CUANTIFICACIÓN DE SEÑALES DE VOZ EN EL DOMINIO WAVELET UTILIZANDO ESQUEMA LIFTING

Universidad
del Cauca

Trabajo de Grado

Lina Virginia Muñoz Garcés
Jhon Fredy Romero Núñez

Directora:
MSc. María Manuela Silva Zambrano

Universidad del Cauca
Facultad de Ingeniería Electrónica y Telecomunicaciones
Departamento de Telecomunicaciones
Grupo de Nuevas Tecnologías en Telecomunicaciones - GNTT
Popayán, Septiembre de 2023

Dedicado a la mejor directora de trabajo de grado
y más grande investigadora
que hemos conocido.

CONTENIDO

LISTA DE TABLAS

LISTA DE FIGURAS

LISTA DE ACRÓNIMOS

ACRM *Absolute Category Rating Methods*. 11, 59, 60

ADC *Analog to Digital Conversion*. 2

BDM *Bark Distortion Measures*. 18, 58

CCR *Comparison Category Rating*. 11

CMOS *Comparison Mean Opinion Score*. 12, 60, 88, 98, 120, 121, 123–125

CRR *Coefficient Relevance Rate*. 47, 49, 50, 52, 55, 56, 60–62, 64, 68, 69, 71–75, 77–79, 82, 87, 101–103, 119

CWT *Continuous Wavelet Transform*. 22, 24, 25

DAM *Diagnostic Acceptability Measure*. 13, 60

DCT *Discrete Cosine Transform*. 9

DFT *Discrete Fourier Transform*. 9

DWT *Discrete Wavelet Transform*. 10, 25–27, 36, 60, 120

FT *Fourier Transform*. 3, 21

FWT *Fast Wavelet Transform*. 28

ITU *International Telecommunication Union*. 59, 120

MOS *Mean Opinion Score*. 12–14, 18, 60, 90, 91, 98, 120–122, 125

MRA *MultiResolution Analysis*. 27, 28

MSE *Mean Square Error*. 10, 15, 16, 58, 59

MSSIM *Mean Structural SIMilarity*. 17, 58

NMSE *Normalized Mean Square Error*. 16, 44, 58, 59, 63–65, 78

PESQ *Perceptual Evaluation of Speech Quality*. 18, 19, 44, 58, 59, 78

PMM *Perceptually-Motivated Measures*. 17

RPM *Relative Preference Methods*. 11, 12, 59, 60

SBI *Scale of Background Intrusiveness*. 14, 60

SSD *Scale of Signal Distortion*. 14

SSNR *Segmental Signal to Noise Ratio*. 15, 16, 58

STFT *Short-time Fourier Transform*. 21

STMESCSINSA *Subjective Test Methodology for Evaluating Speech Communication Systems that Include Noise Suppression Algorithm*. 14, 60

WT *Wavelet Transform*. 21, 26

CAPÍTULO 1

CONVERSIÓN ANALÓGICA-DIGITAL DE SEÑALES DE VOZ

En las telecomunicaciones es fundamental el proceso de conversión de una señal analógica a una señal digital, ya que estas últimas son más fáciles de almacenar y manipular, facilitando una gran variedad de técnicas de transmisión, las cuales buscan ser más robustas frente al ruido y la interferencia. La estimación de la señal transmitida, en el receptor, requiere de un proceso de decisión en un sistema digital, puesto que, a diferencia de un sistema analógico, la señal tiene un número finito de posibilidades para sus valores de amplitud, gracias a lo cual se puede discriminar en mayor medida el efecto del ruido [2].

Este proceso de Conversión Análogo-Digital (ADC, *Analog to Digital Conversion*) tiene, usualmente, tres etapas: muestreo, cuantificación y codificación de fuente. En síntesis, un ADC realiza el paso de una señal continua, tanto en tiempo como en amplitud, a una señal discreta en estos dos parámetros y, finalmente, representa los valores obtenidos mediante una secuencia binaria [3].

Si bien, hay una gran ventaja al digitalizar las señales analógicas, particularmente las señales de voz, en las telecomunicaciones siempre se debe tener en cuenta cuáles son las compensaciones a cambio del beneficio, en este caso, la digitalización genera una pérdida de información ineludible, debido a lo cual, en el diseño de un cuantificador, se busca minimizar la diferencia entre la señal discreta y la original. Por esta razón se deben comprender y aplicar los principios subyacentes a estos procesos [2].

1.1. Muestreo

El muestreo es el primer paso en el proceso de convertir una señal analógica, como las señales de voz, en una señal digital, para ello se genera la discretización de la señal en el tiempo tomando muestras temporales; sin embargo, este proceso no se puede hacer de forma caprichosa, puesto que se debe tener en cuenta una frecuencia de muestreo específica que permita, posteriormente, reconstruir la señal de manera fidedigna. El valor de la frecuencia de muestreo depende del ancho de banda de la señal, por lo que, para casos como el de las señales de voz, se deben tomar decisiones sobre su ancho de banda, i.e., previamente al proceso de muestreo se debe limitar en banda a la señal.

En la Figura 1.1, se explica el proceso de muestreo y la reconstrucción de la señal en el dominio del tiempo. Siendo $x\left(t\right)$ la señal original, la cual es multiplicada por un tren de impulsos $p\left(t\right)$, que tiene una frecuencia de muestreo F_s, obteniendo como resultado de la multiplicación la versión muestreada, $x\left[n\right]$, de la señal original. Adicionalmente se encuentra el proceso de reconstrucción de la señal, en el cual $x\left[n\right]$ pasa por un filtro pasa-bajo (interpolador), para obtener una señal reconstruida, $x'\left(t\right)$, lo más similar posible a la original.

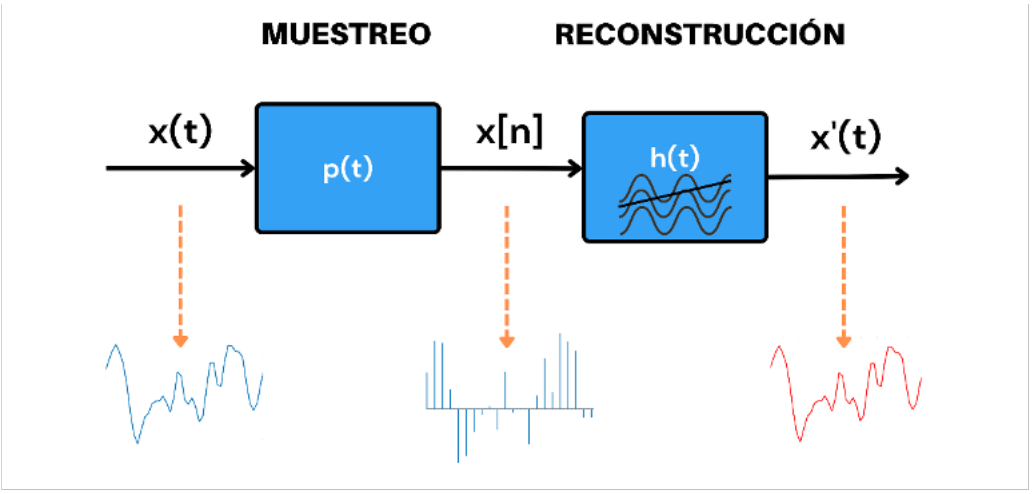


Figura 1.1: Muestreo y reconstrucción.

Matemáticamente, el tren de impulsos $p\left(t\right)$, está descrito de la siguiente manera,

$$p\left(t\right)=\sum_{n=-\infty}^{\infty}\delta\left(t-nT_{s}\right), \tag{1.1}$$

siendo T_s el periodo de la señal.

Es decir, que la señal muestreada está dada por,

$$x\left[n\right]=x\left(t\right)p\left(t\right)=\sum_{n=-\infty}^{\infty}x\left(nT_{s}\right)\delta\left(t-nT_{s}\right). \tag{1.2}$$

En cuanto al dominio de la frecuencia y gracias al análisis realizado con la Transformada de Fourier (FT, *Fourier Transform*) se sabe que:

- Una señal en banda base continua en el tiempo, como lo es la señal original $x\left(t\right)$, tiene un espectro centrado en el origen, sin réplicas espectrales.
- Un tren de impulsos en el tiempo, $p\left(t\right)$, también es un tren de impulsos en el dominio de la frecuencia, esto es,

$$\widetilde{p}\left(f\right)=F_{s}\sum_{l=-\infty}^{\infty}\delta\left(f-lF_{s}\right).$$

- La multiplicación en el dominio del tiempo de dos señales, $x\left(t\right)p\left(t\right)$, corresponde a la convolución de sus espectros en el dominio de la frecuencia, $\widetilde{x}\left(f\right)*\widetilde{p}\left(f\right)$.

- La convolución de cualquier señal con un impulso desplazado en a, resulta en la señal desplazada, $\widetilde{x}\left(f-a\right)=\widetilde{x}\left(f\right)*\delta\left(f-a\right)$.

Entonces, al aplicarle la transformada de Fourier a la señal muestreada se obtiene $\widetilde{x}_n\left(f\right)$, que está compuesta por réplicas espectrales de la señal original, localizadas en múltiplos enteros de F_s.

Es por ello que, para poder obtener una representación confiable de la señal original, es imperativo tener en cuenta la frecuencia a la que se muestrea y cumplir con el límite impuesto por el teorema de muestreo de Nyquist-Shannon, el cual especifica que es posible recuperar una señal original a partir de sus muestras, siempre y cuando la señal esté limitada en banda a w Hz y su frecuencia de muestreo F_s sea por lo menos igual a $2w$ [4].

En el contexto particular de señales de voz, se debe tener en cuenta que ancho de banda de una señal de voz humana es de aproximadamente 8 KHz (w), lo que quiere decir que la frecuencia de muestreo mínima para cumplir con el límite de Nyquist es de 16 KHz ($2w$).

Matemáticamente, se tiene que, la transformada de Fourier de la señal muestreada $\mathcal{F}\left\{x\left[n\right]\right\}$ está dada por,

$$\widetilde{x}_n\left(f\right)=\widetilde{x}\left(f\right)*\left\{F_s\sum_{l=-\infty}^{\infty}\delta\left(f-lF_s\right)\right\}=F_s\sum_{l=-\infty}^{\infty}\widetilde{x}\left(f-lF_s\right). \tag{1.3}$$

Cuando se usan frecuencias mayores a la de Nyquist ($F_s>2w$) se dice que la señal está sobre-muestreada, lo cual, según la ecuación 1.3, distancia las réplicas espectrales y permite utilizar filtros reales en la etapa de interpolación [5].

Por otro lado, cuando se realiza un diezmado con una frecuencia menor al límite de Nyquist ($F_s<2w$) corresponde a sub-muestrear, lo cual impide que la señal pueda ser reconstruida nuevamente, debido a que, al tener una frecuencia menor al doble del ancho de banda de la señal, las réplicas espectrales se comienzan a traslapar, generando el fenómeno de *Aliasing* que produce una distorsión imposible de eliminar. En la Figura 1.2 se observa una señal de voz sub-muestreada, lo cual genera el fenómeno de *Aliasing*, como se evidencia en la Figura 1.3, dado que las réplicas del espectro de la señal original se están superponiendo y, por lo tanto, no se puede recuperar el espectro de la señal original con un filtro pasa-bajo.

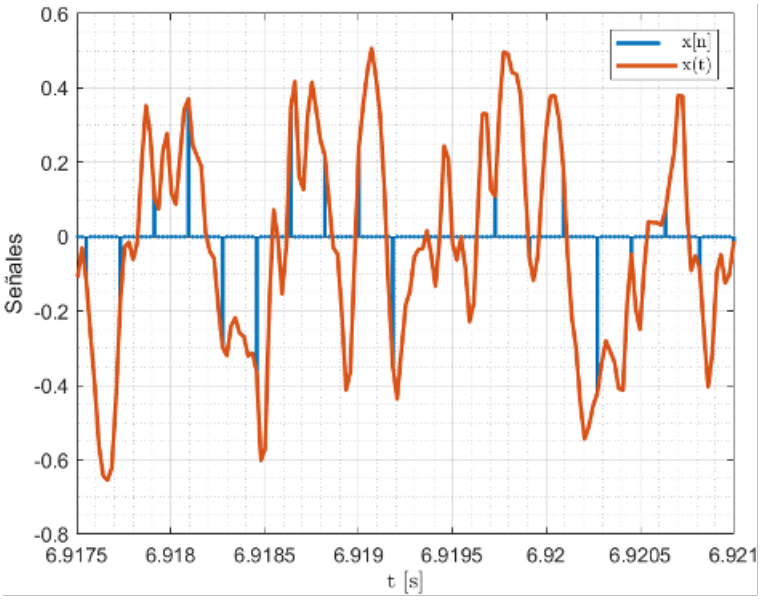


Figura 1.2: Sub-muestreo de una señal.

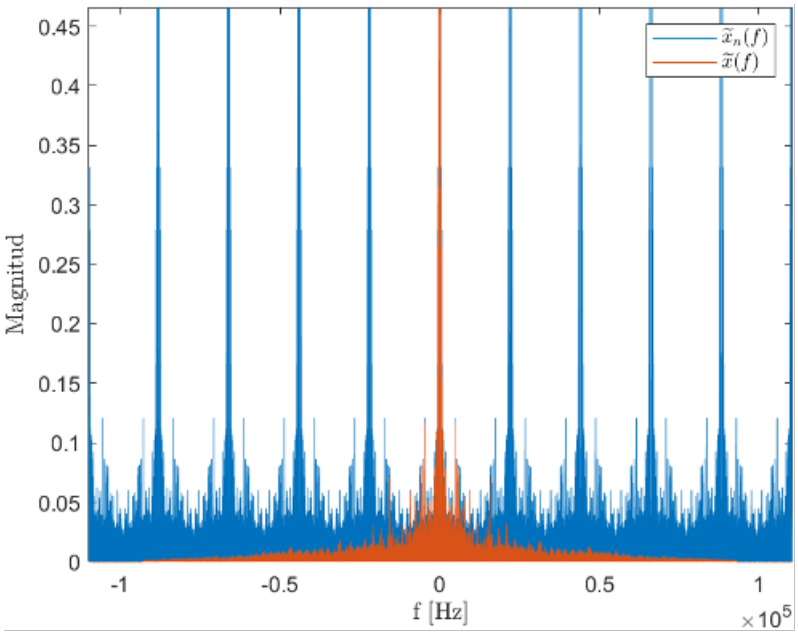


Figura 1.3: Espectro de la señal con fenómeno de *Aliasing*.

1.2. Cuantificación

La cuantificación es el proceso que se realiza sobre una señal de tiempo discreto, a la cual se le restringen sus valores de amplitud [6]. Dicho de otro modo, es el proceso de pasar de una señal con infinitos valores de amplitud posibles, o con un número N de valores posibles de amplitud, a una señal limitada a M amplitudes, siendo $M < N$. Cada uno de los valores de amplitud asignados en este proceso corresponde a un nivel de cuantificación y al conjunto de niveles asignados se le conoce como *alfabeto de cuantificación* [6].

En el marco de las telecomunicaciones el proceso de cuantificación es fundamental, especialmente en la digitalización de las señales, debido a que facilita la transmisión y el procesamiento de la información contenida en dichas señales, dado que el costo computacional disminuye al reducir su número de valores de amplitud.

El proceso de cuantificación tiene ligada una pérdida inherente de información, la cual se puede ver en la diferencia entre las dos señales, i.e., se puede abordar como un ruido que modifica los valores de la señal original, por lo que el cuantificador se debe diseñar buscando disminuir esta pérdida de información para un número dado de niveles de cuantificación, M [7].

La cuantificación se puede abordar de diferentes formas, en la Figura 1.4 se presenta una clasificación según su: dimensión (número de muestras a cuantificar), región (tamaño de las regiones) y representación (dominio en el que se encuentra la señal).

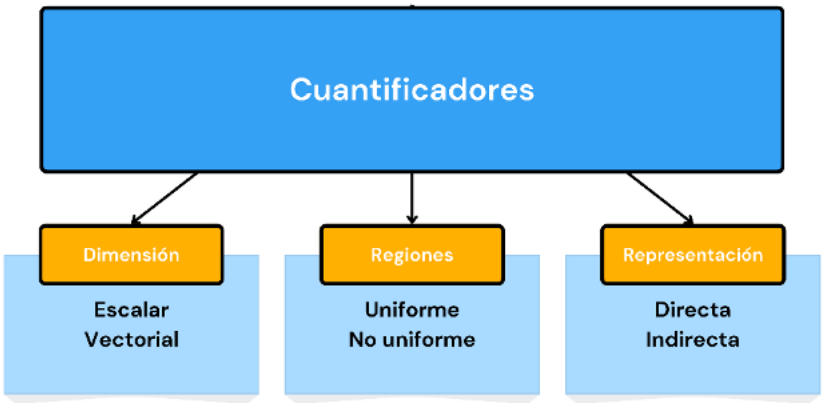


Figura 1.4: Clasificación de cuantificadores.

1.2.1. Cuantificadores Según su Dimensión

Los cuantificadores se pueden clasificar por su dimensión, dependiendo del número de muestras que se clasifican de forma simultánea. Si al cuantificador entra cada muestra de forma independiente es un **cuantificador escalar**; mientras que si a la entrada se tiene un conjunto de muestras (n-tuplas), que son mapeadas en un conjunto más pequeño de valores llamado *codebook*, se hace referencia a un **cuantificador vectorial**, esto genera una señal aún más comprimida después de la cuantificación [8]. El uso de cualquiera de estos dos tipos de cuantificadores (escalar o vectorial) está supeditado a las necesidades del sistema, por ejemplo, qué tanta tolerancia a la distorsión se permite, o qué tan complejo puede llegar a ser el sistema que se quiere. En este trabajo de grado se hace uso de la cuantificación escalar, debido a que el principal objetivo es reducir la distorsión de la señal resultante por medio de un proceso adicional de transformación [9].

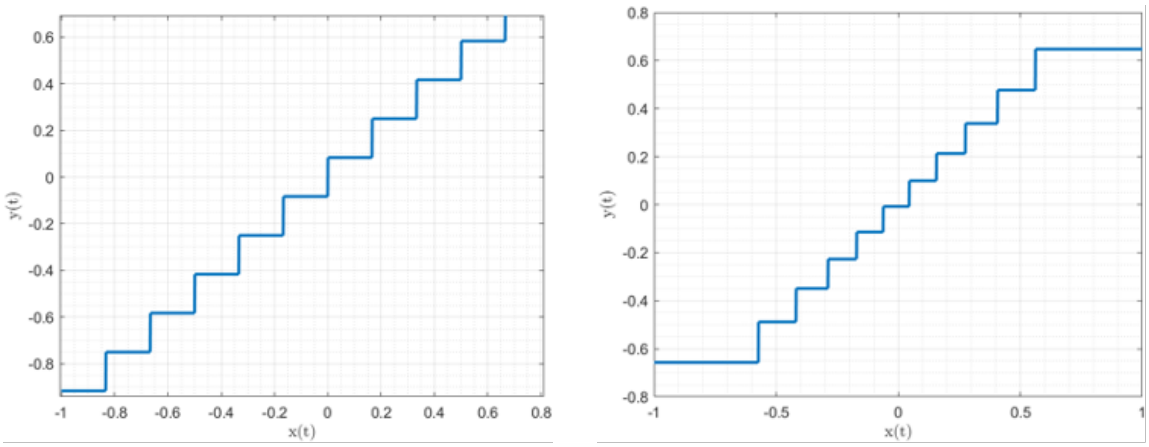
1.2.2. Cuantificadores Según sus Regiones

Cuando un cuantificador hace divisiones de la recta real[1] de igual longitud, el **cuantificador es uniforme**. Por el contrario, si las divisiones varían, siendo más pequeñas en las zonas donde se quiera representar con mayor fidelidad la señal, y más grandes donde se considere que se puede aceptar más distorsión, corresponde a un **cuantificador no uniforme**.

La cuantificación no uniforme presenta ventajas en la codificación de voz, esencialmente por dos razones: uno, un cuantificador no uniforme se adapta de mejor manera a la función de densidad de probabilidad de la señal a cuantificar, por lo que produce una menor distorsión que la que produce un cuantificador uniforme; y dos, con un cuantificador no uniforme se pueden procesar con mayor precisión los valores pequeños de amplitud, lo que, en el caso de las señales de voz, contribuye a una mejor inteligibilidad [10].

Es importante tener en cuenta que la característica de transferencia de un cuantificador describe su comportamiento, evidenciando la relación entre los valores de amplitud de la señal a la entrada y los valores de amplitud de la señal a la salida del cuantificador. Para un cuantificador uniforme, su característica de transferencia se distingue por tener escalones homogéneos, en cuanto a su tamaño, como se observa en la Figura 1.5a; mientras que, el cuantificador no uniforme evidencia una característica de transferencia con escalones de distancias dispares, como se observa en la Figura 1.5b, debido a que para los valores en los que se requiere mayor precisión en la señal se disponen escalones más pequeños, permitiendo aumentar la exactitud, en comparación con el caso uniforme.

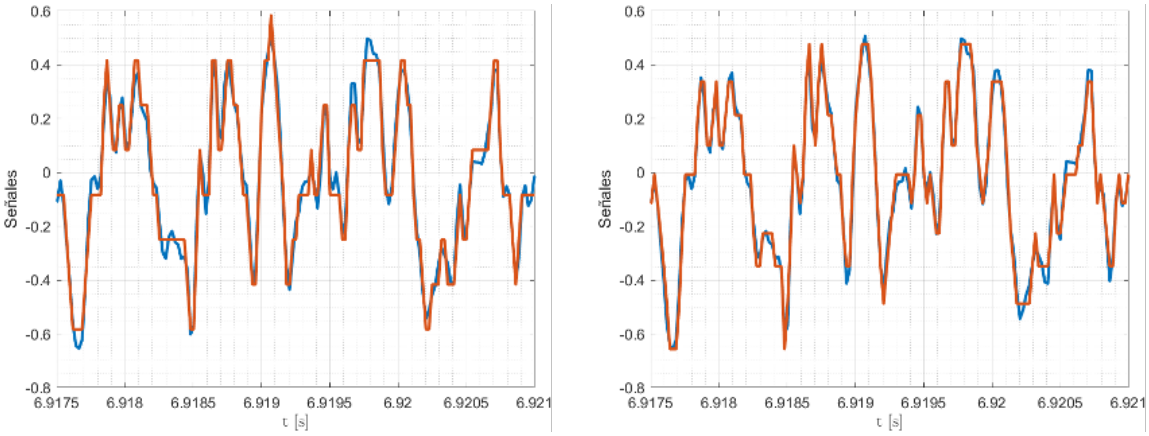
[1] Son todos los posibles valores que puede tomar la amplitud de la señal de entrada al cuantificador y que, posteriormente, son mapeados con el alfabeto de cuantificación.



