,983 | 0,993 | 0,943 | 0,944 | 0,945 | 0,945 | 0,946 |
| sym35 | | 0,819 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,98 | 0,981 | 0,908 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,95 | 0,763 | 0,89 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,944 | 0,945 | 0,946 |
| db36 | 72 | 0,819 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,981 | 0,981 | 0,909 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,95 | 0,76 | 0,888 | 0,954 | 0,983 | 0,993 | 0,943 | 0,944 | 0,945 | 0,946 | 0,946 |
| sym36 | | 0,819 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,98 | 0,98 | 0,981 | 0,909 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,95 | 0,762 | 0,89 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,945 | 0,945 | 0,946 |
| db37 | 74 | 0,819 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,98 | 0,98 | 0,981 | 0,909 | 0,82 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,95 | 0,759 | 0,888 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,945 | 0,945 | 0,946 |
| sym37 | | 0,819 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,98 | 0,98 | 0,981 | 0,909 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,95 | 0,763 | 0,889 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,945 | 0,945 | 0,946 |
| db38 | 76 | 0,82 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,975 | 0,979 | 0,98 | 0,98 | 0,981 | 0,909 | 0,82 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,951 | 0,759 | 0,888 | 0,954 | 0,982 | 0,993 | 0,943 | 0,944 | 0,945 | 0,946 | 0,946 |
| sym38 | | 0,82 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,98 | 0,98 | 0,981 | 0,909 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,951 | 0,763 | 0,89 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,945 | 0,945 | 0,946 |
| db39 | 78 | 0,82 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,98 | 0,98 | 0,981 | 0,909 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,951 | 0,76 | 0,888 | 0,955 | 0,982 | 0,993 | 0,943 | 0,944 | 0,945 | 0,945 | 0,946 |
| sym39 | | 0,82 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,98 | 0,981 | 0,909 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,951 | 0,762 | 0,89 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,945 | 0,945 | 0,946 |

| | | Número de etapas ($\xi$) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | | | | | | | | | | | 2 | | | | | | | | | | 3 | | | | | | | | | | |
| Familia wavelet | Lf | G | Uniforme | | | | | No Uniforme | | | | | G | Uniforme | | | | | No Uniforme | | | | | G | Uniforme | | | | | No Uniforme | | | | |
| | | | 8 | 16 | 32 | 64 | 128 | 8 | 16 | 32 | 64 | 128 | | 8 | 16 | 32 | 64 | 128 | 8 | 16 | 32 | 64 | 128 | | 8 | 16 | 32 | 64 | 128 | 8 | 16 | 32 | 64 | 128 |
| db40 | 80 | 0,82 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,981 | 0,981 | 0,91 | 0,82 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,951 | 0,759 | 0,888 | 0,955 | 0,982 | 0,993 | 0,943 | 0,944 | 0,945 | 0,946 | 0,946 |
| sym40 | | 0,82 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,98 | 0,981 | 0,91 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,951 | 0,762 | 0,89 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,945 | 0,945 | 0,946 |
| db41 | 82 | 0,821 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,981 | 0,981 | 0,91 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,951 | 0,759 | 0,888 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,945 | 0,946 | 0,946 |
| sym41 | | 0,821 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,98 | 0,98 | 0,981 | 0,91 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,951 | 0,763 | 0,89 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,945 | 0,945 | 0,946 |
| db42 | 84 | 0,821 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,981 | 0,981 | 0,91 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,952 | 0,759 | 0,888 | 0,954 | 0,982 | 0,993 | 0,943 | 0,944 | 0,945 | 0,946 | 0,946 |
| sym42 | | 0,821 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,98 | 0,981 | 0,91 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,951 | 0,762 | 0,89 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,945 | 0,945 | 0,946 |
| db43 | 86 | 0,821 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,981 | 0,981 | 0,911 | 0,82 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,952 | 0,759 | 0,888 | 0,954 | 0,982 | 0,993 | 0,943 | 0,944 | 0,945 | 0,946 | 0,946 |
| sym43 | | 0,822 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,98 | 0,981 | 0,911 | 0,819 | 0,922 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,952 | 0,762 | 0,89 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,945 | 0,945 | 0,946 |
| db44 | 88 | 0,822 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,981 | 0,981 | 0,911 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,952 | 0,759 | 0,888 | 0,954 | 0,982 | 0,993 | 0,943 | 0,944 | 0,945 | 0,946 | 0,946 |
| sym44 | | 0,822 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,981 | 0,98 | 0,981 | 0,911 | 0,819 | 0,923 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,952 | 0,762 | 0,89 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,945 | 0,946 | 0,946 |
| db45 | 90 | 0,822 | 0,869 | 0,948 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,98 | 0,98 | 0,981 | 0,911 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,952 | 0,759 | 0,888 | 0,955 | 0,982 | 0,993 | 0,943 | 0,944 | 0,945 | 0,946 | 0,946 |
| sym45 | | 0,822 | 0,869 | 0,947 | 0,98 | 0,992 | 0,997 | 0,976 | 0,979 | 0,98 | 0,98 | 0,981 | 0,911 | 0,82 | 0,923 | 0,97 | 0,988 | 0,995 | 0,966 | 0,966 | 0,967 | 0,967 | 0,968 | 0,952 | 0,763 | 0,889 | 0,955 | 0,983 | 0,993 | 0,943 | 0,944 | 0,945 | 0,945 | 0,946 |