(a) (b)

Figura 1.5: Característica de transferencia de un cuantificador (a) uniforme (b) no uniforme.

En las características de transferencia mostradas anteriormente se evidencia que el cuantificador no uniforme presenta una mayor precisión en las regiones de valores cercanos a cero, debido a que, para el caso de señal de voz, son las regiones en donde ocurren con mayor probabilidad los valores de amplitud de la señal, es decir, es donde están contenidas la mayor parte de las muestras de la señal, y por ello permite obtener una mayor similitud entre las señales de entrada y salida del cuantificador.



(a) (b)

Figura 1.6: Señal resultante de un cuantificador (a) uniforme (b) no uniforme.

Comparando gráficamente los dos cuantificadores (uniforme y no uniforme), en la Figura 1.6b, se evidencia que la señal de color naranja, obtenida con el cuantificador no uniforme mantiene la forma más cercana a la señal original correspondiente a la de color azul, a diferencia del cuantificador uniforme en la Figura 1.6a, cuya señal resultante de color naranja no se ciñe tan bien a la señal original de color azul.

1.2.3. Cuantificadores Según su Representación

Los cuantificadores se pueden clasificar por su forma de representar la información de la señal que entra al cuantificador. Ésta se puede dar de manera **directa** realizando la cuantificación sobre la señal original de tiempo discreto; o **indirecta**, pasando previamente la señal a un dominio transformado.

1.2.4. Cuantificadores en un Dominio Transformado

Los cuantificadores en un dominio transformado son aquellos que agregan una etapa de transformación previa al proceso de cuantificación, buscando reducir el ruido de cuantificación. Idealmente, dichas transformadas deben ser invertibles, evitando que este paso adicional aumente la distorsión sobre la señal resultante [11]. Existen múltiples alternativas de transformación de dominio, algunos de los dominios más estudiados en el procesamiento de las señales de voz, aparte del dominio temporal, son: dominio cepstral, dominio no lineal y dominio espectral o de la frecuencia [12], en el cual más específicamente se destacan los cuantificadores basados en la Transformada Discreta de Fourier (DFT, *Discrete Fourier Transform*) o la Transformada Discreta de Coseno (DCT, *Discrete Cosine Transform*), donde la cuantificación se realiza sobre los coeficientes obtenidos. Sin embargo, el uso de un cuantificador basado en la DFT equivale a un aumento en el número de bits que representan cada muestra, ya que por cada muestra de amplitud original se debe cuantificar una parte real y una parte imaginaria. Por otro lado, los cuantificadores basados en la DCT buscan evitar el problema de la duplicación de bits descrito anteriormente, por lo que representan la señal original únicamente con funciones coseno, dejando de lado la parte impar de la señal, asociada a las funciones seno, lo cual implica que en el proceso de reconstrucción no se tenga en cuenta toda la información original y se genere una distorsión adicional.

La cuantificación de señales de voz tiene como desafíos que los valores de amplitud de estas señales no tienen una distribución uniforme y son señales no estacionarias, lo que implica una modificación de las características estadísticas de su función de densidad de probabilidad a través del tiempo y un comportamiento poco predecible, lo cual hace poco práctico usar técnicas de optimización

de niveles de cuantificación basadas en su función de densidad de probabilidad. Es por estas razones que se considera pertinente someter las señales de voz a una transformación de dominio antes de su cuantificación, permitiendo diseñar un cuantificador que se adapte a la señal transformada y de esta manera reducir la distorsión para un número de niveles de cuantificación dado.

No obstante, adicionar una transformación al proceso de cuantificación implica algún tipo de contrapartida, por lo que constantemente se estudian nuevos dominios transformados, con la finalidad de encontrar transformaciones cuya implementación sea más práctica, como lo son los dominios tiempo-escala, dentro de los que se destaca la Transformada Discreta *Wavelet* (DWT, *Discrete Wavelet Transform*).

Entre las ventajas que ofrece la DWT se tiene que es completamente invertible y que sus coeficientes son números reales, por lo que no duplica el número de muestras de la señal. Por otro lado, es ampliamente utilizada en aplicaciones de compresión, dado que, generalmente, unos pocos coeficientes contienen la mayor cantidad de la rasgos de la señal, contribuyendo así a disminuir la fidelidad al momento de realizar dicho proceso.

Existen diferentes algoritmos para implementar computacionalmente la DWT. El algoritmo de Mallat ha sido utilizado para definir un método de cuantificación de señales de voz basado en la DWT, el cual toma ventaja de la información proporcionada por los coeficientes *Wavelet* y *Scaling* para minimizar la distorsión de la señal reconstruida respecto a la original. No obstante, el algoritmo de Mallat consiste en filtrajes iterativos, lo cual puede resultar inconveniente dependiendo de la longitud de la señal, es por esto que se deben analizar otras alternativas para implementar la DWT [13].

1.3. Medidas de Distorsión

Las medidas de distorsión son estándares de medición que tienen como objetivo medir la fidelidad de una señal procesada, con respecto a una señal original. Para esto, se comparan la señal original y la señal procesada, y se provee una puntuación cuantitativa que describe el grado de similaridad/fidelidad, o de manera contraria, el nivel de error/distorsión entre ellas. Una de las medidas de distorsión más utilizadas es el Error Medio Cuadrático (MSE, *Mean Square Error*).

Las medidas de distorsión se pueden clasificar en dos categorías generales: medidas subjetivas y medidas objetivas. A continuación, se explican las más importantes en el contexto de las señales de voz.

1.3.1. Medidas Subjetivas

Las medidas subjetivas implican comparaciones entre señales de voz originales y procesadas realizadas por un grupo de personas a las cuales se les pide calificar la calidad de la señal en una escala predeterminada [14].

Las medidas subjetivas se categorizan entre los Métodos de Preferencia Relativa (RPM, *Relative Preference Methods*) y los Métodos de Calificación de Categoría Absoluta (ACRM, *Absolute Category Rating Methods*).

Métodos de Preferencia Relativa - RPM

Estos métodos son los más simples, pues fuerzan a escoger de forma comparativa entre un par de opciones. Normalmente se le presenta a los evaluadores un par de señales producidas por un sistema A y un sistema B, respectivamente. Luego de escuchar y comparar las señales, se les pide que indiquen cuál de las dos señales prefieren. Los resultados se dan en términos del porcentaje de veces que el sistema A fue preferido sobre el sistema B. El principal problema encontrado en estos métodos es que no es fácil comparar el rendimiento de un sistema A con sistemas B obtenidos en otros laboratorios y con otras condiciones.

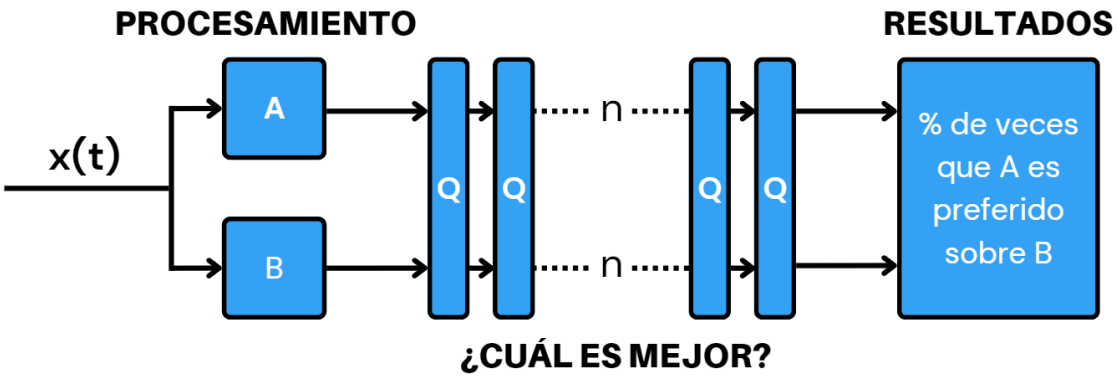


Figura 1.7: Diagrama de Bloques RPM.

También existen variantes, en las que no solamente se pregunta *¿Cuál es mejor?*, sino también *¿Qué tanto es mejor?*. Entre las medidas que siguen esta metodología, resalta una, llamada Calificación de Comparación Categórica (CCR, *Comparison Category Rating*), la cual fue diseñada para cuantificar la intensidad de la diferencia de preferencias. En la Tabla 1.1 se muestran las calificaciones categóricas propuestas por esta medida [15], [16]. Esta escala también es ampliamente

conocida como la Puntuación Media de Opinión de Comparación (CMOS, *Comparison Mean Opinion Score*).

Tabla 1.1: Calificaciones de comparación categórica usados en CMOS.

Calificación	**La calidad del segundo estimulo comparado con el primero es:**
6	Mucho mejor
5	Mejor
4	Poco mejor
3	Casi igual
2	Poco peor
1	Peor
0	Mucho peor

Métodos de Calificación de Categoría Absoluta - ACRM

Estos métodos intentan dar solución a los problemas inherentes a los RPM, mediante el uso de medidas de opinión de calidad absoluta, en los cuales la opinión de la calidad general es solicitada a los evaluadores sin necesidad de comparaciones con referencias. Las medidas más significativas que siguen esta metodología son expuestas a continuación.

Puntuación Media de Opinión - MOS

La Puntuación Media de Opinión (MOS, *Mean Opinion Score*) solicita a los evaluadores que califiquen la calidad de la señal usando una escala numérica de 5 puntos. En la Tabla 1.2 se muestran la escala de puntajes y su correspondiente significado categórico. Este método es recomendado por el Subcomité IEEE de Métodos Subjetivos [17] y también en los estándares ITU-R BS.562-3 [15], ITU-T P.830 [16], ITU-T P.10 [18] e ITU-T P.800 [19].

Tabla 1.2: Escala de calificación MOS.

Calificación	Calidad de la señal	Nivel de distorsión
5	Excelente	Imperceptible
4	Buena	Apenas perceptible, pero no molesta
3	Justa	Perceptible y un poco molesta
2	Pobre	Molesta, pero no detestable
1	Mala	Muy molesta y detestable

Medida Diagnóstica de Aceptabilidad - DAM

MOS se basa en la calificación de la calidad general de la señal; sin embargo, no provee información alguna sobre los criterios de juicio de calidad de los evaluadores. Por ejemplo, dos evaluadores pueden basar su calificación en diferentes características de la señal y aun así dar el mismo puntaje de calidad general. Es por esta razón que MOS es considerado una medida unidimensional y se presenta como alternativa multidimensional la Medida Diagnóstica de Aceptabilidad (DAM, *Diagnostic Acceptability Measure*) [20].

DAM evalúa la calidad de la señal en tres diferentes escalas clasificadas como paramétrica, metamétrica e isométrica. Estas tres escalas cubren un total de 16 diferentes medidas de calidad basadas en diferentes atributos de la señal y de su fondo. Las escalas metamétrica e isométrica basan su evaluación en cuanto a la inteligibilidad, el placer y la aceptabilidad de la señal, mientras que la escala paramétrica se basa en las distorsiones presentes en la señal, en los silencios y en su fondo.

Ya que DAM considera un mayor número de parámetros y analiza la calidad de los periodos de silencio, permite esperar resultados más exactos; sin embargo, en comparación con MOS, DAM requiere de una mayor cantidad de tiempo y de evaluadores cuidadosamente seleccionados tras una serie de evaluaciones, capacitaciones y calibraciones.

Metodología de Prueba Subjetiva para la Evaluación de Sistemas de Comunicaciones de Voz que Incluyen Algoritmos de Supresión de Ruido - STMESCSINSA

Las medidas expuestas hasta el momento fueron diseñadas principalmente para la calificación de codificadores de voz; sin embargo, los codificadores de voz son calificados, por lo general, en espacios silenciosos y por lo tanto presentan diferentes tipos de distorsión que aquellos que tienen adicionalmente algoritmos de supresión de ruido. En este último escenario (particularmente en condiciones con

una alta presencia de ruido), mientras el ruido de fondo es suprimido, también puede ser degradada la señal de voz. Esta situación complica la calificación de calidad de los evaluadores, pues ya no es claro si ésta se basa en el componente de distorsión de la señal, en el componente de distorsión del ruido o en ambos.

Es por esto que se crea la Metodología de Prueba Subjetiva para la Evaluación de Sistemas de Comunicaciones de Voz que Incluyen Algoritmos de Supresión de Ruido (STMESCSINSA, *Subjective Test Methodology for Evaluating Speech Communication Systems that Include Noise Suppression Algorithm*), que incluye el ruido de fondo para la calificación de calidad de la señal de voz y es presentada en el estándar ITU-T P.835 [21].

Esta metodología requiere que el evaluador califique la señal de voz individualmente, el ruido de fondo individualmente y el efecto en general de la voz y el ruido en cuanto a su calidad. Para estas calificaciones se usan respectivamente la Escala de Distorsión de la Señal (SSD, *Scale of Signal Distortion*) presentada en la Tabla 1.3, la Escala de Intrusividad del Fondo (SBI, *Scale of Background Intrusiveness*) presentada en la Tabla 1.4 y la escala de MOS presentada anteriormente en la Tabla 1.2.

Tabla 1.3: Descripción SSD.

Calificación	**Descripción**
5	Muy natural, sin degradación
4	Altamente natural, poca degradación
3	Algo natural, algo degradada
2	Altamente innatural, altamente degradada
1	Muy innatural, muy degradada

Tabla 1.4: Descripción SBI.

Calificación	**Descripción**
5	Imperceptible
4	Algo perceptible
3	Perceptible pero no intrusivo
2	Altamente evidente, algo intrusivo
1	Muy evidente, muy intrusivo

1.3.2. Medidas Objetivas

Las medidas objetivas implican una comparación matemática entre la señal de voz original y la procesada, las cuales cuantifican la calidad por medio de la medida de la "distancia" entre la señal original y la procesada [14].

Las medidas objetivas se pueden clasificar entre las que se basan en comparación directa y las que se basan en parámetros, estos parámetros pueden ser estadísticos o constantes que buscan imitar la percepción humana de los sonidos [13].

Medidas de Comparación Directa

La comparación directa busca estimar la diferencia total entre las muestras de la señal original y las muestras de la señal procesada, sin tener en cuenta la ponderación de diferentes características.

Error Medio Cuadrático - MSE

Una de las medidas de distorsión más famosas es el Error Cuadrático Medio (MSE, *Mean Square Error*). Si se asume que $\mathbf{x} = \{x_i | i = 1, 2, \ldots, N\}$ y que $\mathbf{y} = \{y_i | i = 1, 2, \ldots, N\}$ son dos señales discretas de longitud finita, donde N es el número de muestras de las señales **x** y **y**, x_i y y_i son los valores de la i-ésima muestra en **x** y en **y** respectivamente. El MSE de la señal está descrito de la siguiente manera,

$$MSE\left(\mathbf{x}, \mathbf{y}\right) = \frac{1}{N} \sum_{i=1}^{N} \left(x_i - y_i\right)^2, \tag{1.4}$$

donde $e_i = x_i - y_i$ es el error de la señal, dado que es la diferencia entre la señal original y la señal procesada [22], [3], [13].

Relación Señal a Ruido Segmentada - SSNR

La Relación Señal a Ruido Segmentada (SSNR, *Segmental Signal to Noise Ratio*) es quizás una de las medidas objetivas más sencillas que se utilizan para evaluar los algoritmos de mejora o codificación del habla. Para que esta medida sea significativa, es necesario que la señal original y la señal procesada estén alineadas en el tiempo y que se corrija cualquier error de fase presente. La SSNR se define de la siguiente manera,

$$SSNR\left(\mathbf{x}, \mathbf{y}\right) = \frac{10}{M} \sum_{j=0}^{M-1} log\left(\frac{\sum_{i=N_j}^{N_j+N-1} {x_i}^2}{\sum_{i=N_j}^{N_j+N-1} \left(x_i - y_i\right)^2}\right), \tag{1.5}$$

donde x_i y y_i son las muestras i-ésimas de la señal original y de la señal procesada respectivamente, N es la longitud de la trama[2] (típicamente 15 o 20 ms de duración), N_j es un índice de tiempo que indica el momento inicial del bloque de muestras a procesar y M es el número de tramas en las señal.

Sin embargo, la SSNR presenta un problema potencial y es que la energía de la señal del habla durante los intervalos de silencio (que son abundantes en el habla conversacional) es muy pequeña, lo que da lugar a grandes valores negativos de SSNR, que sesgan la medida global. Además, en [14] se demuestra que no presenta una buena correlación con las medidas subjetivas y propone no usarla en el contexto de señales de voz.

Error Medio Cuadrático Normalizado - NMSE

Una de las principales desventajas del MSE es que es susceptible a los cambios de escala y desfases, pues al evaluar la diferencia de las i-ésimas muestras de la señal original y de la señal procesada, no se está teniendo en cuenta si ambas señales están sincronizadas, así como tampoco se está teniendo en cuenta su escala; por lo que la misma señal con un leve desplazamiento (representado como un desfase) obtendría valores altos de MSE, indicando una falsa diferencia de la señal, misma consecuencia que se da en el caso donde se compara la misma señal con diferentes escalas, pues la diferencia de tamaño de las señales también resulta en valores altos de MSE; sin embargo, esto no implica una diferencia en cuanto a información contenida en las señales. Teniendo esto en mente, se presenta el Error Cuadrático Medio Normalizado (NMSE, *Normalized Mean Square Error*), el cual es una variación del MSE, donde los problemas antes mencionados son suprimidos mediante la sincronización[3] y normalización de las señales a evaluar [23]. El NMSE se define por la siguiente ecuación,

$$NMSE\left(\mathbf{x},\mathbf{y}\right)=1-\frac{1}{2}\sqrt{\frac{\sum_{i=1}^{N}\left(x_i-y_i\right)^2}{\sum_{i=1}^{N}\left(x_i\right)^2}}. \tag{1.6}$$

[2]Hace referencia a las particiones de menor duración en las que se divide la señal a cuantificar.

[3]La sincronización y normalización de las señales son condiciones necesarias que el evaluador debe realizar de forma manual previa a la utilización de la NMSE.

Medidas Basadas en Parámetros Estadísticos

Este tipo de medidas no realiza una comparación directa entre las señales, pues lo que compara son los cambios sufridos sobre la señal original a partir de la señal procesada, para esto realiza una ponderación de los parámetros estadísticos de ambas señales.

Media de Similitud Estructural - MSSIM

La Media de Similitud Estructural (MSSIM, *Mean Structural SIMilarity*) se basa en la idea de que la medida del cambio en las características estructurales es una buena aproximación al cambio de calidad percibida. La MSSIM se calcula realizando la ponderación de tres medidas diferentes: luminosidad (ℓ), contraste (c) y estructura ($\int$) [23], [24].

$$\ell\left(\mathbf{x},\mathbf{y}\right)=\frac{2\mu_x\mu_y+C_1}{{\mu_x}^2+{\mu_y}^2+C_1}, \tag{1.7}$$

$$c\left(\mathbf{x},\mathbf{y}\right)=\frac{2\sigma_x\sigma_y+C_2}{{\sigma_x}^2+{\sigma_y}^2+C_2}, \tag{1.8}$$

$$\int\left(\mathbf{x},\mathbf{y}\right)=\frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3}, \tag{1.9}$$

donde μ_x y μ_y son las medias de la señal original y la señal procesada respectivamente; σ_x y σ_y sus desviaciones estándar; σ_{xy} su covarianza; y C_1, C_2 y C_3 son constantes introducidas para evitar inestabilidad.

Finalmente, la MSSIM 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.

$$MSSIM\left(\mathbf{x},\mathbf{y}\right)=\ell\left(\mathbf{x},\mathbf{y}\right)^{\alpha}c\left(\mathbf{x},\mathbf{y}\right)^{\beta}\int\left(\mathbf{x},\mathbf{y}\right)^{\gamma}. \tag{1.10}$$

Motivadas por la Percepción - PMM

Las medidas expuestas anteriormente son atractivas porque son fáciles de implementar; sin embargo, su habilidad para determinar si la señal realmente tiene alta calidad para el usuario final es limitada, ya que no emulan el procesamiento de la señal presente en la percepción humana de la voz. Las Medidas Motivadas por la Percepción (PMM, *Perceptually-Motivated Measures*) intentan solventar esta limitante basándose en modelos de la percepción humana de la voz [25], [26]. Las medidas más significativas son expuestas a continuación.

Medida de Distorsión Bark - BDM

La Medida de Distorsión Bark (BDM, *Bark Distortion Measures*) tiene en cuenta dos aspectos: que la resolución de frecuencia del oído no es uniforme y, por lo tanto, el análisis frecuencial de las señales acústicas no se basa en una escala de frecuencias lineal; y que la sonoridad está relacionada con la intensidad de la señal de forma no lineal. La BDM para la trama k-ésima se basa en la diferencia entre el espectro de sonido, y se define como,

$$BDM\left(M\right) = \sum_{i=1}^{N_b} \left[\tilde{S}_M\left(i\right) - \overline{\tilde{S}_M}\left(i\right)\right]^2, \tag{1.11}$$

donde $\tilde{S}_M\left(b\right)$ y $\overline{\tilde{S}_M}\left(b\right)$ son los espectros sonoros de la señal original y la señal procesada respectivamente y N_b es el número de bandas críticas.

Un problema presentado en esta medida es que en los momentos de silencio se obtienen valores muy altos, pero esto se puede resolver usando detectores de voz/silencio. Adicionalmente, se ha encontrado que la BDM tiene una alta correlación ($\rho = 0{,}9$) con la medida subjetiva MOS [14].

Evaluación Perceptiva de la Calidad de Voz - PESQ

La mayoría de las medidas objetivas expuestas funcionan de forma correcta para un limitado número de distorsiones; no obstante, no se consideran las más comunes cuando una señal de voz es transmitida por un canal de telecomunicaciones, como lo son: la pérdida de paquetes, los retrasos de la señal y las distorsiones.

Es por esto que la Medida de Evaluación Perceptiva de la Calidad de Voz (PESQ, *Perceptual Evaluation of Speech Quality*) [27], es seleccionada por la recomendación ITU-T P.862.

Esta medida propone que, en primera medida, la señal original y la señal procesada sean ecualizadas y filtradas a un nivel de escucha estándar. Luego, las señales son sincronizadas en el tiempo para corregir errores de desfases y se obtiene su espectrograma, similar a como lo hace BDM. Finalmente, la diferencia absoluta entre los espectrogramas de la señal original y la señal procesada es usada como medida del error audible en la etapa final, donde, a diferencia de BDM, se discrimina entre las diferencias positivas y negativas, puesto que PESQ reconoce una diferencia positiva como la adición de una componente de ruido y una diferencia negativa como la omisión o dura atenuación de un componente. Debido a que los componentes omitidos no son tan fácilmente percibidos y pueden llevar a una forma de distorsión poco objetiva, se aplican distintos pesos a las diferencias positivas y negativas, permitiendo obtener una precisa predicción de medidas subjetivas como la MOS.

En el diagrama mostrado en la Figura 1.8 se sintetiza la estructura en bloques necesaria para la obtención de la PESQ, siguiendo el proceso anteriormente descrito. Es importante resaltar que se incluyen etapas de preprocesamiento (filtraje y ecualización) y alineación (sincronización en el tiempo) para garantizar una comparación justa entre las señales de audio.

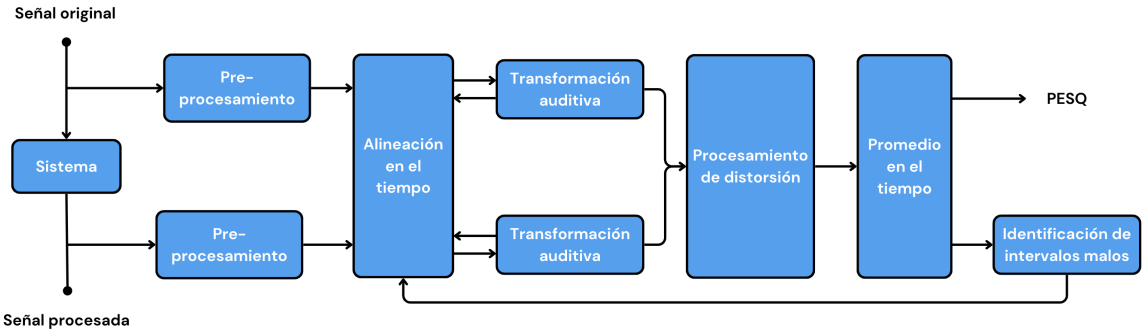


Figura 1.8: Diagrama de Bloques PESQ.

Al realizar comparaciones en diversas condiciones se ha determinado que la PESQ tiene una alta correlación ($\rho = 0{,}92$) con las medidas subjetivas y que, por ende, es la medida objetiva más confiable para la evaluación de la calidad de señales de voz [14].

CAPÍTULO 2

TRANSFORMADA WAVELET

Para el análisis y procesamiento de señales es útil tener un cambio de perspectiva, por lo que se suelen utilizar transformaciones. La FT fue durante muchos años la más usada, debido a que es la forma más directa de obtener información en el dominio de la frecuencia. Cuando se abarcan contextos que implican señales no estacionarias, como lo es el caso de las señales se voz, la representación de la información con la FT tiene limitaciones, debido a que, esta transformada puede brindar perfectamente información de lo que pasa en el tiempo o en la frecuencia, pero no puede relacionar los dos dominios, como lo denota Heisenberg en su principio de incertidumbre.

Como solución a la incertidumbre en uno de los dominios, y para analizar cómo cambia el contenido de frecuencia de una señal no estacionaria en diferentes instantes de tiempo [28], surge la Transformada de Fourier en Tiempo Corto (STFT, *Short-time Fourier Transform*), la cual implementa el análisis en ventanas específicas de tiempo con el propósito de proporcionar la información de frecuencia localizada [29], es decir, la FT se hace por segmentos de la señal en lugar de sobre toda la señal, permitiendo observar (según el tamaño de la ventana) las componentes espectrales que tienen lugar en dicha ventana de tiempo, posteriormente se mueve a otro segmento de tiempo y así sucesivamente. Esto resulta en una secuencia de FT de múltiples señales enventanadas, obteniendo así la información de ambos dominios.

Sin embargo, la STFT está supeditada a la duración y tipo de ventana. Una ventana de gran duración permite tener una buena resolución en la frecuencia, sacrificando la resolución en el tiempo, mientras que una ventana de corta duración permite tener una buena resolución en el tiempo, sacrificando la resolución en la frecuencia. Por otro lado, dado que el proceso de enventanado implica una multiplicación en el tiempo, en el dominio de la frecuencia se tiene una modificación por la convolución del espectro de la señal con el espectro del tipo de ventana utilizada.

La STFT no se adapta correctamente a todas las aplicaciones del procesamiento de señales, puesto que para una ventana dada se mantiene una resolución fija, sin importar si la señal tiene cambios abruptos o segmentos que requieren mayor resolución, por lo que si la ventana no se adapta correctamente a ellos, se puede llegar a conclusiones imprecisas. Es por ello que aparece la Transformada *Wavelet* (WT, *Wavelet Transform*), la cual permite trabajar con diferentes resoluciones y compensar la incertidumbre existente en alguno de los dominios.

2.1. Transformada Wavelet Continua

La Transformada *Wavelet* Continua (CWT, *Continuous Wavelet Transform*) permite realizar un análisis de tiempo-frecuencia y el filtrado de componentes de frecuencia localizados en el tiempo, por medio de una ventana con una anchura de tiempo más corta para las frecuencias altas y más ancha para frecuencias más bajas [30]. Esta ventana ajustable se realiza modificando la escala de una función llamada *Wavelet madre*, $\psi(t)$, la cual es una onda de duración finita, aproximadamente acotada en frecuencia. Por lo anterior, la CWT constituye una herramienta de análisis para fenómenos transitorios, no estacionarios o variables en el tiempo [31].

La CWT es una transformada biparamétrica dado que crea un conjunto de funciones de base a partir de la combinación de dos parámetros, escala y traslación, que modifican a $\psi(t)$. La propiedad de escalamiento hace referencia a la posibilidad de estirar o encoger la *Wavelet madre* en el tiempo, siendo a el factor de escalamiento, la expresión de *Wavelet* con escalamiento está dada de la siguiente manera [32],

$$\psi\left(\frac{t}{a}\right),\ a > 0. \tag{2.1}$$

Cuanto mayor sea el factor de escalamiento, a, más estirada estará la *Wavelet*, lo que permite analizar frecuencias más bajas y detectar cambios más lentos de la señal. Por otro lado, con un a de valores menores, se obtiene una *Wavelet* más comprimida, lo que permite el análisis de frecuencias altas y a su vez se ajusta a los cambios abruptos que pueda tener la señal no estacionaria. En la Figura 2.1 se muestra una *Wavelet de Morlet*, con dos factores de escalamiento diferentes; la *Wavelet* de color azul posee un a mayor al de la *Wavelet* de color rojo, es decir, la *Wavelet* de color azul se ajusta mejor a las frecuencias bajas.

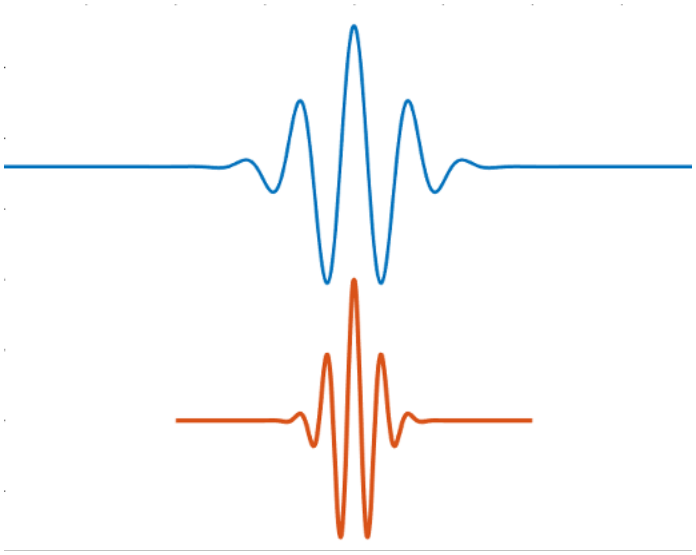


Figura 2.1: *Wavelet de Morlet* a diferentes escalas.

En cuanto a la propiedad de traslación, es la que permite retrasar o adelantar la *Wavelet* a lo largo de la señal que se quiere analizar. La traslación en b segundos de la *Wavelet* madre se define matemáticamente de la siguiente manera,

$$\psi\left(t-b\right). \tag{2.2}$$

Explicando la expresión anterior, la *Wavelet* es desplazada y centrada en b. La variación de este parámetro permite analizar la señal original por tramos, como se puede observar en la Figura 2.2, en donde la *Wavelet* (onda de color rojo) se va deslizando por la señal de color azul para poder realizar el análisis.

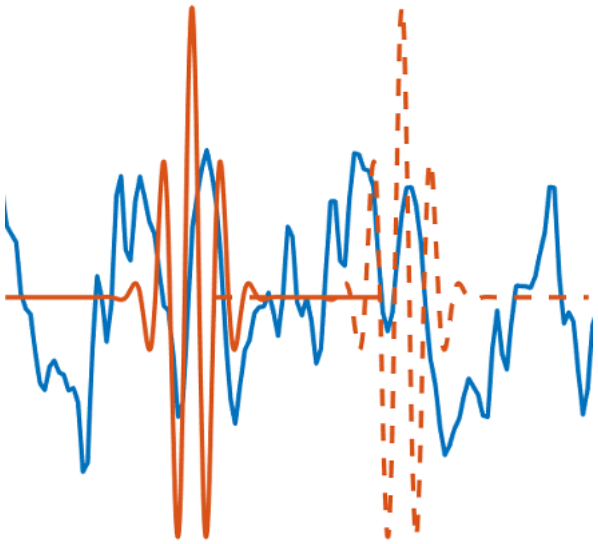


Figura 2.2: Traslación de la *Wavelet*.

Aplicando simultáneamente escalamiento y traslación, la *Wavelet* se describe matemáticamente de la siguiente manera,

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right). \tag{2.3}$$

Se define entonces la CWT como una superficie, la cual es obtenida matemáticamente como [31],

$$W_x(a,b) = \int_{-\infty}^{\infty} x(t)\ \ \psi_{a,b}(t)\, dt. \tag{2.4}$$

Esta transformada es invertible, por lo que la señal original se obtiene de la siguiente manera,

$$x(t) = C\int_0^{\infty}\int_{-\infty}^{\infty} W_x(a,b)\ \ \psi_{a,b}(t)\ db\ da, \tag{2.5}$$

donde C es un valor constante que está asociado a la familia *Wavelet* utilizada, pero que no dispone de mayor relevancia.

No obstante, las *Wavelet* madre son construidas sabiendo *a priori* que deben cumplir con tres condiciones de admisibilidad [33], [34]:

1. El área bajo la curva debe ser igual a cero.

$$\int_{-\infty}^{\infty} \psi\ (t)\ dt = 0. \tag{2.6}$$

Esta condición conlleva a que $\widetilde{\psi}\ (0) = 0$ y, por lo tanto, se puede considerar a la *Wavelet* como una función en pasa-banda.

2. La duración debe ser finita.

$$\psi\,(t) = 0,\ \forall t \notin [t_1, t_2] \tag{2.7}$$

3. Buena localización espectral.

$$E_\psi \approx \int_{f_1}^{f_2} \left|\widetilde{\psi}\ (f)\right|^2\ df,\ -\infty < f_1 < f_2 < \infty. \tag{2.8}$$

Las condiciones 2 y 3 muestran que la *Wavelet* es una función finita en el tiempo, que en el dominio de la frecuencia tiene la mayoría de su energía concentrada en un rango finito de componentes.

Teóricamente, la CWT presenta grandes beneficios en el análisis de señales no estacionarias, pero su carácter de continuo genera que, en la práctica, no sea computacionalmente realizable, por lo que la DWT surge como una forma de evitar la redundancia presente en la CWT, buscando así llegar a un algoritmo implementable en la práctica.

2.2. Transformada Wavelet Discreta

La DWT, además de ser discreta y computacionalmente realizable, se encarga de remover la redundancia presente en la CWT, mediante la determinación de los valores de traslación y de escala estrictamente necesarios para conservar la totalidad de la información de la señal.

La discretización de la CWT se logra mediante la limitación de los parámetros a y b, pues así se tiene una variación discreta a nivel de escala y traslación de la señal, respectivamente. De esta forma, se define la variación de los parámetros a y b en función de una base arbitraria z, en la práctica se trabaja con $z = 2$, pues este valor asegura que no se presente ni traslape ni separación entre dos bandas adyacentes en el dominio de la frecuencia [13].

$$a = z^j = 2^j, \tag{2.9}$$

$$b = z^j k = 2^j k. \tag{2.10}$$

Así, j y k, conocidos como el nivel de resolución y el nivel de traslación respectivamente, pertenecen al conjunto de los números enteros $\mathbb{Z}$. La creación de la familia *Wavelet* para la DWT se obtiene cuando las ecuaciones 2.9 y 2.10 son reemplazadas en la definición de la WT, lo cual resulta en,

$$\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{t-2^j k}{2^j}\right) = 2^{-j/2}\psi\left(2^{-j}t-k\right). \tag{2.11}$$

Es de gran importancia resaltar que a pesar de haber restringido los valores de los parámetros a y b a desplazamientos en función de potencias enteras de dos, los niveles de resolución j y traslación k aún toman valores infinitos.

Es entonces como la aplicación de la DWT para la obtención de los coeficientes *Wavelet* ${W_x}^j[k]$ se define como,

$$W_x^j[k] = 2^{-j/2}\int_{-\infty}^{\infty} x(t)\,\psi\left(2^{-j}t-k\right)dt. \tag{2.12}$$

Y su proceso inverso como,

$$x(t) = \sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} 2^{-j/2} W_x^j[k]\,\psi\left(2^{-j}t-k\right). \tag{2.13}$$

2.3. Análisis Multiresolución

La DWT idealmente busca un cubrimiento total del espectro a partir de la creación de particiones de éste, mediante la utilización de una familia *Wavelet* y la variación necesaria de su parámetro de escala o nivel de resolución, tal y como se representa en la Figura 2.3; sin embargo, como la *Wavelet* es una señal pasa-banda, siempre existirá en el espectro un espacio no cubierto adyacente al origen, pues este espacio solo es cubierto cuando la variación del parámetro de escala tiende a infinito, lo cual es computacionalmente imposible.

Figura 2.3: Partición del espectro mediante la función *Wavelet* en la DWT.

Es entonces como, con el fin de conseguir una implementación no teórica de la DWT, se realiza un Análisis MultiResolución (MRA, *MultiResolution Analysis*) el cual permite obtener la información contenida en el espacio no cubierto por la función *Wavelet*. Aquí se introduce el concepto de función *Scaling*, $\varphi(t)$.

La función $\varphi(t)$ es creada con el fin de poder reconstruir la señal original sin pérdida de información, ya que a diferencia de la función *Wavelet*, es una función en banda-base que tiene un área bajo la curva diferente de cero. El hecho de que $\varphi(t)$ sea una función en banda-base implica que las particiones del espectro creadas mediante la variación del nivel de resolución están contenidas una dentro de la siguiente, y no están adyacentes como sucede con las funciones *Wavelet*. Esto es visible en la Figura 2.4.

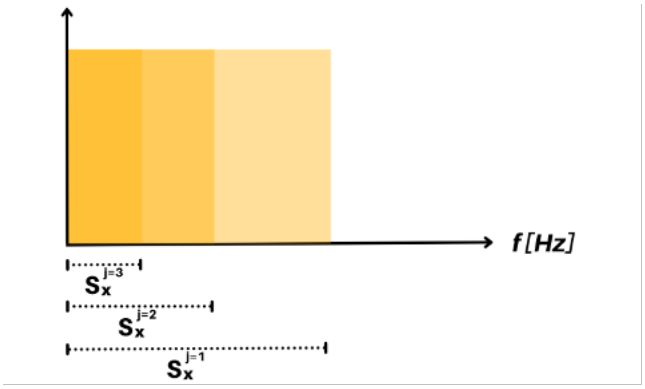


Figura 2.4: Partición del espectro mediante la función *Scaling* en la DWT.

Así, los coeficientes *Scaling*, $S_x^j[k]$, en el nivel de resolución j, se obtienen de la proyección de la señal sobre la partición del mismo nivel de resolución, de la siguiente manera,

$$S_x^j[k] = \int_{-\infty}^{\infty} x(t)\ \varphi_{j,k}(t)\, dt. \tag{2.14}$$

El proceso inverso, para recuperar la proyección de la señal original sobre la partición del nivel de resolución j, se define así,

$$x^j(t) = \sum_{k=-\infty}^{\infty} S_x{}^j[k]\varphi_{j,k}(t). \tag{2.15}$$

De esta forma, se constituye la aplicación del MRA como la utilización complementaria de las particiones del mismo nivel de resolución generadas por las funciones *Wavelet* y *Scaling*, tal y como se muestra en la Figura 2.5.

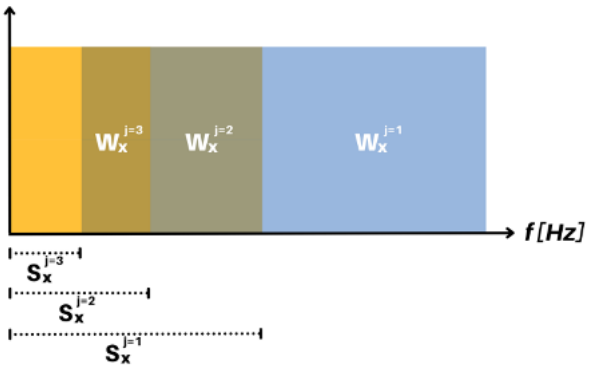


Figura 2.5: Uso complementario de las particiones del espectro generadas por las funciones *Wavelet* y *Scaling* en la DWT[4].

Entonces, es posible representar a la señal original como una combinación de coeficientes *Wavelet* y *Scaling*, como se muestra a continuación,

$$x\left(t\right)=\sum_{k=-\infty}^{\infty}{S_x}^{n}\left[k\right]\varphi_{n,k}\left(t\right)+\sum_{j=1}^{n}\sum_{k=-\infty}^{\infty}{W_x}^{j}\left[k\right]\psi_{j,k}\left(t\right). \tag{2.16}$$

La implementación computacional del MRA se lleva a cabo a través de filtros digitales, diezmadores y/o sobre-muestreadores en un proceso conocido como Transformada Rápida *Wavelet* (FWT, *Fast Wavelet Transform*) o algoritmo de Mallat [35].

2.3.1. Algoritmo de Mallat

El algoritmo de Mallat es un método de análisis de señal en el dominio *Wavelet*, el cual utiliza bancos de filtros de respuesta al impulso finita y permite realizar el MRA principalmente para las familias *Wavelet* ortogonales de primera generación.

Este algoritmo funciona a través de un proceso iterativo de filtrado, sub-muestreo y descomposición, como se muestra en la Figura 2.6. De manera general, este proceso tiene la siguiente secuencia:

1. La señal de entrada, $x\left[n\right]$, es filtrada por un filtro pasa alto, más específicamente un filtro *Wavelet*, $g\left[-n\right]$, con el objetivo de extraer la información de las altas frecuencias.

[4]Las funciones *Wavelet* y *Scaling* de un mismo nivel de resolución j son ortogonales entre sí.

2. La señal $x\,[n]$ es filtrada por un filtro pasa bajo, más específicamente un filtro *Scaling*, $h\,[-n]$, con el objetivo de obtener la información de las bajas frecuencias.

3. Se sub-muestrean en un factor de dos las señales resultantes de los filtros $g\,[-n]$ y $h\,[-n]$.

4. Se repiten los pasos 1, 2 y 3 sobre las señales resultantes de los de los filtros *Scaling* cuantas veces se considere necesario, obteniendo una descomposición satisfactoria de la señal original en sub-bandas con información de distintas bandas de frecuencia.

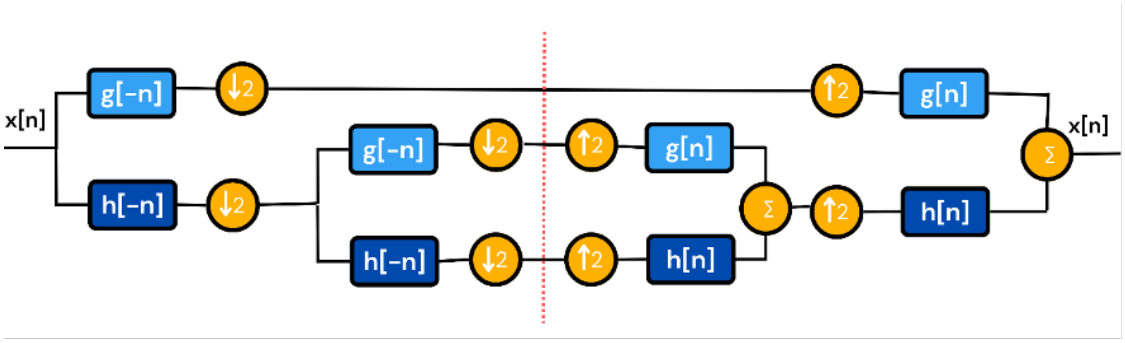


Figura 2.6: Diagrama de bloques del algoritmo de Mallat.