## F.2. Medidas Objetivas Cuantificación Uniforme en el Dominio del Tiempo

En la Tabla F.7 se muestran los resultados obtenidos para las 5 medidas objetivas al realizar la cuantificación uniforme en el dominio del tiempo. Los resultados obtenidos con este tipo de cuantificación son más cercanos a los de la cuantificación uniforme de los coeficientes en el dominio *wavelet*, sin embargo, para esta cuantificación se obtienen valores ligeramente superiores.

Tabla F.7. Medidas objetivas cuantificación en el dominio del tiempo.
Elaboración propia.

| N | M-NRMSE | MAE | SNR | $\rho$ | SSMIP |
|---|---|---|---|---|---|
| 8 | 0.9029 | 0.0142 | 33.744 | 0.9804 | 0.862 |
| 16 | 0.9514 | 0.0071 | ∞ | 0.9949 | 0.9525 |
| 32 | 0.9758 | 0.0035 | ∞ | 0.9987 | 0.9867 |
| 64 | 0.9881 | 0.0018 | ∞ | 0.9997 | 0.9966 |
| 128 | 0.9943 | 0.0009 | ∞ | 0.9999 | 0.9992 |

## F.3. Medidas Subjetivas

Para la evaluación subjetiva de la calidad de las señales de voz reconstruidas se utilizó la MOS. En total se plantearon 6 pruebas y cada prueba está compuesta por 3 versiones de un mismo audio. En la Tabla F.8 se muestran los resultados de los 27 evaluadores.

Tabla F.8. Resultados medidas subjetivas.
Elaboración propia.