Las divisiones en el espectro generadas por las familias *Wavelet* ortogonales, permiten analizar el espectro de la señal de manera completa, desde $F_s/2$ hasta el origen, sin redundancia. Al aumentar el número de etapas del algoritmo de Mallat se incrementa el número de divisiones del espectro, en la Figura 2.7 se muestran las divisiones ideales para una familia *Wavelet* ortogonal.

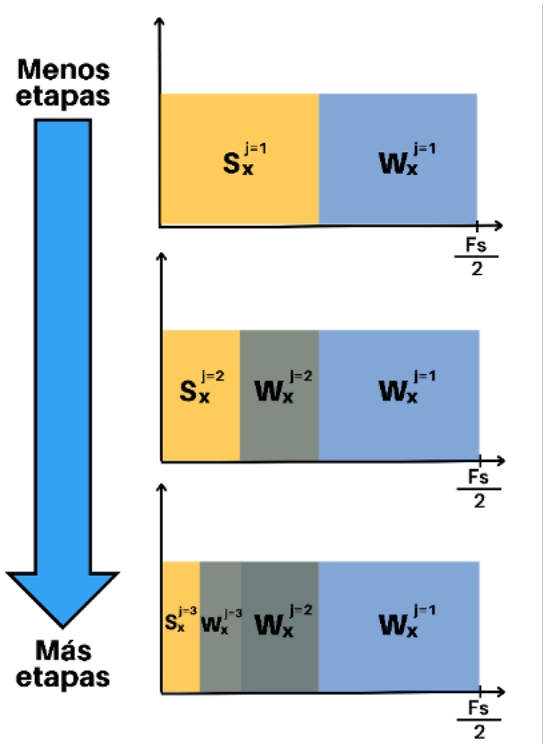


Figura 2.7: Divisiones del espectro de la DWT.

A medida que se van aumentando las etapas, la ventana *Scaling* se va reduciendo cada vez más, permitiendo tener un análisis más riguroso en las frecuencias que se encuentran alrededor del origen, cuyo aporte de energía, por lo general, es el que más datos de la señal contiene, como a medida que se va incrementando el nivel de resolución se va aumentando el tamaño de la ventana, el número de coeficientes de etapas superiores es menor.

2.3.2. Esquema Lifting

El esquema básico de *Lifting*, propuesto por *Sweldens* [36], es un nuevo mecanismo para separar los coeficientes *Scaling* de los coeficientes *Wavelet*, que a diferencia del algoritmo de Mallat, no se basa en convoluciones sucesivas, donde la separación de los coeficientes antes mencionados se logra gracias a la utilización de un banco de filtros de análisis compuesto por un filtro pasa bajo y un filtro pasa alto.

El esquema *Lifting* propone factorizar el banco de filtros de análisis en una secuencia finita de matrices triangulares que se alternan entre las superiores y las inferiores. Es así como se puede dejar a un lado el proceso de separación de

coeficientes por filtraje y, en su lugar, realizar la separación mediante multiplicaciones entre las matrices antes mencionadas, buscando agilizar el proceso y disminuir la posibilidad de que se creen distorsiones.

La computación de la matriz triangular superior es conocida como *primal Lifting*, mientras que la computación de la matriz triangular inferior es conocida como *dual Lifting*. Estos dos procesos también son usualmente conocidos como predicción, **P** y actualización, **U**, los cuales son vectores que varían según la familia *Wavelet* y multiplican a la señal de entrada que corresponda [37], [38], [39].

En la Figura 2.8 se presenta un diagrama del esquema básico de *Lifting*, en el cual se tienen sus tres bloques principales.

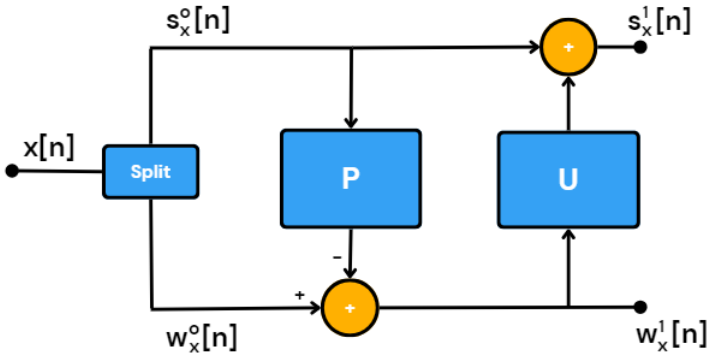


Figura 2.8: Esquema básico de *Lifting*. Adaptado de [1].

En el primer bloque, denominado *Split*, ocurre la separación de las muestras pares e impares de la señal. De esta forma, se define $s_x^0[n]$ como el conjunto de muestras pares de la señal, y $w_x^0[n]$ como el conjunto de muestras impares de la señal, esto es:

$$s_x^0[n] = \text{even}\{x[n]\} = x[2n]. \tag{2.17}$$

$$w_x^0[n] = \text{odd}\{x[n]\} = x[2n+1]. \tag{2.18}$$

El siguiente bloque, denotado por la letra **P**, representa el proceso de predicción. Los coeficientes *Wavelet* resultan de la diferencia entre la muestra actual y la predicción realizada sobre ésta a partir de sus muestras aledañas, así, este bloque se aprovecha de la correlación existente entre muestras vecinas para encontrar los cambios más rápidos de la señal, los cuales están relacionados con las componentes de alta frecuencia. En el caso particular de las señales de voz se tiene

una alta correlación entre muestras consecutivas, por lo que el primer grupo de coeficientes *Wavelet* no tiene un gran porcentaje de energía.

Finalmente, el último bloque, denotado por la letra **U**, representa el proceso de actualización. Dada una muestra par, se ha predicho que la siguiente muestra impar tiene un valor similar y, por lo tanto, se almacena la diferencia de estos valores. Es así entonces como se actualiza la entrada par con el fin de reflejar el conocimiento de la señal y analizar bajas frecuencias.

A continuación, se describe el funcionamiento del esquema de *Lifting* para la familia *Wavelet* de *Haar*, en donde se usa como señal de entrada el vector,

$$x\left[n\right] = \left[7, 2, 9, 5, 3, 1, 0, 4\right], \tag{2.19}$$

y bloques **P** y **U**, con valores de 1 y $\frac{1}{2}$ respectivamente, como se muestra en el Figura 2.9.

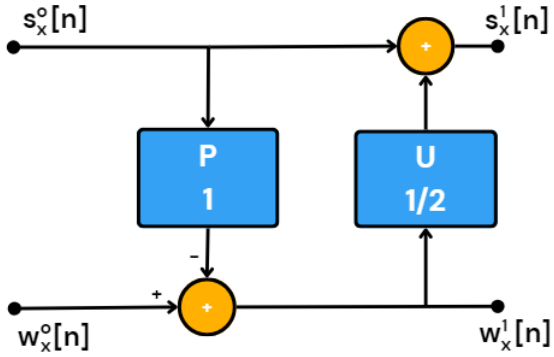


Figura 2.9: Esquema básico de *Lifting* para la familia *Wavelet* de *Haar*

Al ingresar la señal $x[n]$ al bloque *Split*, se obtiene:

$$s_x^0\left[n\right] = \left[7, 9, 3, 0\right], \tag{2.20}$$

$$w_x^0\left[n\right] = \left[2, 5, 1, 4\right]. \tag{2.21}$$

Por tanto, al realizar el proceso de predicción, **P**, se obtiene que el grupo de coeficientes *Wavelet* w_x^1, está dado por,

$$w_x^1\left[n\right] = w_x^0\left[n\right] - \mathbf{P}\left(s_x^0\left[n\right]\right), \tag{2.22}$$

$$w_x^1\left[n\right] = w_x^0\left[n\right] - s_x^0\left[n\right], \tag{2.23}$$

$$w_x^1[n] = \left[\frac{9}{2}, 7, 2, 2\right]. \tag{2.24}$$

Una vez obtenido el valor de w_x^1, se realiza el proceso de actualización, **U** para obtener los valores del grupo de coeficientes *Scaling*, $s_x^1[n]$ definidos como,

$$s_x^1[n] = s_x^0[n] + \mathbf{U}\left(w_x^1[n]\right) \tag{2.25}$$

$$s_x^1[n] = x[2n] + \frac{w_x^1[n]}{2} \tag{2.26}$$

$$s_x^1[n] = [-5, -4, -2, 4] \tag{2.27}$$

El algoritmo descrito representa un esquema *Lifting* de una etapa. Un esquema *Lifting* puede tener tantas etapas como se considere necesario, donde las señales $w_x^1[n]$ son almacenadas y la señal $s_x^1[n]$ es usada como entrada a la siguiente etapa del esquema. Este proceso es representado en la Figura 2.10.

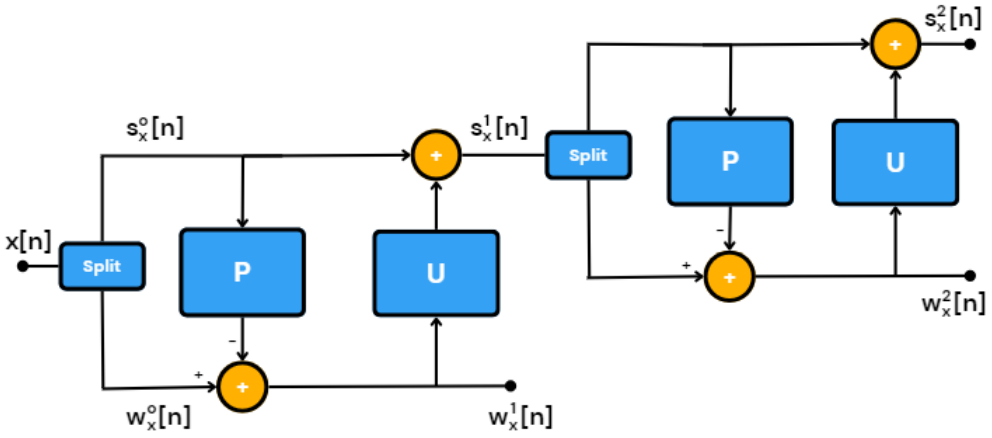


Figura 2.10: Esquema básico de *Lifting* de dos etapas. Adaptado de [1].

Así, el esquema *Lifting* consiste esencialmente en realizar la separación de las muestras pares e impares de la señal a tratar (proceso conocido como *Lazy Wavelet Transform*), y, alternando, aplicar el *primal Lifting* y el *dual Lifting*, produciendo así dos sub-bandas: una pasa baja y otra pasa alta [37]; las cuales corresponden a los coeficientes *Scaling*, $s_x^1[n]$, y *Wavelet*, $w_x^1[n]$, respectivamente [38].

El esquema básico de *Lifting* también es un procedimiento invertible, lo cual se logra reflejando el diagrama de bloques mostrado en la Figura 2.8, esto es, invirtiendo los signos de las operaciones y realizando una mezcla de las dos señales resultantes, tal y como se indica en la Figura 2.11.

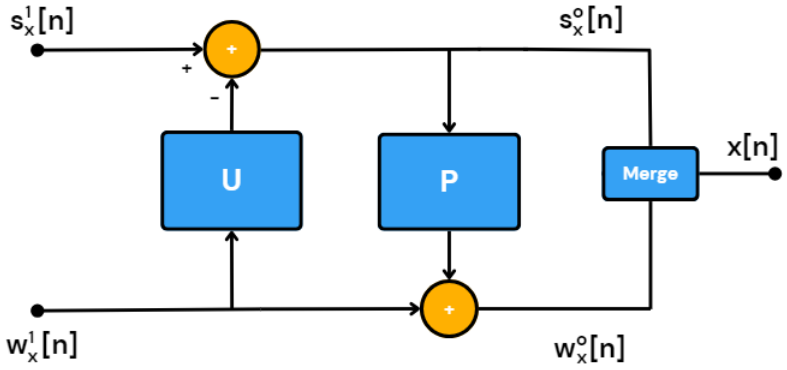


Figura 2.11: Esquema básico inverso de *Lifting*. Adaptado de [1].

De forma general, la transformación directa realizada por el esquema básico de *Lifting* se define como:

$$w_x^1[n] = w_x^0[n] - \mathbf{P}\left(s_x^0[n]\right). \tag{2.28}$$

$$s_x^1[n] = s_x^0[n] + \mathbf{U}\left(w_x^1[n]\right). \tag{2.29}$$

Y su proceso inverso como:

$$w_x^0[n] = w_x^1[n] + \mathbf{P}\left(s_x^0[n]\right). \tag{2.30}$$

$$s_x^0[n] = s_x^1[n] - \mathbf{U}\left(w_x^1[n]\right). \tag{2.31}$$

El esquema *Lifting* presenta ventajas sobre algoritmos basados en convoluciones, dado que requiere hasta un 50 % menos de recursos computacionales [37], y ofrece la posibilidad de implementar la transformada *Wavelet* entera, la cual se ajusta a procesamiento de señales sin pérdidas [38].

CAPÍTULO 3

DISEÑO E IMPLEMENTACIÓN[5]

[5]El código realizado durante el desarrollo de este trabajo de grado se puede encontrar en el siguiente repositorio: https://github.com/jfredyromero/TesisDreamTeam.git

Para diseñar un cuantificador de señales de voz en el dominio *Wavelet* utilizando el esquema de *Lifting*, es fundamental comprender los aspectos clave y consideraciones necesarias para el cuantificador *per se* y para los procesos aplicados a la señal previa y posteriormente al cuantificador, como los son el pre-procesamiento y la codificación de fuente respectivamente.

En la Figura 3.1 se muestra el diagrama de bloques general del cuantificador de señales de voz en el dominio *Wavelet* utilizando el esquema de *Lifting*, donde $x\,[n]$ corresponde a la señal de voz original, $x_i\,[n]$ son las tramas en las que se divide la señal de voz, $c_i\,[\kappa]$ corresponde al arreglo que contiene los diferentes grupos de coeficientes que se obtienen al aplicar la DWT, $c_i'\,[\kappa]$ es la versión cuantificada de los coeficientes en el dominio *Wavelet*, $\{b_i\}$ son secuencias binarias que describen los valores cuantificados de las tramas y $\{b_t\}$ es la secuencia total necesaria para representar la información de la señal de voz, la cual no puede exceder un límite superior impuesto por la restricción del número de niveles de cuantificación.

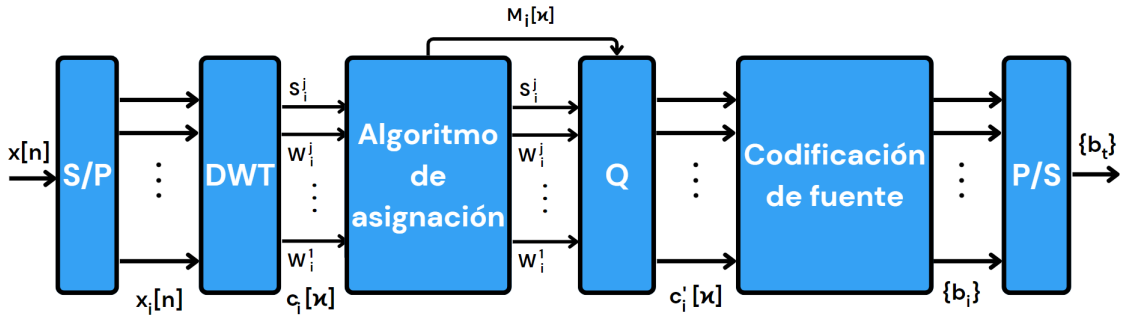


Figura 3.1: Diagrama de bloques general del cuantificador de señales de voz en el dominio *Wavelet* utilizando el esquema de *Lifting*.

En la Figura 3.2 se muestra el diagrama de bloques que sintetiza el funcionamiento del algoritmo inverso.

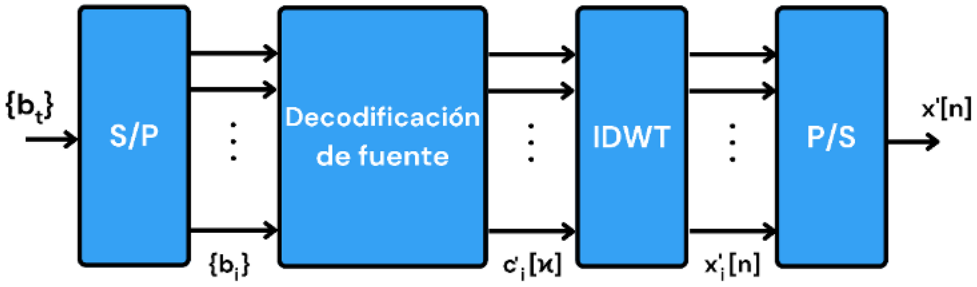


Figura 3.2: Diagrama de bloques inverso del cuantificador de señales de voz en el dominio *Wavelet* utilizando el esquema de *Lifting*.

La primera restricción que se tiene en cuenta durante el desarrollo de este proyecto es el enfoque hacia los servicios de voz, los cuales son susceptibles al retardo y, en consecuencia, no es posible esperar a tener la señal de voz completa para procesarla y enviarla. Por lo tanto, se realiza la división de las señales de voz, $x\,[n]$, en tramas, $x_i\,[n]$, y se procesa cada trama de manera individual.

En cuanto al cuantificador, que tiene como propósito principal representar la señal de voz cuantificada en el dominio *Wavelet*, $c_i'\,[\kappa]$, lo más similar posible a la señal original muestreada en el dominio *Wavelet*, $c_i\,[\kappa]$, se tiene como restricción el número de niveles de cuantificación, que, a su vez, determina la cantidad de bits disponibles para la transmisión o almacenamiento de la señal. Para esta restricción se asume que el codificador de fuente asigna palabras código de longitud fija.

La codificación busca convertir la señal de entrada en una representación más compacta utilizando un número de bits definido por las restricciones establecidas por el número de niveles de cuantificación, M, ya que esos M niveles deben ser representados de forma binaria, es decir, se debe encontrar el número de bits necesarios para representar los M posibles valores. Lo anterior se encuentra de la forma $m = \lceil \log_2 (M) \rceil$ y, por tanto, cada muestra de la señal tiene un peso de $\lceil \log_2 (M) \rceil$, con lo cual se puede calcular el máximo número de bits que se pueden asignar a cada trama de la señal. En otras palabras, la carga útil de la trama, ρ, es equivalente a la multiplicación del número de muestras totales de la trama ζ_i, por el número de bits correspondientes a cada muestra, como se observa en la ecuación 3.1.

$$\rho = \zeta_i \cdot \lceil \log_2 (M) \rceil, \tag{3.1}$$

siendo,

$$\zeta_i = F_s \cdot L, \tag{3.2}$$

donde, F_s es igual a la frecuencia de muestreo y L la longitud de la trama.

Para mejorar la eficiencia en la distribución de los recursos dados para una trama, este trabajo de grado emplea la transformada *Wavelet*, y en particular del esquema *Lifting*, para asignar de forma no uniforme los recursos disponibles en función de las diferencias entre los grupos de coeficientes, buscando que la señal de voz resultante tenga una buena calidad para el número M de niveles de cuantificación.

3.1. Metodología

Para el desarrollo del trabajo de grado se usa una adaptación de la metodología propuesta por Alan Hevner en 2007, la cual es basa en un sistema de retroalimentación dividido en tres ciclos de trabajo para cada fase de una investigación científica. Estos ciclos se plantean de la siguiente manera: el ciclo de rigor proporciona teorías y métodos; el ciclo de diseño aporta un bucle más estrecho de actividad de investigación para la construcción de artefactos y procesos de diseño; el ciclo de relevancia aporta a la investigación los requisitos y evaluación del entorno contextual e introduce los artefactos de investigación en las pruebas de cómputo con la experiencia y los conocimientos de los fundamentos base de conocimientos y añade los nuevos conocimientos generados por la investigación.

En la Figura 3.3, se observa como los diferentes ciclos tienen conexión para generar una retroalimentación entre ellos en el desarrollo de la investigación.

Figura 3.3: Metodología del trabajo de grado.

Para implementar la metodología descrita se proponen tres fases dadas por las etapas de un sistema, mostrado en la Figura 3.4, donde en cada una de ellas se desarrollan los tres ciclos de la metodología, tal como se evidencia en la Figura 3.5.

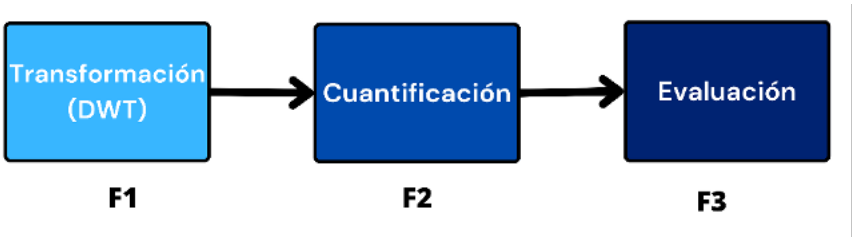


Figura 3.4: Fases para la creación del sistema.

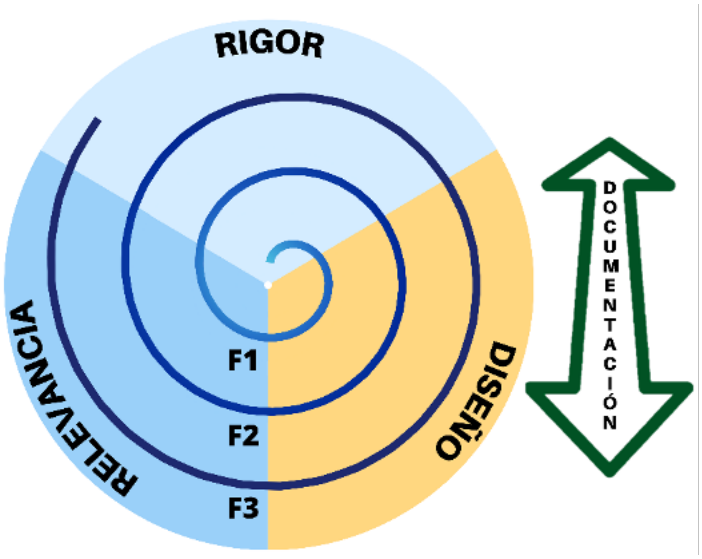


Figura 3.5: Metodología aplicada al trabajo de grado.

3.2. Requerimientos Funcionales

Es esencial identificar y especificar los requerimientos funcionales que guiarán el diseño y desarrollo del cuantificador de señales de voz en el dominio *Wavelet* utilizando el esquema *Lifting*.

Estos requerimientos abarcan desde la adquisición y pre-procesamiento de las señales de voz hasta la evaluación de la calidad y la medición de la distorsión generada durante el proceso de cuantificación y su cumplimiento se reflejará a lo largo del trabajo de grado, permitiendo que el cuantificador desarrollado logre los resultados deseados.

A continuación, se detallan los requerimientos funcionales específicos que se deben cumplir:

1. Leer y preprocesar las señales de voz.
2. Aplicar la transformada *Wavelet* y la transformada *Wavelet* inversa a las señales de voz.
3. Asignar el número de niveles de cuantificación a cada grupo de coeficientes *Wavelet*.
4. Realizar la cuantificación uniforme en función del número de niveles de cuantificación sobre cada uno de los grupos de coeficientes.

5. Medir la distorsión generada en el proceso de cuantificación.

3.2.1. Lectura y Pre-Procesamiento Señales de Voz

Para el diseño y desarrollo del cuantificador de señales de voz en el dominio *Wavelet* utilizando el esquema de *Lifting*, es necesario tener un repositorio de señales de voz que sirva como base de datos para las pruebas, así como realizar un pre-procesamiento adecuado a las señales de voz incluidas en dicho repositorio.

Repositorio de Señales de Voz

Para el repositorio de señales de voz, se consideraron dos opciones: utilizar un repositorio existente o crear uno propio. Después de evaluar las dos opciones, se opta por desarrollar un repositorio propio, para asegurar una cantidad considerable de personas con diferentes rangos de edad, que pronunciaran las mismas frases, con el fin de obtener una variedad representativa de muestras de voz que abarca las diversas características vocales de una población de hombres y mujeres de todas las edades y así obtener resultados más generalizables y aplicables. En el Apéndice A se detalla el proceso completo de construcción del repositorio, que incluye información sobre la selección de los participantes, el consentimiento informado y la construcción de las señales de voz.

Una vez creado el repositorio, se debe tener en cuenta los aspectos que son necesarios para el pre-procesamiento de las señales que además están relacionados con la digitalización y compresión de la señal dada por el entorno de trabajo MATLAB. Al pasar por este proceso, la señal se convierte en una secuencia discreta tanto en valores de amplitud como en muestras en el dominio del tiempo.

Frecuencia de Muestreo

La frecuencia de muestreo, F_s, indica la cantidad de muestras tomadas por segundo para representar la señal de voz. Es un parámetro fundamental para el procesamiento digital de señales, ya que determina la precisión con la que se captura la información en el dominio del tiempo. Es importante destacar que la elección de la frecuencia de muestreo adecuada depende de la naturaleza de la señal y de los requisitos del procesamiento posterior. Por lo tanto, en este trabajo de grado se ha elegido una frecuencia de muestreo de 16 KHz, debido a su amplia adopción y capacidad para capturar de manera efectiva la información relevante de las señales de voz. Esta frecuencia de muestreo es comúnmente utilizada en aplicaciones de voz, ya que abarca las frecuencias necesarias para representar adecuadamente las características de la voz humana, hasta 8 KHz,

por lo que según lo planteado en el teorema de muestreo de Nyquist-Shannon resulta en una frecuencia de muestreo de 16 KHz [40].

División de la Señal por Tramas

Como se mencionó anteriormente, en este trabajo de grado se busca lograr un procesamiento de la señal que se acerque lo más posible al entorno real. Para lograrlo, es necesario determinar la longitud de las tramas utilizadas en el análisis.

Inicialmente se elige una longitud de trama, L, de 64 ms la cual es ampliamente utilizada, debido a que ofrece un equilibrio adecuado entre la resolución temporal, la resolución frecuencial y la eficiencia de procesamiento [41]. Sin embargo, es importante destacar que esta elección será analizada más detalladamente en etapas posteriores de esta investigación.

Niveles de Cuantificación

Para seleccionar el número de niveles de cuantificación M, se tienen en cuenta las siguientes consideraciones:

- M se encuentra limitado por el número valores de amplitud de la señal, N.
- M debe ser un número potencia de 2, ya que esto permite una representación binaria óptima en la codificación de fuente de longitud fija.
- Se eligen valores de M hasta 64, ya que este trabajo de grado quiere demostrar que la implementación de un algoritmo de cuantificación con el esquema de *Lifting* puede presentar mejores resultados de calidad para señales cuantificadas con mayor compresión con respecto a otros cuantificadores con las mismas condiciones, en específico en el tiempo o con el algoritmo de Mallat.
- Se seleccionan valores de M mayores a 2, ya que la distorsión de la señal para niveles de cuantificación tan bajos no es recomendable en los servicios de comunicaciones.

Por tanto,

$$M = 2^m,\ \forall\ m \in \mathbb{N},\ 1 < m < 7, \tag{3.3}$$

siendo m, el número de bits disponibles para cuantificar cada muestra.

3.2.2. Implementación de la DWT

Una vez realizado el pre-procesamiento de la señal se transforma cada una de las tramas al dominio *Wavelet*, esta transformada puede ser implementada de diferentes maneras, siendo dos de las más populares el esquema de *Lifting* y algoritmo de Mallat. Para cualquiera de los dos métodos se debe establecer el número de niveles de resolución que se aplicarán a la señal.

Para este trabajo de grado, se tiene como propósito distribuir el número bits disponibles para la cuantificación dependiendo de la importancia de los grupos de coeficientes de cada trama, por tanto, es importante tener en cuenta el tamaño de la trama, ζ_i, para escoger el número de niveles de resolución usados en la transformada debido a que, con niveles de resolución muy grandes, la señal se divide tantas veces que se pueden presentar los siguientes casos:

- La señal ya no tiene más muestras para seguir siendo dividida.
- Se obtienen grupos de coeficientes con muy pocas muestras, que no tienen mucha relevancia en la señal.

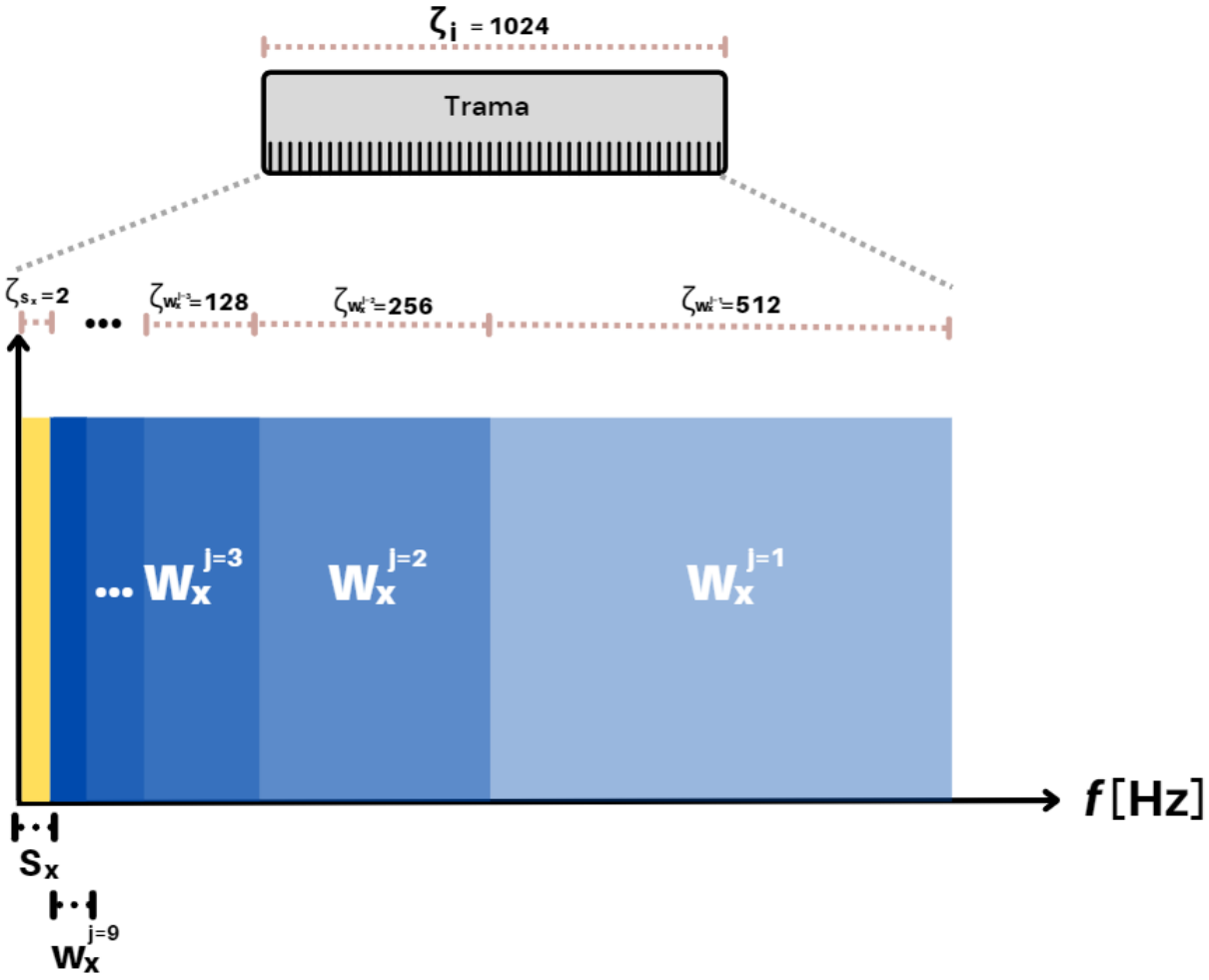


Figura 3.6: Muestras por coeficiente.

Es por eso, que se establece como requisito que existan al menos 2 muestras en los coeficientes *Scaling* para el nivel de resolución seleccionado. Por ejemplo, si la trama consta de 1024 muestras ($\zeta_i = 1024$) y se realiza la transformada *Wavelet* con 9 niveles de resolución, el grupo de coeficientes *Scaling* estará compuesto por 2 muestras como se observa en la Figura 3.6. De lo contrario, si se escogen 10 niveles de resolución el coeficiente *Scaling* trendrá una sola muestra, la cual, a pesar de tener un gran aporte en la energía de la señal, no es adecuada para ser considerada de forma individual como grupo de coeficientes, dado que se debe diseñar para ella un cuantificador específico y se requieren muchos recursos para representar este único valor.

Luego de definir los niveles de resolución que se van a usar, se aplica la transformada *Wavelet* y su inversa a la señal de voz, $x\,[n]$, utilizando tanto el esquema de *Lifting* como el algoritmo de Mallat, con el fin de analizar si para estos dos méto-

dos el proceso de transformación es completamente reversible, i.e., sin pérdidas de información. Por lo tanto, se analiza si al final del proceso de transformación se puede obtener nuevamente $x\,[n]$, como se observa en la Figura 3.7.

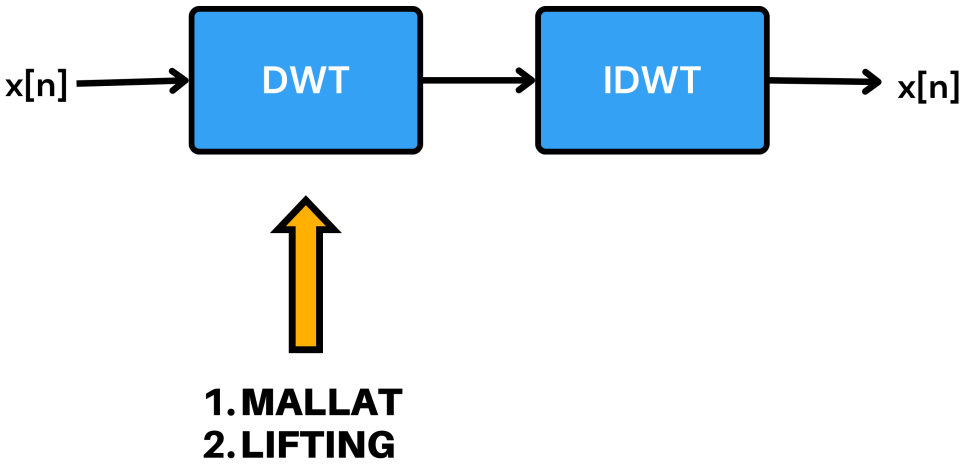


Figura 3.7: Validación transformada *Wavelet* con dos diferentes métodos para transformar.

En la Figura 3.8, se evidencia la calidad[6] de las señales después de realizar el proceso de transformación y posteriormente revertir este proceso, utilizando el esquema de *Lifting* y el algoritmo de Mallat, para todas las familias *Wavelet* compartidas por estos dos métodos. Es importante resaltar que este proceso se hace variando los niveles de resolución del uno al nueve.

[6]En este trabajo de grado los resultados presentados en términos de *calidad* hacen referencia al valor obtenido del promedio de la PESQ y el NMSE, entre la señal original y la señal procesada.

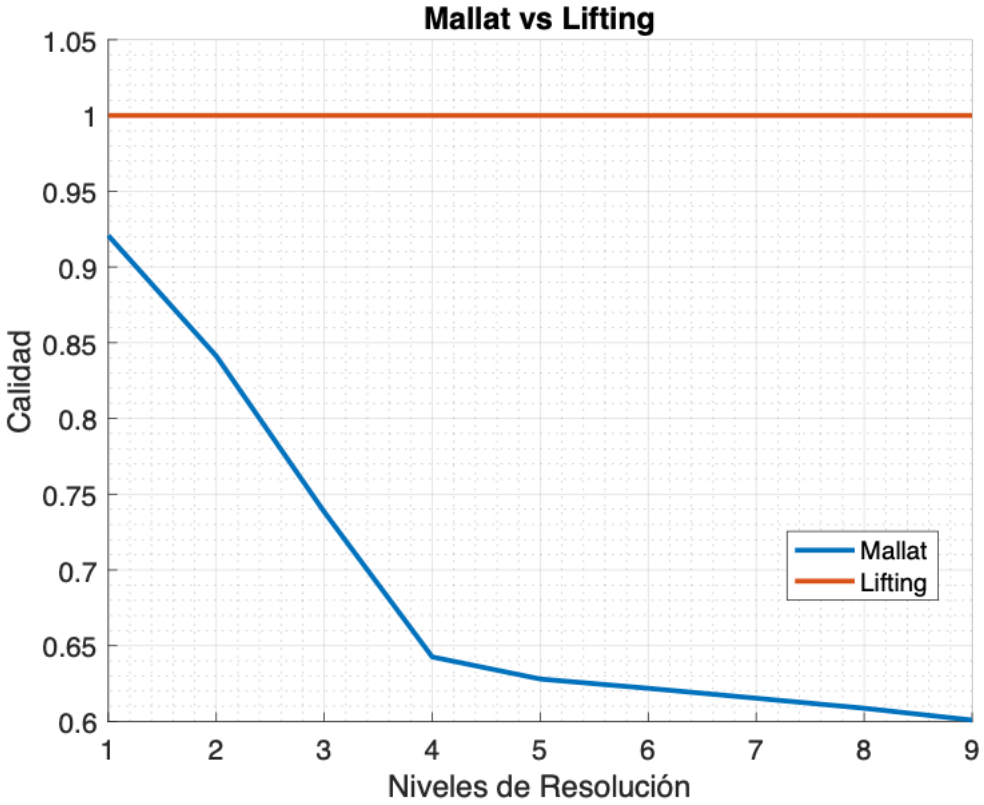


Figura 3.8: Fidelidad de la señal reconstruida vs Niveles de resolución para *Lifting* y Mallat.

Los resultados obtenidos en la Figura 3.8, determinan que:

- Transformar una señal con el esquema de *Lifting* permite recuperar la señal original en su totalidad, por tanto, es un método que permite transformar la señal al dominio *Wavelet* sin inducirle distorsión, aspecto fundamental para esta investigación, ya que un cuantificador por defecto induce distorsión a la señal, por lo que se deben evitar al máximo otros factores adicionales que continúen deteriorando la calidad.
- Transformar una señal con el algoritmo de Mallat no permite recuperar sin distorsión la señal original. Incluso, se evidencia que la fidelidad de la señal recuperada disminuye progresivamente a medida que los niveles de resolución aumentan. Esto quiere decir que el proceso realizado por el algoritmo de Mallat es un proceso que introduce distorsión, el cual se acentúa cuando se realiza una descomposición con un mayor número de niveles de resolución. Si se analizan los resultados de forma más detallada, es posible concluir que este comportamiento no es propio del 100 % de familias *Wavelet*, sino que existen algunas familias *Wavelet* que, junto con el algoritmo de

Mallat, también permiten la recuperación de la señal original sin distorsión alguna (*Haar*, *bior1.1* y *rbio1.1*).

- Tras analizar y comparar todas las familias *Wavelet* con estas familias excepcionales, se concluye que la diferencia puede estar relacionada con el tamaño de los filtros asociados a cada familia, pues el algoritmo de Mallat consiste en la separación de coeficientes mediante convoluciones sucesivas, las cuales intrínsecamente conllevan la respuesta transitoria de los filtros que depende de su longitud. El manejo de la respuesta transitoria se hace más delicado entre más similares sean las longitudes del filtro y de la señal, por lo que, desde la perspectiva de la distorsión que se agrega, los casos con familias de filtros más complejos o descomposiciones con mayor número de niveles de resolución son más críticos.

En la Tabla 3.1 se muestran las 6 familias *Wavelet* que se usan para el desarrollo de las pruebas realizadas en este trabajo de grado, las cuales se escogen entre el conjunto de familias que se pueden implementar tanto para el esquema *Lifting* como para el algoritmo de Mallat, buscando que éstas tengan un número diferente de pasos *Lifting* y matrices (**U** y **P**), lo cual equivale a utilizar filtros de diferente longitud ($g\,[n]$ y $h\,[n]$). Adicionalmente, se utilizan familias ortogonales y biortogonales.

Tabla 3.1: Familias usadas en las pruebas.

Familia Wavelet	**Pasos Lifting**	**Tamaño de las Matrices**	**Longitud de los Filtros**
Haar	2x1	1	2
rbio4.4	4x1	2	10
bior5.5	6x1	2	12
sym6	7x1	2	12
db7	8x1	2	14
bior6.8	6x1	2 y 4	18

Como se explicó al inicio de este capítulo, después de obtener la señal original muestreada en el dominio *Wavelet*, $c_i\,[\kappa]$, se procede a cuantificarla. Para ello se debe tener muy claro el propósito de la transformada *Wavelet* y la manera en que se usa para beneficio de la cuantificación. Es por esto que el cuantificador de señales de voz en el dominio *Wavelet* utilizando esquema *Lifting* tiene como finalidad discriminar los grupos de coeficientes que son más relevantes para cada trama de la señal de voz y de esta manera asignarles más niveles de cuantificación, obteniendo así, una señal cuantificada con mayor precisión en los valores

de amplitud de los coeficientes que aportan más significativamente a la calidad de la señal resultante.

Por consiguiente, este trabajo de grado propone un algoritmo que permite distribuir los niveles de cuantificación a cada grupo de coeficientes de cada trama, considerando las características de dichos coeficientes en diferentes enfoques y teniendo en cuenta la limitante del número de niveles de cuantificación que tiene cada trama.

3.2.3. Algoritmo de Asignación de Bits

Este algoritmo de asignación busca distribuir los bits disponibles para cada trama, de manera que los coeficientes más relevantes dispongan de más bits ($M = 2^m$, m: número de bits disponibles para cuantificar cada muestra) para cuantificar sus muestras. Teniendo en cuenta la restricción del número de niveles de cuantificación, la estructura de este algoritmo se basa en los siguientes principios:

- Se debe procurar representar la información de todos los coeficientes de una trama.
- La información contenida en los grupos de coeficientes de una trama aporta en diferente medida en la calidad de la señal de voz reconstruida, por lo cual es posible crear una escala de importancia o de prioridad, con el fin de disminuir la distorsión en los grupos de coeficientes más cruciales.
- Se debe procurar que por la distribución desigual de niveles de cuantificación no se desperdicien recursos, en consecuencia, se debe hacer una validación final para distribuir los bits sobrantes [7].

Por lo anterior, el algoritmo de asignación de bits está compuesto por 4 bloques principales como se muestra en la Figura 3.9. De manera general, el bloque 1 realiza la primera asignación de bits a los coeficientes con los denominados bits de reserva, r. El bloque 2 hace uso de diferentes métodos para calcular los Porcentajes de Relevancia de cada grupo de Coeficientes (CRR, *Coefficient Relevance Rate*). El bloque 3 realiza la asignación de bits con respecto a los porcentajes dados por el bloque 2 y, finalmente, el bloque 4 asigna los bits sobrantes para que no exista desperdicio de bits.

[7] Se define como bits sobrantes a aquellos bits que, después de la repartición inicial a cada grupo de coeficientes, quedan sin asignar debido a que resultan insuficientes para cubrir la cantidad de muestras contenidas en el grupo de coeficientes al que originalmente se designaron.