| MOS | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Evaluador | Prueba | | | | | | | | | | | | | | | | | |
| | 1 | | | 2 | | | 3 | | | 4 | | | 5 | | | 6 | | |
| 1 | 4 | 3 | 3 | 5 | 4 | 4 | 1 | 2 | 5 | 1 | 2 | 5 | 4 | 5 | 5 | 4 | 5 | 5 |
| 2 | 5 | 5 | 5 | 5 | 5 | 5 | 3 | 3 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 3 | 4 | 3 | 3 | 4 | 4 | 3 | 1 | 1 | 3 | 2 | 2 | 4 | 3 | 3 | 4 | 4 | 4 | 5 |
| 4 | 4 | 3 | 3 | 4 | 4 | 3 | 1 | 1 | 3 | 2 | 2 | 4 | 3 | 3 | 4 | 4 | 4 | 5 |
| 5 | 4 | 3 | 2 | 4 | 3 | 2 | 1 | 1 | 5 | 2 | 2 | 4 | 3 | 3 | 4 | 4 | 4 | 4 |
| 6 | 5 | 5 | 5 | 5 | 5 | 4 | 3 | 4 | 5 | 3 | 3 | 5 | 4 | 4 | 5 | 5 | 5 | 5 |
| 7 | 5 | 4 | 4 | 5 | 3 | 2 | 1 | 1 | 4 | 1 | 2 | 5 | 3 | 3 | 4 | 5 | 5 | 5 |
| 8 | 4 | 4 | 4 | 5 | 4 | 4 | 2 | 3 | 5 | 2 | 3 | 5 | 4 | 5 | 5 | 4 | 5 | 5 |
| 9 | 5 | 4 | 3 | 4 | 4 | 3 | 1 | 1 | 4 | 1 | 2 | 5 | 3 | 3 | 5 | 4 | 5 | 5 |
| 10 | 5 | 5 | 5 | 5 | 5 | 5 | 2 | 3 | 5 | 3 | 3 | 5 | 4 | 5 | 5 | 5 | 5 | 5 |
| 11 | 3 | 2 | 1 | 3 | 2 | 2 | 1 | 1 | 3 | 1 | 1 | 3 | 2 | 2 | 3 | 3 | 4 | 4 |
| 12 | 4 | 3 | 3 | 4 | 4 | 3 | 1 | 1 | 3 | 2 | 2 | 4 | 3 | 3 | 4 | 4 | 4 | 5 |
| 13 | 4 | 3 | 3 | 5 | 4 | 4 | 1 | 2 | 5 | 1 | 2 | 5 | 4 | 5 | 5 | 4 | 5 | 5 |
| 14 | 5 | 4 | 4 | 5 | 3 | 2 | 1 | 1 | 4 | 1 | 2 | 5 | 3 | 3 | 4 | 5 | 5 | 5 |
| 15 | 4 | 4 | 4 | 5 | 4 | 4 | 2 | 3 | 4 | 1 | 2 | 5 | 3 | 3 | 5 | 4 | 5 | 5 |
| 16 | 4 | 3 | 3 | 4 | 4 | 3 | 1 | 1 | 4 | 1 | 2 | 5 | 3 | 3 | 4 | 5 | 5 | 5 |
| 17 | 4 | 3 | 3 | 5 | 4 | 4 | 1 | 2 | 5 | 1 | 2 | 5 | 4 | 5 | 5 | 4 | 5 | 5 |
| 18 | 5 | 4 | 4 | 5 | 3 | 2 | 1 | 1 | 4 | 1 | 2 | 5 | 3 | 3 | 4 | 5 | 5 | 5 |
| 19 | 4 | 3 | 2 | 4 | 3 | 2 | 1 | 1 | 3 | 1 | 1 | 3 | 2 | 2 | 3 | 3 | 4 | 4 |
| 20 | 4 | 3 | 3 | 4 | 4 | 3 | 1 | 1 | 5 | 1 | 2 | 5 | 4 | 5 | 5 | 4 | 5 | 5 |
| 21 | 4 | 3 | 3 | 4 | 4 | 3 | 1 | 1 | 3 | 2 | 2 | 4 | 3 | 3 | 4 | 4 | 4 | 5 |
| 22 | 4 | 4 | 4 | 5 | 4 | 4 | 2 | 3 | 4 | 1 | 2 | 5 | 3 | 3 | 5 | 4 | 5 | 5 |
| 23 | 5 | 4 | 4 | 5 | 3 | 2 | 1 | 1 | 4 | 2 | 2 | 4 | 3 | 3 | 4 | 4 | 4 | 5 |
| 24 | 4 | 3 | 3 | 4 | 4 | 3 | 1 | 1 | 3 | 2 | 2 | 4 | 3 | 3 | 4 | 4 | 4 | 5 |
| 25 | 5 | 5 | 5 | 5 | 5 | 4 | 3 | 4 | 5 | 3 | 3 | 5 | 4 | 4 | 5 | 5 | 5 | 5 |
| 26 | 5 | 4 | 4 | 5 | 3 | 2 | 1 | 1 | 4 | 2 | 2 | 4 | 3 | 3 | 4 | 4 | 4 | 5 |
| 27 | 4 | 3 | 3 | 4 | 4 | 3 | 1 | 1 | 3 | 2 | 2 | 4 | 3 | 3 | 4 | 4 | 4 | 5 |
| **Media** | 4,3333 | 3,5926 | 3,4444 | 4,5185 | 3,8148 | 3,1481 | 1,3704 | 1,7037 | 4,037 | 1,7037 | 2,1481 | 4,5185 | 3,2963 | 3,5185 | 4,3704 | 4,2222 | 4,5926 | 4,8889 |
| **Varianza** | 0,3077 | 0,6353 | 0,9487 | 0,3362 | 0,5413 | 0,9003 | 0,4729 | 1,0627 | 0,6524 | 0,6781 | 0,3618 | 0,4131 | 0,4473 | 0,9516 | 0,396 | 0,3333 | 0,2507 | 0,1026 |