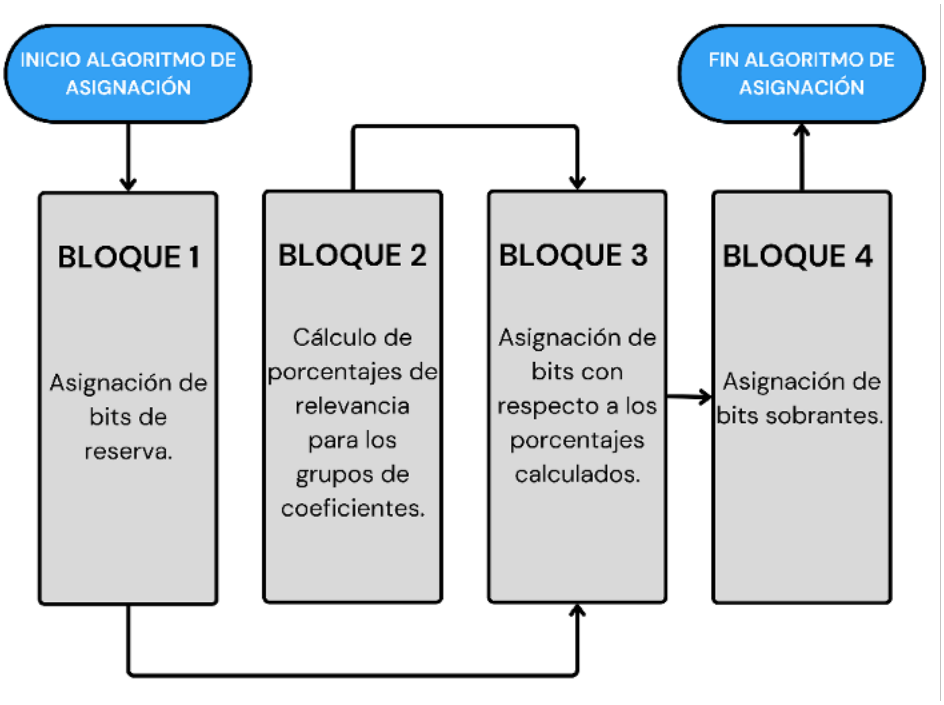


Figura 3.9: Diagrama general del algoritmo.

En la Figura 3.10 se observa más detalladamente el funcionamiento del bloque 1, que tiene como entradas: el número de niveles de cuantificación, M, y los grupos de coeficientes de la trama, $S_i^j,\ W_i^j,\ \ldots\ ,\ W_i^1$, a los cuales se les asignan los bits de reserva de manera equitativa.

Es importante resaltar que los bits de reserva, r, son relevantes porque inicialmente se busca que todas las muestras de la señal tengan asignado un número mínimo de bits por muestra para realizar la cuantificación, con el fin de mantener la calidad de la señal resultante. Por otro lado, permiten disminuir el costo computacional, dado que la repartición de bits por coeficiente empieza con un número diferente a cero, i.e., se tiene un menor número de recursos por distribuir.

Una vez asignados los bits de reserva a los coeficientes de la señal se actualiza el vector de bits asignados por coeficiente, $\beta c_i\,[\kappa]$, y con ello se puede realizar la sumatoria de elementos del vector $\beta c_i\,[\kappa]$ para obtener el resultado correspondiente al número de bits asignados a la trama en general, βu_i, i.e., la cantidad de bits utilizados del total disponible.

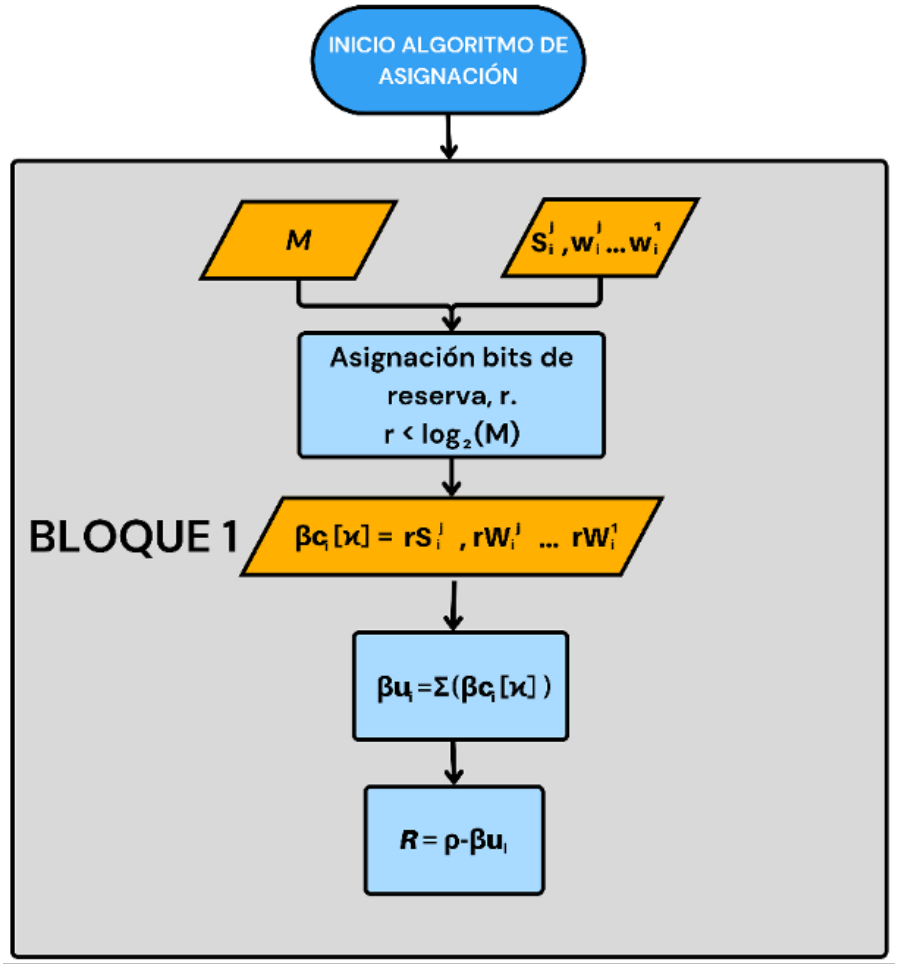


Figura 3.10: Bloque 1. Asignación bits de reserva.

Finalmente se realiza el cálculo del número de bits sobrantes para asignación de la trama, R, que resulta de la resta entre el máximo número de bits que se pueden asignar a cada trama de la señal, ρ, y βu_i.

Para aprovechar las características de la transformada *Wavelet* que permite la división de la señal en coeficientes, en este trabajo de grado se proponen tres métodos para calcular los CRR (Percepción, Energía y Heurístico), los cuales permiten que se asigne mayor cantidad de niveles de cuantificación a los coeficientes que aportan más a la calidad de la señal. Es por ello que en el bloque 2 se tiene como entrada los coeficientes de la trama, a los cuales se les aplica el método para calcular los CRR, obteniendo como resultado un vector de porcentajes de relevancia $P_i[\kappa]$ como se observa en la Figura 3.11.

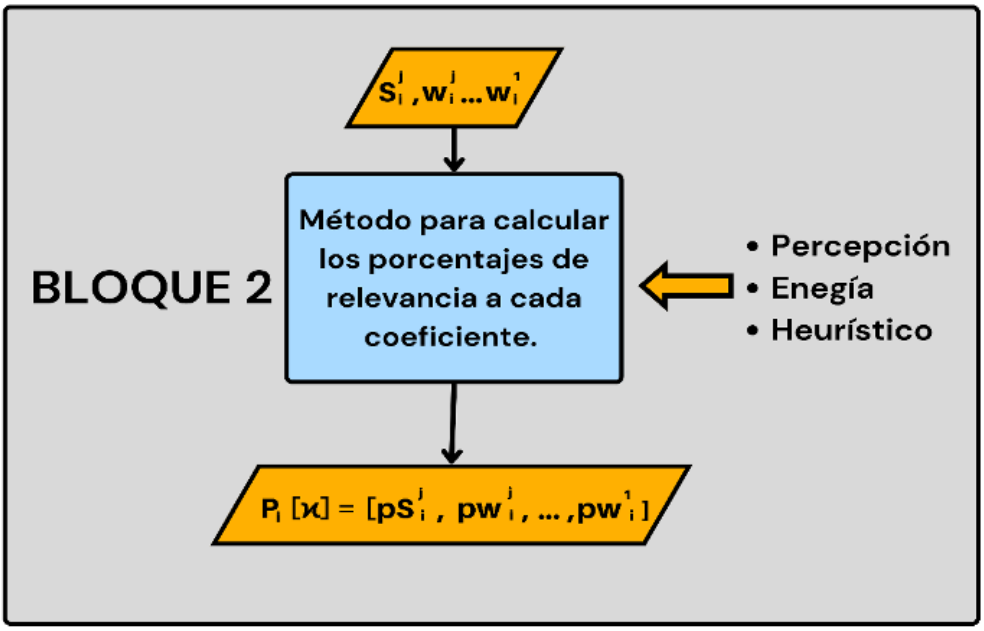


Figura 3.11: Bloque 2 - Cálculo de los CRR.

Una vez obtenido el vector $P_i[\kappa]$, se puede realizar el proceso de asignación de bits con respecto a los porcentajes calculados. Este proceso se realiza en el bloque 3, como se observa en la Figura 3.12, en el cual se ordenan de manera descendente los valores de $P_i[\kappa]$, y se almacenan en un nuevo vector llamado valores ordenados. Adicionalmente, se crea un vector ubicaciones ordenadas que contiene las posiciones de los valores en el vector original, $P_i[\kappa]$, pero en relación con el ordenamiento descendente. Es decir, muestra la posición que tenía cada valor en $P_i[\kappa]$ antes de ser ordenados.

Seguidamente se reparten los bits sobrantes, R, a cada grupo de coeficientes, en función de los porcentajes dados en $P_i[\kappa]$, obteniendo un vector $\beta p_i[\kappa]$, que contiene los bits por muestra asignados a cada grupo de coeficientes con los CRR. No obstante, el resultado de estos productos no siempre es un múltiplo entero de las longitudes de los grupos de coeficientes, las cuales siempre son una potencia de 2.

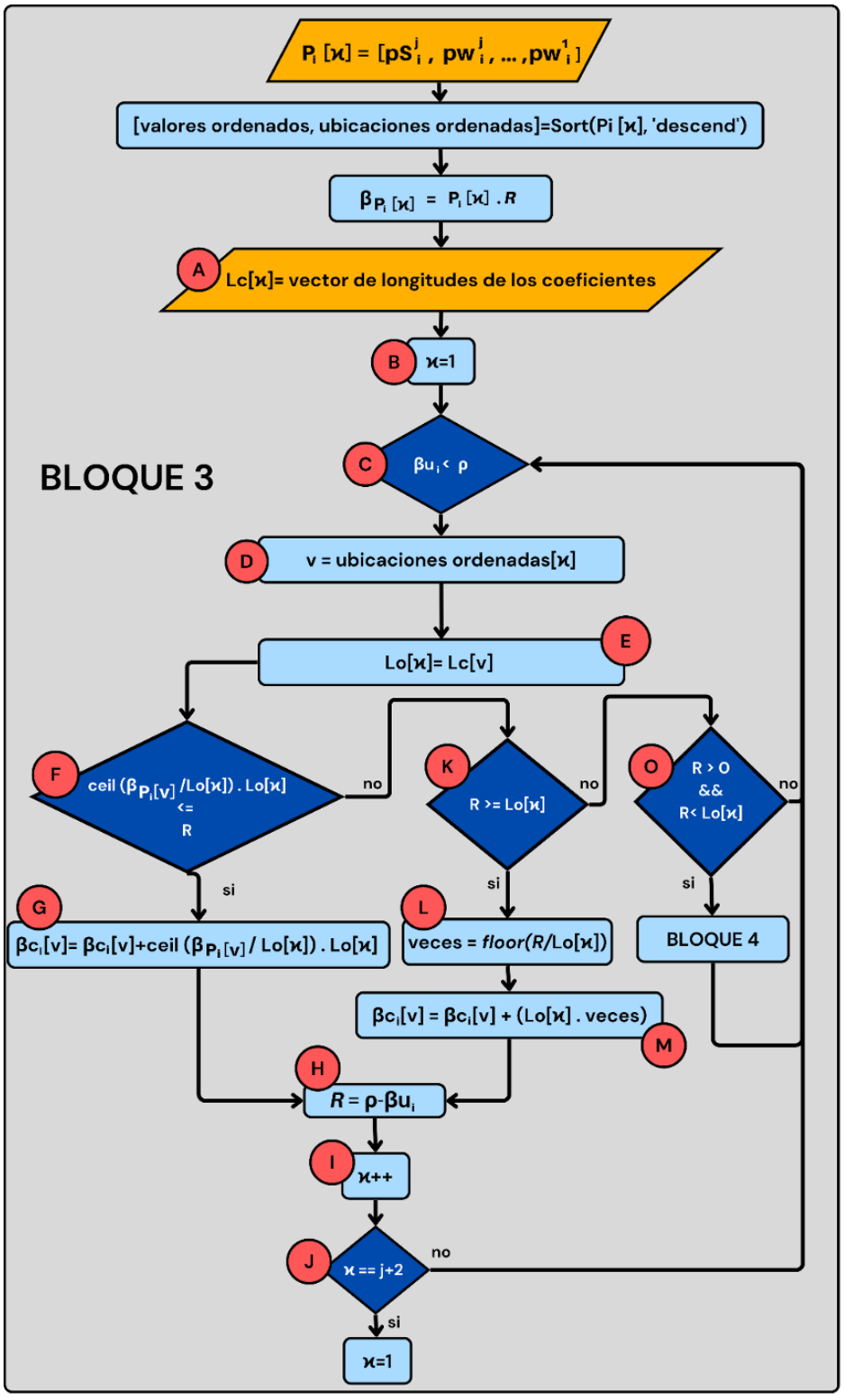


Figura 3.12: Bloque 3 - Asignación de bits con respecto a los porcentajes calculados.

Dado que existe la posibilidad de que a partir de los CRR no se repartan de forma exacta los bits disponibles, se agrega una lógica al algoritmo que permite repartir los bits, teniendo en cuenta la longitud de cada grupo de coeficientes y su relevancia, así:

(A) Se crea un vector de longitudes de los grupos de coeficientes, $Lc\left[\kappa\right]$.

(B) Se inicializa la variable κ en 1, la cual va a recorrer los vectores que contienen a los grupos de coeficientes, es decir, esta variable tiene como máximo valor $j+1$, donde j es el número de niveles de resolución.

(C) Se inicia un ciclo, que se va a repetir mientras se cumpla la condición de que el número de bits asignados a la trama sea menor al número de bits máximos que se le puede asignar a cada trama[8],

$$\beta u_i < \rho. \tag{3.4}$$

(D) Se crea la variable v que permite recorrer el vector con las $ubicaciones\ ordenadas$.

(E) Se crea la variable $Lo\left[\kappa\right]$, la cual almacena la longitud de un coeficiente que se obtiene a partir de la lista de ubicaciones ordenadas, que es recorrida con ayuda de la variable κ, es decir que, se van almacenando en $Lo\left[\kappa\right]$ las longitudes de los coeficientes de mayor relevancia hasta llegar al de menor relevancia.

(F) Se evalúa la primera condición, dada por la parte entera superior de la división de los bits asignados al coeficiente en la posición de relevancia κ, $\beta p_i\left[v\right]$, y $Lo\left[\kappa\right]$, para poder asignar un número entero de bits a cada uno de los coeficientes del grupo κ–ésimo, este resultado debe ser menor que los bits sobrantes, R.

(G) En caso de que la condición de F se cumpla, se asigna al grupo de coeficientes evaluado, los bits necesarios para que todas sus muestras tengan la misma cantidad de bits, así,

$$\beta c_i\left[v\right] = \beta c_i\left[v\right] + \left\lceil \frac{\beta p_i\left[v\right]}{Lo\left[\kappa\right]} \right\rceil \cdot Lo\left[\kappa\right]. \tag{3.5}$$

(H) Se actualiza el valor de los bits sobrantes R.

[8]Los recursos asignados a una trama se reparten entre los grupos de coeficientes, estos recursos se pueden ver en términos de niveles de cuantificación o del número de bits utilizados para representar el valor de cada coeficiente; no obstante, la restricción con respecto al número máximo de recursos que se puede usar en cada trama siempre se cumple.

(I) Se incrementa el valor de κ.

(J) Se valida si κ excede el número de grupos de coeficientes dado, si es así se asigna nuevamente el valor de 1 a la variable, si no el algoritmo se devuelve al paso C.

(K) En caso de que la condición F no se cumpla, se evalúa la segunda condición, la cual verifica que los bits sobrantes R sean mayores o iguales al tamaño del grupo de coeficientes evaluado $Lo\left[\kappa\right]$, esto con el fin de poder asignarle al menos un bit a cada coeficiente, en caso de que no se le haya podido asignar el número de muestras sugerido por los porcentajes de relevancia.

(L) En caso de que la condición K se cumpla, se evalúa cuantos bits se le puede asignar a cada muestra del coeficiente de la siguiente manera,

$$\text{veces} = \left\lfloor \frac{R}{Lo\left[\kappa\right]} \right\rfloor . \tag{3.6}$$

(M) Se asigna al coeficiente evaluado, los bits disponibles para asignarle a cada una de las muestras el mismo número de bits, así,

$$\beta c_i\left[v\right] = \beta c_i\left[v\right] + \left(Lo\left[\kappa\right] \cdot \text{veces}\right). \tag{3.7}$$

(N) Se repiten los pasos H, I, J.

(O) En caso de que la condición K no se cumpla, se evalúa la última condición que es el caso en el que los bits sobrantes R, son mayores a 0 pero son menores a la longitud del coeficiente evaluado $Lo\left[\kappa\right]$. Esta condición se desarrolla en el bloque 4 del algoritmo.

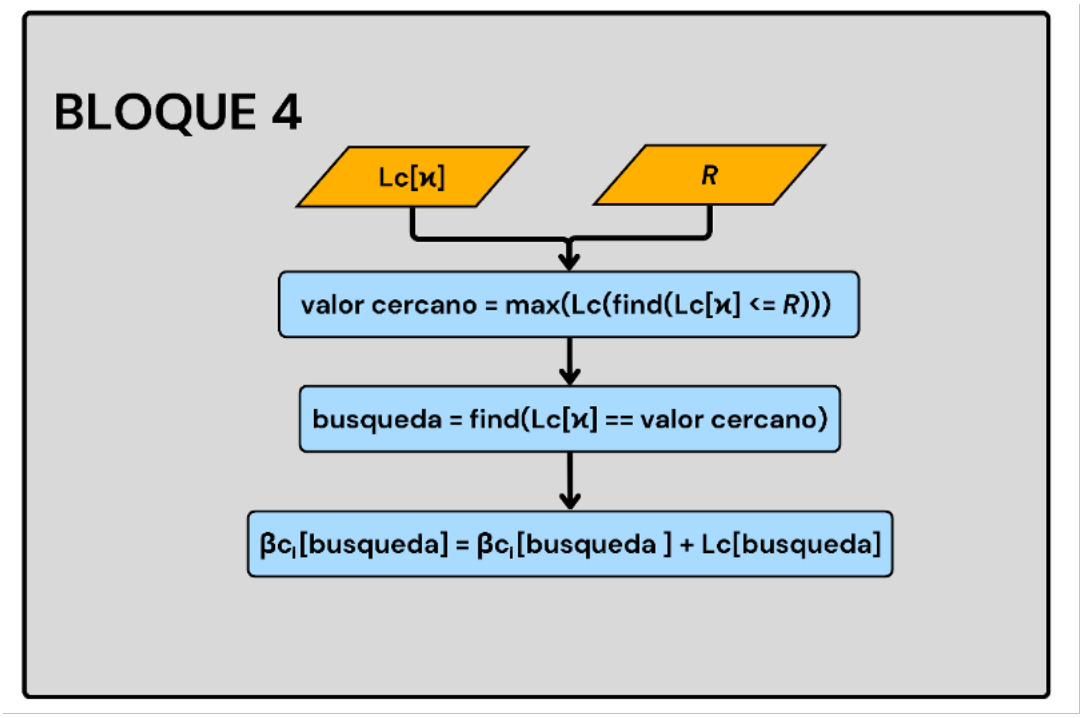


Figura 3.13: Bloque 4 - Asignación bits sobrantes.

En la Figura 3.13 se observa el bloque 4 del algoritmo, en el cual se asignan los bits sobrantes, ya que al no ser suficientes para ser asignados al coeficiente evaluado se busca un grupo de coeficientes al que sí se le puedan asignar, encontrando el grupo con longitud más cercana al valor de los bits sobrantes R, así,

$$\text{valor cercano} = \max\left\{Lc\left[\text{find}\left\{Lc\left[w\right] \leq R\right\}\right]\right\}. \tag{3.8}$$

La variable valor cercano contiene la longitud del coeficiente que mejor se ajusta al valor de los bits sobrantes R. Para poder encontrar la posición del coeficiente se hace la búsqueda del valor en $Lc\left[w\right]$ y se guarda la posición en la variable busqueda, la cual es una variable que se usa para poder asignar los bits al coeficiente, así,

$$\beta c_i\left[\text{busqueda}\right] = \beta c_i\left[\text{busqueda}\right] + Lc\left[\text{busqueda}\right]. \tag{3.9}$$

De este modo el algoritmo se asegura que todos los bits disponibles para cada trama se asignen a los coeficientes, sin desperdicio y teniendo en cuenta que unos coeficientes tienen más relevancia que otros.

Una vez asignados los bits de cuantificación disponibles para cada trama a los coeficientes $\beta c_i\left[\kappa\right]$, se identifica cuantos bits corresponden a cada muestra, $m_i\left[k\right]$

dividiendo los bits asignados para cada coeficiente entre su respectiva longitud, así,

$$m_i\left[\kappa\right] = \frac{\beta c_i\left[\kappa\right]}{Lc\left[\kappa\right]}. \tag{3.10}$$

Con el valor de $m_i\left[\kappa\right]$, se puede saber finalmente cuantos niveles de cuantificación dispone cada grupo de coeficientes, $M_i\left[\kappa\right]$, recordando que $M_i = m_i$.

A continuación, se realiza un ejemplo numérico con porcentajes de relevancia arbitrarios, una trama con un número de muestras totales $\zeta_i = 32$, un número de niveles de cuantificación $M = 8$ y un número de niveles de resolución $j = 4$.

En la Figura 3.14 se muestran en color amarillo los bits de reserva asignados de forma constante a todos los grupos de coeficientes (bloque 1). A partir de los CRR (bloque 2) se asignan unos bits adicionales para cada grupo de coeficientes (bloque 3), los cuales se muestran en color azul. Finalmente, los bits sobrantes se reparten entre los grupos de coeficientes buscando que no se desperdicien recursos (bloque 4), estos bits se muestran en color rojo.

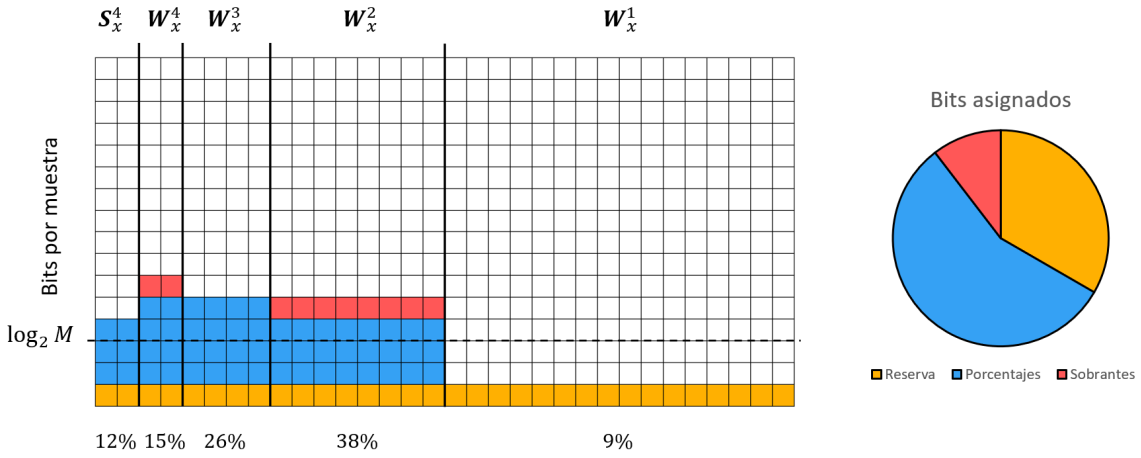


Figura 3.14: Ejemplo de asignaciones de bits.

Con el fin de complementar el ejemplo mostrado en la Figura 3.14, se utilizan los datos consignados en la Tabla 3.2 para mostrar matemáticamente la forma en la que, para este caso particular, se obtienen las diferentes clases de bits para cada grupo de coeficientes. El número de niveles de cuantificación por cada grupo de coeficientes corresponde a 2 elevado a la suma de las 3 clases de bits, e.g., para W_x^4 los niveles de cuantificación a utilizar son $M = 2^6 = 64$.

Tabla 3.2: Valores del ejemplo de asignación de bits.

Grupo de Coeficientes	$Lc\,[\kappa]$	CRR	βp_i	Bits de Reserva	Bits Porcentajes	Bits Sobrantes
S_x^4	2	12 %	7.68	1	3	0
W_x^4	2	15 %	9.6	1	4	1
W_x^3	4	26 %	16.64	1	4	0
W_x^2	8	38 %	24.32	1	3	1
W_x^1	16	9 %	5.76	1	0	0

3.2.4. Cuantificador Uniforme

El proceso de cuantificación es uniforme para las muestras de cada grupo de coeficientes en particular; sin embargo, ya que cada grupo de coeficientes tiene su cuantificador, se considera que sobre cada trama se realiza un proceso de cuantificación no uniforme, dado que por cada trama existen $j+1$ cuantificadores. Por tanto, el bloque del cuantificador uniforme hace uso de parámetros adaptativos que varían según las características de cada grupo de coeficientes.

Como se observa en la Figura 3.15 los parámetros adaptativos utilizados para cada grupo de coeficientes en el proceso de cuantificación son: amplitud máxima, $A_M\,[\kappa]$, amplitud mínima, $A_m\,[\kappa]$, y número de niveles de cuantificación, $M_i\,[\kappa]$.

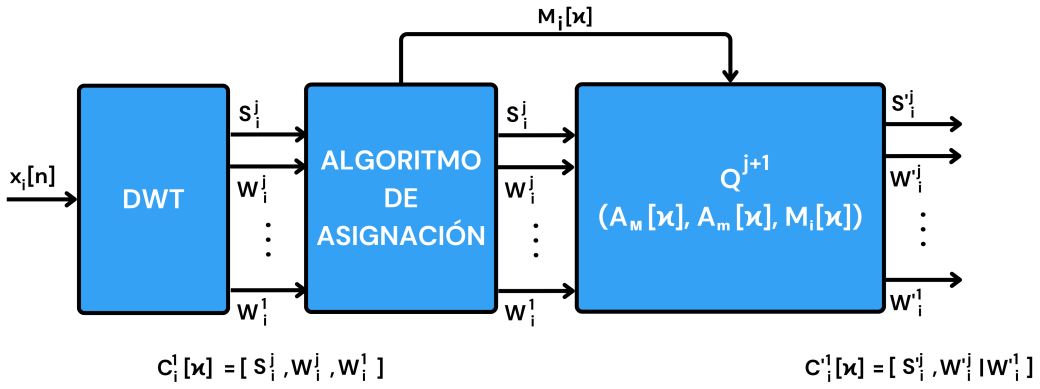


Figura 3.15: Cuantificador uniforme adaptativo.

En cuanto al proceso de cuantificación uniforme *per se*, presenta como característica principal que los niveles de cuantificación se encuentran equi-espaciados en la recta real, esa distancia entre cada nivel es llamada tamaño de escalón, Δ.

El valor de Δ se encuentra restando el máximo, $A_M[\kappa]$, y mínimo, $A_m[\kappa]$, lo cual corresponde al valor del rango dinámico. Finalmente, este valor se divide entre el número de niveles de cuantificación, $M_i[\kappa]$, así,

$$\Delta = \frac{A_M[\kappa] - A_m[\kappa]}{M_i[\kappa]}. \tag{3.11}$$

A partir de la ecuación 3.11 se deduce que entre mayor sea el valor del número de niveles de cuantificación, menor es la distorsión que introduce el cuantificador, dado que el valor de Δ es menor y, por lo tanto, las aproximaciones que realiza el cuantificador son más pequeñas.

En cuanto a los valores de los niveles de cuantificación, v_M, se ubican en el punto medio de cada escalón, es decir, el primer nivel de cuantificación, v_1, está dado por $A_m[\kappa] + {}^{\Delta}/_{2}$ y de ahí en adelante quedan equi-espaciados cada Δ, hasta llegar a $v_M = A_M[\kappa] - {}^{\Delta}/_{2}$.

Finalmente, las regiones de cuantificación sin intervalos de extensión Δ, que se centran en el nivel de cuantificación correspondiente, i.e., la primera región de cuantificación está comprendida desde $A_m[\kappa]$ hasta $A_m[\kappa] + \Delta$ y los valores de amplitud englobados en este rango, toman el valor de v_1 a la salida del cuantificador. La siguiente región está comprendida entre $A_m[\kappa] + \Delta$ y $A_m[\kappa] + 2\Delta$ y así sucesivamente, teniendo en cuenta que los valores de cada intervalo no comprenden el límite inferior, pero si comprenden el límite superior [42].

Por lo anterior se resalta que, los $j + 1$ cuantificadores que tiene cada una de las tramas de la señal de voz resultan del hecho de que la característica de transferencia del cuantificador depende de los parámetros: $A_M[\kappa]$, $A_m[\kappa]$ y $M_i[\kappa]$; los cuales se actualizan en función de los datos de cada grupo de coeficientes.

3.2.5. Medidas de Distorsión

Para evaluar la distorsión de la señal cuantificada con el algoritmo de asignación de bits propuesto y sus diferentes variantes, se utilizan medidas objetivas y subjetivas, las cuales se nombran a continuación.

Medidas Objetivas

Para escoger estas medidas, se determina que la elección debe tener en cuenta una medida de comparación directa, proporcionando un valor numérico que representa la similitud entre las señales original y cuantificada; y otra medida basada en percepción, para tener en cuenta de forma aproximada las características perceptuales del sistema auditivo humano. Adicionalmente, se busca que las

medidas escogidas sean acotadas, para que, al obtener los resultados individuales de las medidas, se pueda hacer la conversión a porcentajes y de esta forma realizar un promedio entre la medida de comparación directa y la medida de percepción, obteniendo así un resultado global, el cual tiene en cuenta esos dos aspectos fundamentales para determinar la calidad de la señal resultante y de esta manera se busca obtener un resultado más acertado.

Teniendo en cuenta las consideraciones anteriores, se descartan las siguientes medidas abarcadas en el Capítulo 1:

- **MSE.** A pesar de ser una de las medidas de distorsión más famosas, la cual evidencia la diferencia entre la señal original y la señal procesada tiene la desventaja de ser susceptible a los cambios de escala y desfases, pues al evaluar la diferencia de las n-ésimas muestras de la señal original y de la señal procesada, no se tiene en cuenta si ambas señales están sincronizadas, así como tampoco se tiene en cuenta su escala; por lo que la misma señal con un leve desplazamiento (representado como un desfase) obtendría valores altos de MSE, indicando una falsa diferencia de la señal, misma consecuencia que se da en el caso donde se compara la misma señal con diferentes escalas, pues la diferencia de tamaño de las señales también resulta en valores altos de MSE; sin embargo, esto no implica una diferencia en cuanto a información contenida en las señales.

- **SSNR.** Presenta un problema potencial y es que la energía de la señal del habla durante los intervalos de silencio (que son abundantes en el habla conversacional) es muy pequeña, lo que da lugar a grandes valores negativos de SSNR, que sesgan la medida global. Además, en la literatura se demuestra que no presenta una buena correlación con las medidas subjetivas y se propone no usarla en el contexto de señales de voz.

- **MSSIM.** Está orientada principalmente a medir la calidad de las imágenes, es decir que mide la similitud entre dos imágenes. Es una métrica que tiene en cuenta la percepción visual humana y que se utiliza para comparar una imagen de referencia con una imagen distorsionada. Por lo que no se encuentra diseñada para medir el tipo de señales evaluadas en este trabajo de grado, el cual abarca señales de voz.

- **BDM.** Aunque es motivada por la percepción, presenta una desventaja en los momentos de silencio ya que se obtienen valores desfavorables, generando así una sensación de distorsión incluso en señales con poca distorsión real.

Es por esto que para este trabajo de grado escogen el NMSE y la PESQ como medidas para evaluar la calidad de los audios obtenidos después de realizar el proceso de cuantificación.

Recordando que el NMSE, es una medida de comparación directa que mejora los problemas de normalización descritos con la MSE, i.e., presenta unos valores acotados entre 0 y 1. Si la señal cuantificada es idéntica a la señal original, el NMSE será igual a uno. Por otro lado, si la señal cuantificada no se parece en nada a la señal original, el NMSE será igual a cero. Esto proporciona una medida estandarizada de la calidad de la cuantificación que puede ser fácilmente interpretada.

Por otro lado, la PESQ es una medida de percepción desarrollada por la Unión Internacional de Telecomunicaciones (ITU, *International Telecommunication Union*) y se implementa a partir de la recomendación ITU-T P.862 [43]. Esta medida tiene una alta correlación con las medidas subjetivas y, por ende, es la medida objetiva más confiable para la evaluación de la calidad de señales de voz. Adicionalmente se ha convertido en una de las medidas de calidad de voz más utilizadas y reconocidas en la industria de las telecomunicaciones y se encuentra acotada en un rango de -0.5 a 4.5, donde los valores más altos indican una mejor calidad de voz percibida por el oyente.

Medidas Subjetivas

Las medidas subjetivas analizadas en el Capítulo 1, contemplan los RPM, los ACRM y existen algunas diferencias importantes entre ellos, cuando se tiene en cuenta los siguientes aspectos:

- **Escala de medida.** en los RPM, la escala de medida es relativa y se basa en la comparación entre dos o más estímulos. Por lo tanto, los participantes deben indicar cuál de los estímulos les gusta más o cuál prefieren. En cambio, en los ACRM, la escala de medida es absoluta y se basa en la evaluación de un solo estímulo, donde los participantes deben calificar el estímulo en una escala predeterminada.
- **Sensibilidad.** Los RPM son más sensibles que los ACRM, ya que permiten una mayor discriminación entre estímulos. Como los participantes comparan dos o más estímulos, se les permite evaluar las diferencias sutiles entre ellos. Por otro lado, en los ACRM, los participantes deben evaluar un solo estímulo y, por lo tanto, las diferencias entre estímulos pueden ser menos obvias.
- **Facilidad de uso.** Los ACRM son más fáciles de usar que los RPM, ya que los participantes solo deben calificar un estímulo en una escala predeterminada. En cambio, los RPM requieren más tiempo y esfuerzo del participante, ya que deben comparar dos o más estímulos y tomar una decisión.

Por estas razones se decide realizar las rubricas de medición subjetivas con las dos medidas, ya que los RPM son útiles para comparar y clasificar estímulos

en función de la preferencia relativa y los ACRM son adecuados para evaluar la calidad o la aceptabilidad absoluta de un solo estímulo.

Finalmente se realiza una adaptación de los CMOS para los RPM y se escoge MOS como medida dentro de los ACRM descartando el STMESCSINSA, ya que tiene en cuenta la SBI y esto puede generar confusiones en los calificadores, debido a que es posible que en las grabaciones originales se perciban algunos pequeños ruidos de fondo. Además, se descarta la medida DAM puesto que requiere de una mayor cantidad de tiempo y de evaluadores cuidadosamente seleccionados tras una serie de validaciones, capacitaciones y calibraciones.

En el Apéndice B se detalla el procedimiento por medio del cual se aplican las rúbricas seleccionadas para este trabajo de grado.

3.3. Variables para Considerar en el Proceso de Diseño

En el proceso de diseño de un cuantificador de señales de voz en el dominio *Wavelet* con el esquema *Lifting*, existen diversas variables que desempeñan un papel fundamental en el costo de procesamiento y en la calidad de la señal resultante. Por ello, en esta sección, se estudian detalladamente estas variables, realizando la variación de sus posibles valores para finalmente analizar su impacto en el resultado final de la cuantificación, y de esta manera dar una recomendación de diseño[9].

En la Figura 3.16 se describen las variables que afectan a cada bloque. Los niveles de resolución, j, afectan directamente a la DWT. Por otro lado, los bits de reserva, r, junto con el método para encontrar los CRR, representan un impacto en el bloque del algoritmo de asignación.

[9]Todas pruebas realizadas en este apartado usan el Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Esquema *Lifting* para procesar los 300 audios del repositorio y posteriormente se promedian estos resultados. Además, se realiza el mismo proceso para 6 familias *Wavelet*, considerando las variaciones de parámetros que se necesiten para cada prueba en particular.

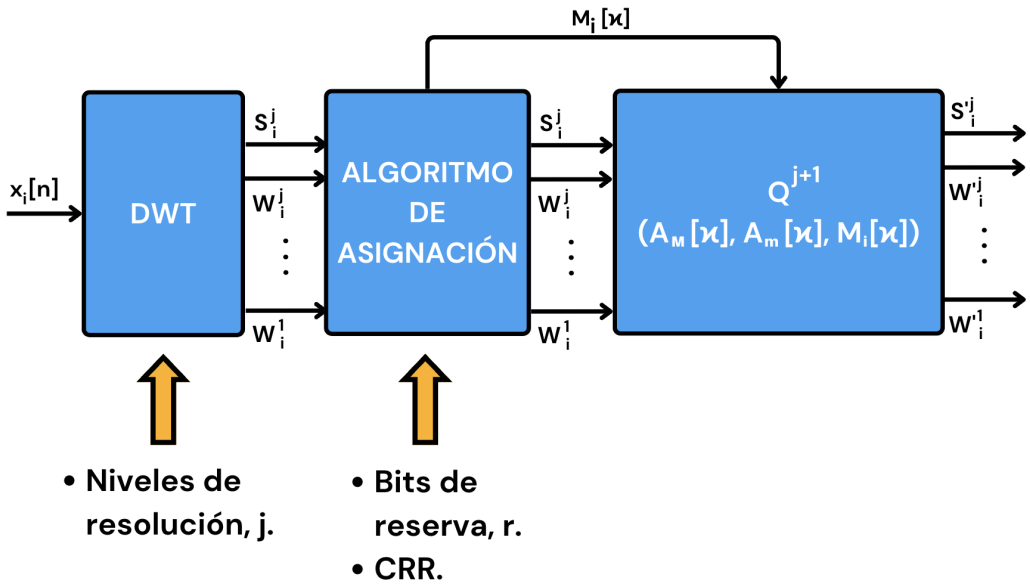


Figura 3.16: Variables para el diseño del Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Esquema *Lifting*.

3.3.1. Porcentajes de Relevancia por Grupo de Coeficientes - CRR

Como lo dice su nombre, estos porcentajes permiten que el algoritmo de asignación reparta los bits de cuantificación, $m_i\,[k]$, según la relevancia de los grupos de coeficientes, la cual es analizada desde diferentes enfoques. En este trabajo de grado se proponen tres métodos que permiten asignar los porcentajes desde enfoques diferentes. A continuación, se describen.

Método CRR de Energía

Partiendo del principio de la conservación de la energía de Rayleigh se estima el aporte de energía de cada uno de los grupos de coeficientes a la energía total de la trama, por tanto, los grupos de coeficientes con más energía tienen asociado un CRR más alto y, por consiguiente, un mayor número de niveles de cuantificación.

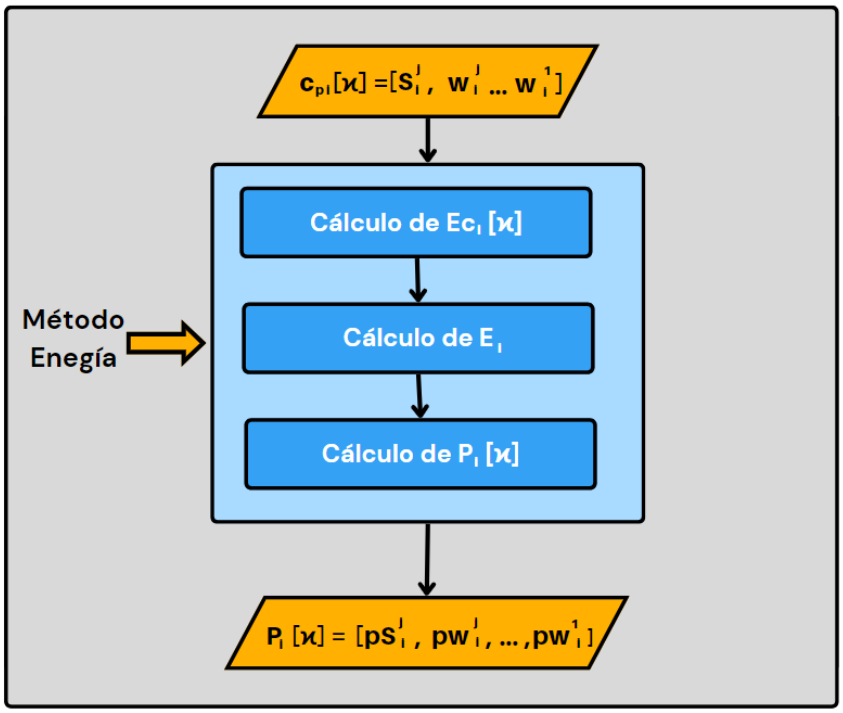


Figura 3.17: Método CRR de energía.

En la Figura 3.17 se observa el diagrama de flujo de este método, comenzando por el cálculo de energías de cada uno de los coeficientes de las tramas de la señal. Cabe resaltar que $c_{pi}[k]$ almacena los valores de las muestras de cada coeficiente en la posición κ, y $Ec_i[\kappa]$, es un arreglo que contiene la energía de cada grupo de coeficientes.

$$c_{pi}[k] = c_i[\kappa], \tag{3.12}$$

$$Ec_i[\kappa] = \sum_k c_{pi}[k]^2. \tag{3.13}$$

Adicionalmente, se calcula la energía total de cada trama en E_i.

$$E_i = \sum_\kappa Ec_i[\kappa]. \tag{3.14}$$

Para finalmente obtener los CRR, $P_i[\kappa]$, que corresponde al porcentaje que representa $Ec_i[\kappa]$, con respecto a E_i.

$$P_i[\kappa] = \frac{Ec_i[\kappa]}{E_i} \tag{3.15}$$

Método CRR de Percepción

Este enfoque se centra en estimar el aporte de cada grupo de coeficientes en la calidad de la señal de voz resultante, para lo cual de forma sistemática se eliminan uno a uno los grupos de coeficientes de la trama y a partir de los grupos restantes se realiza la transformada *Wavelet* inversa, obteniendo una señal en el dominio del tiempo, la cual se compara (NMSE) con la señal original. De esta forma se evalúa la importancia de cada uno de los coeficientes en la calidad de la señal.

En la Figura 3.18 se ejemplifica el funcionamiento de este método CRR con el fin de ilustrar su lógica. A la señal original, de color azul, se le aplica la transformada *Wavelet* y, posteriormente, se anula uno de los grupos de coeficientes, en este caso $W_x{}^2$, el cual se encuentra representado en color gris. Usando los coeficientes restantes se realiza la transformada inversa para obtener una señal en el dominio del tiempo, mostrada en color naranja, para finalmente calcular la calidad de la señal procesada (señal naranja) con respecto a la original (señal azul) y de esta manera estimar la relevancia del grupo de coeficientes eliminado.

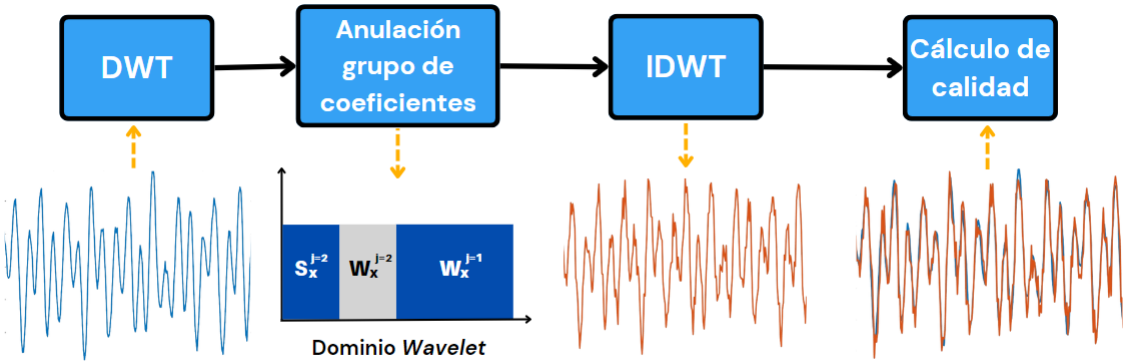


Figura 3.18: Funcionamiento método CRR de percepción.

En la Figura 3.19 se muestra el funcionamiento de este método a nivel técnico, el cual tiene como entrada el vector de grupos de coeficientes $c_i\,[\kappa]$.

Se inicializa la variable $\kappa = 1$, para comenzar el bucle, que se repite mientras se cumple la condición de que κ no exceda el número de grupos de coeficientes. Si la condición se cumple, se hace una copia del vector de grupos de coeficientes c_i en la c_{Ci}, la cual va a ir almacenando los grupos de coeficientes originales a excepción del grupo de coeficientes en la posición κ, al cual se le asignan ceros en todas sus muestras, esto es,

$$c_{Ci}\left[\kappa\right] = 0.$$

Seguidamente se hace la transformada inversa a partir de los coeficientes almacenados en $c_{C\,i}$ y la señal del tiempo resultante se guarda en la variable $x_R[n]$. Además, se calcula el NMSE entre la $x_R[n]$ y la señal original muestreada $x[n]$, con el propósito de analizar la distorsión que introduce anular el grupo de coeficientes evaluado. Cabe resaltar que en el NMSE un valor muy cercano a 1 implica que el coeficiente eliminado no es muy relevante para la calidad de la trama, por eso se determina que la relevancia se calcula como el complemento del NMSE y se almacena en la variable $rate$.

$$\mathsf{rate}\left[\kappa\right] \; = 1 - NMSE\left\{x_R\left[n\right],\; x\left[n\right]\right\}. \tag{3.16}$$

A continuación, se aumenta el valor de κ y se vuelve a evaluar la condición del bucle. En caso de que κ sea mayor al número de grupos de coeficientes, se cierra el bucle. Y, finalmente, se normalizan los valores del vector rate para sacar los valores definitivos de los CRR y se asignan a $P_i\left[\kappa\right]$.

$$P_i\left[\kappa\right] = \frac{\mathsf{rate}}{\sum \mathsf{rate}}. \tag{3.17}$$

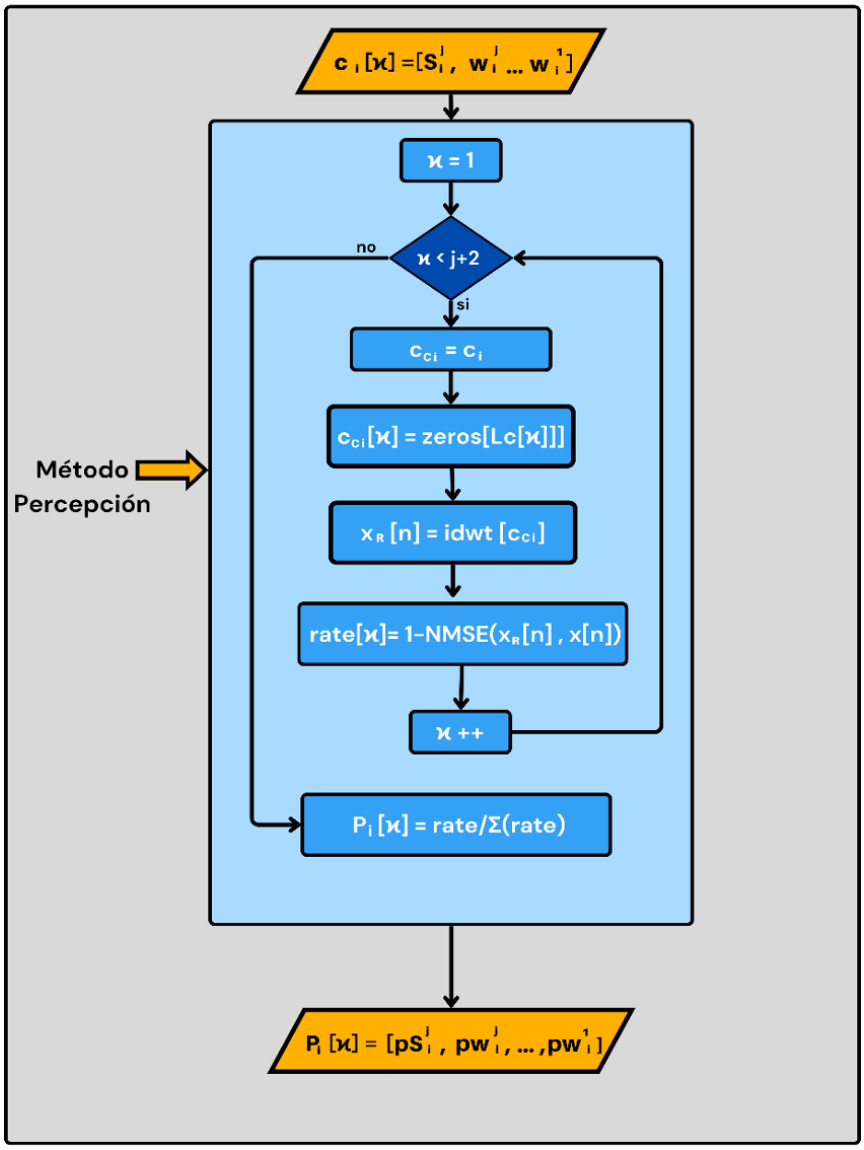


Figura 3.19: Método CRR de percepción.

Método CRR Heurístico

Este método calcula los porcentajes de relevancia para cada trama por medio de un proceso heurístico. Comenzando por una asignación de 0 bits a cada grupo de coeficientes, se empieza a variar de forma sistemática la cantidad de bits que se asignan a cada uno y se calcula el NMSE en el tiempo para validar si las variaciones están siendo fructíferas. De esta forma, continúa variando estos valores hasta no encontrar una mejor distribución y, es entonces, cuando finaliza calculando el mejor vector de porcentajes encontrado $P_{Fi}\,[\kappa]$.

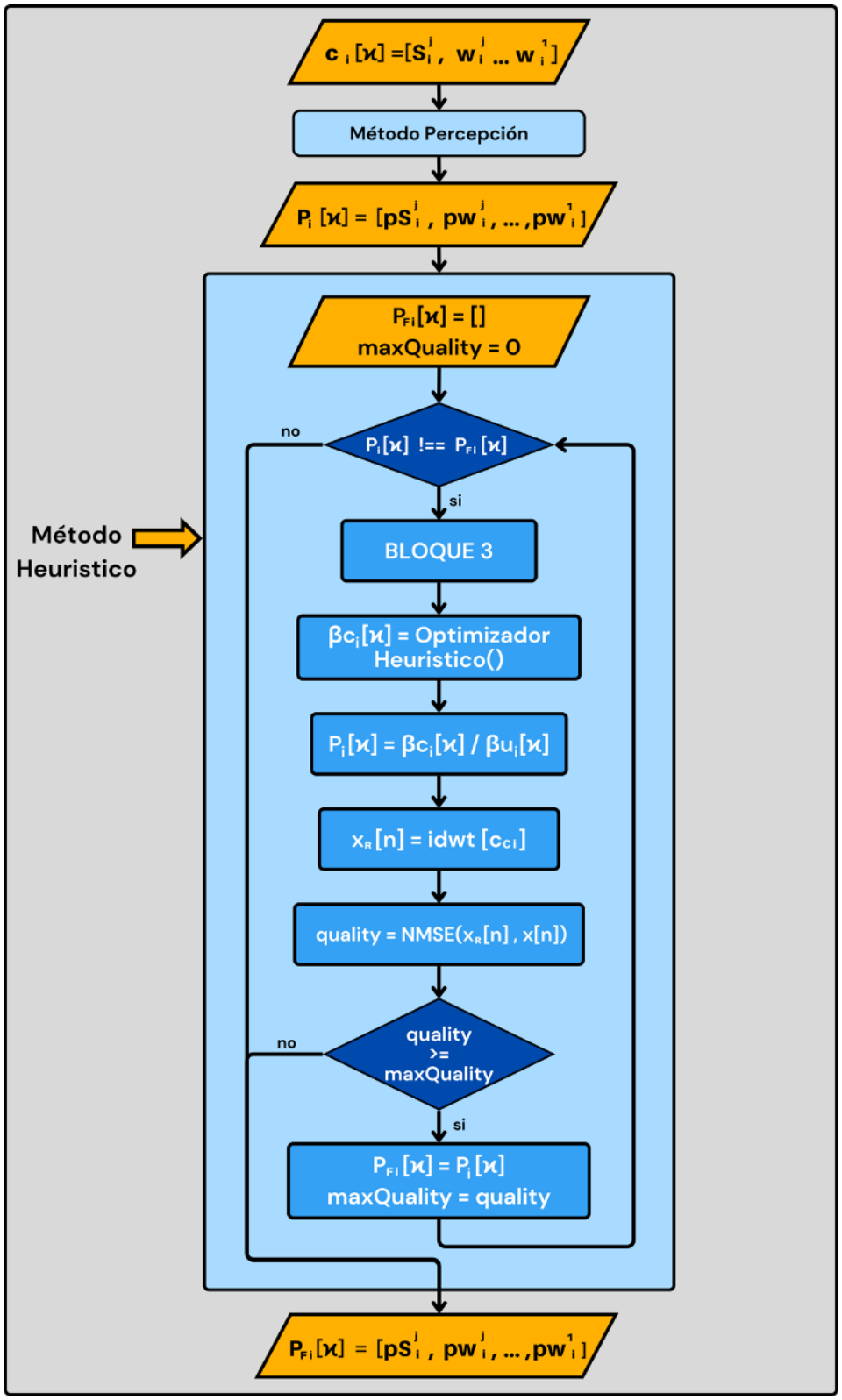


Figura 3.20: Método CRR heurístico.

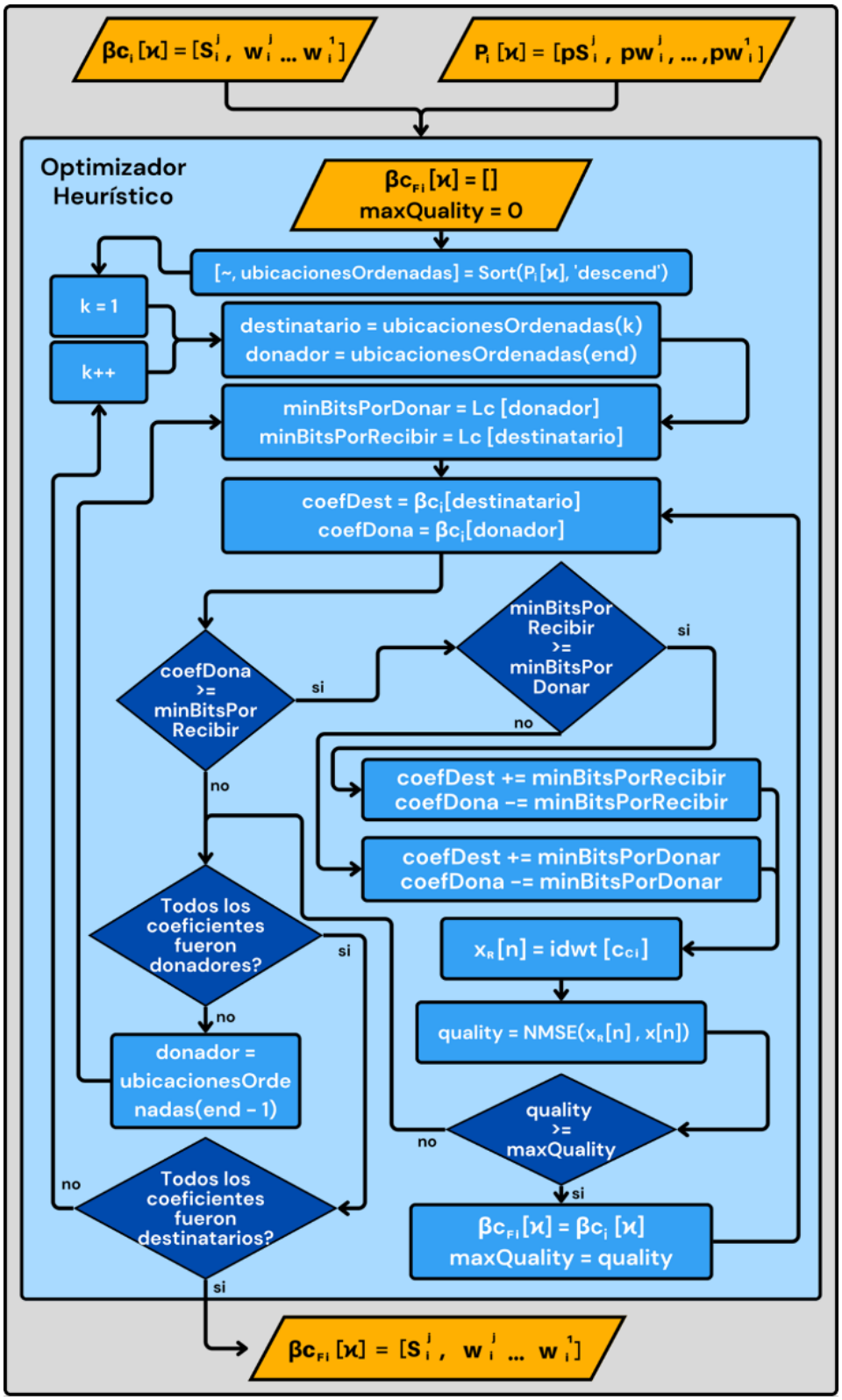


Figura 3.21: Optimizador heurístico.

En la Figura 3.20 se muestra con más detalle el funcionamiento de este método, el cual de forma inicial usa la misma lógica del método CRR de percepción, pues requiere un vector de porcentajes iniciales para identificar la relevancia de los coeficientes.

Posterior a ello, se crean dos variables que servirán para llevar registro de los valores encontrados, esto es: los mejores porcentajes, $P_{Fi}\left[\kappa\right]$, y la mejor calidad obtenida, $maxQuality$. Después, siempre que los porcentajes sean actualizados, se realiza la repartición de los bits con los mejores porcentajes encontrados hasta el momento y se optimiza la distribución con la función llamada $Optimizador\ Heuristico$, la cual se desglosa en la Figura 3.21.

Finalmente, se calculan los porcentajes de relevancia basándose en la mejor distribución de bits encontrada hasta el momento, y se computa el nuevo valor de calidad. Si este nuevo valor de calidad es mayor al que se encuentra almacenado en la variable $maxQuality$, entonces se actualizan los mejores porcentajes $P_{Fi}\left[\kappa\right]$ y la mejor calidad obtenida $maxQuality$, para empezar nuevamente el ciclo. Estas iteraciones solo terminan cuando no se encuentran mejores porcentajes a los que se tienen, o cuando la mejor calidad obtenida es menor o igual a la almacenada previamente.

En la Figura 3.21 se muestra a detalle el funcionamiento del $OptimizadorHeuristico$, el cual recibe como parámetros de entrada el vector de bits asignados por grupo de coeficientes, $\beta c_i\left[\kappa\right]$, y el vector con los porcentajes de relevancia de los coeficientes $P_i\left[\kappa\right]$.

Inicialmente, se crean dos variables que servirán para llevar registro del mejor vector de bits asignados por grupo de coeficientes, $\beta c_{Fi}\left[\kappa\right]$, y la mejor calidad obtenida, $maxQuality$. Luego de definir un orden de prioridades basado en los porcentajes de relevancia, se definen los coeficientes que actuarán como donador y destinatario, y la cantidad de bits mínimos que están dispuestos a donar y/o recibir respectivamente. Este último parámetro es de gran importancia, pues los bits deben de ser donados de tal forma que al final la cantidad de bits asignados a cada trama continúe siendo múltiplo del tamaño de muestras que posee cada grupo de coeficientes.

Tanto el donador como el destinatario deben tener siempre un número entero de bits por muestra, por lo que a continuación se ejemplifican los dos posibles escenarios en los cuales se deben hacer validaciones para cumplir este axioma. Así, si el grupo de coeficientes donador posee 128 muestras y el destinatario 32, el donador debe ceder como mínimo 128 bits (o un número múltiplo de 128), pues de no ser así resultaría con un número de bits que no se puede repartir uniformemente entre todas sus muestras. Si, por el contrario, el coeficiente donador posee

32 muestras y el destinatario 128, el donador debe ceder como mínimo 128 muestras para que el destinatario resulte con un número de bits que se pueda repartir de forma uniforme.

Tras hacer estos cambios en la distribución de bits asignados por coeficiente, solo resta validar si el cambio aportó o no a mejorar la calidad de la trama. De ser así, se actualizan el mejor vector de bits asignados por grupo de coeficientes, $\beta c_{Fi}[\kappa]$, y la mejor calidad obtenida $maxQuality$, para empezar nuevamente el ciclo; si no, se actualiza al siguiente coeficiente donador y se repite nuevamente todo el proceso. Esto se repite hasta que todos los coeficientes hayan sido donadores y destinatarios de bits.

3.3.2. Bits de Reserva

Se entienden como bits de "colchón", i.e., el valor mínimo que tiene cada grupo de coeficientes, los cuales juegan un papel esencial en el proceso de cuantificación. Estos bits se asignan inicialmente para garantizar que cada muestra de la señal cuantificada tenga un número mínimo de bits disponibles para representar su valor. De esta manera, se preserva la calidad perceptual de la señal resultante, evitando una degradación significativa por la falta de bits en algunas muestras. Además, la asignación inicial de bits de reserva permite que la cantidad de recursos que se deben distribuir por medio de los CRR sea menor, lo cual simplifica el proceso y se traduce en un menor costo de procesamiento durante la asignación.

Para analizar cuál es el número de bits de reserva más conveniente para el diseño del cuantificador, se realizan pruebas con los tres métodos de CRR propuestos, variando el número de bits de reserva asignados para ver cómo esto repercute en la calidad de la señal resultante.

En la Figura 3.22 se representa por medio de una gráfica de calor la calidad de la señal resultante en función de los bits de reserva que se asignan dependiendo del número de bits disponibles por muestra, para el método CRR de percepción. Teniendo en cuenta que los valores con mejor calidad son los que más se acercan al color amarillo y los que mayor distorsión tienen los que se acercan al color azul oscuro, se puede afirmar que para todas las variaciones de M se obtienen mayores valores de calidad para los bits de reserva correspondientes a $\log_2(M) - 1$. Es decir que se tienen mejores resultados cuando se asignan como bits de reserva el máximo de bits disponibles por muestra menos uno, esto es: $m - 1$.

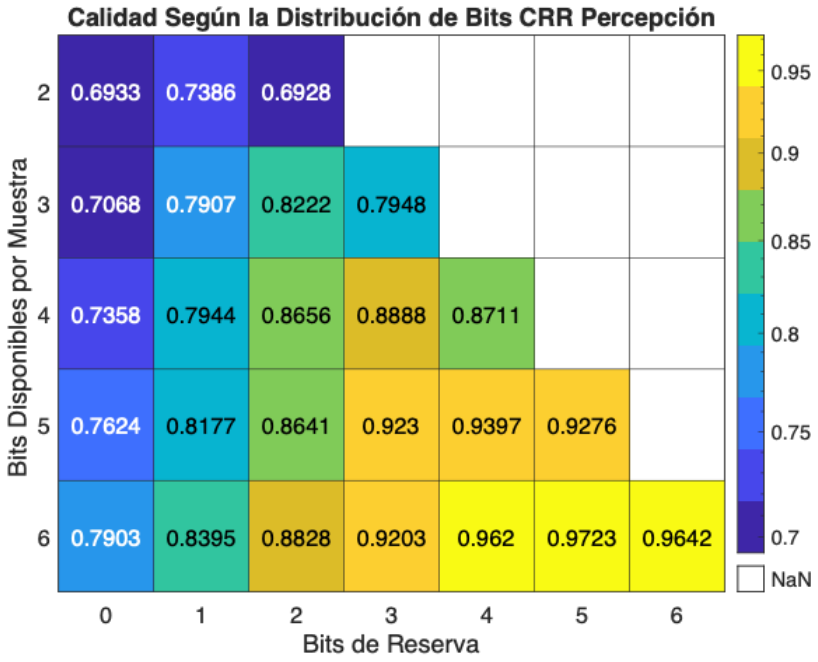


Figura 3.22: Calidad de la señal según bits reserva para método de percepción.

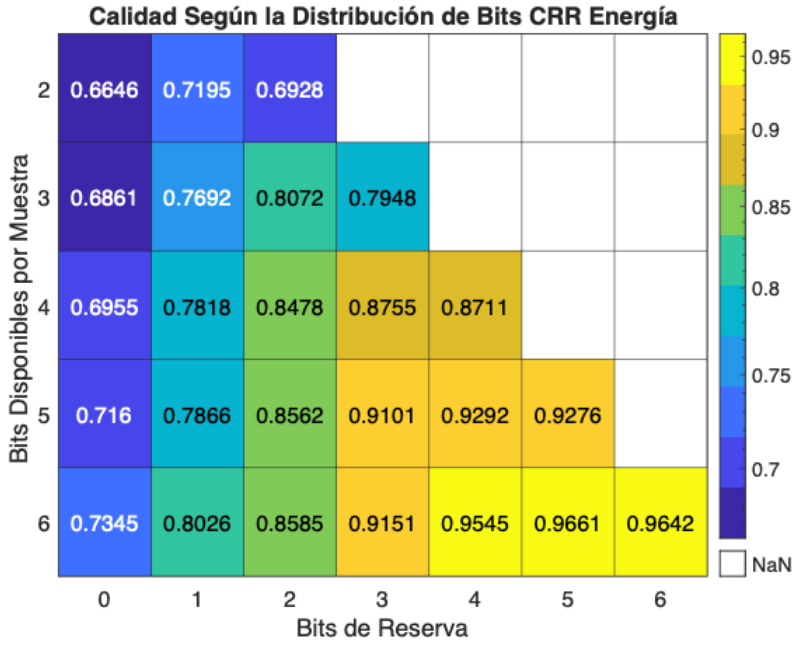


Figura 3.23: Calidad de la señal según bits reserva para método de energía.

Para el método CRR de energía se obtiene el mismo comportamiento que para el método CRR de percepción. Como se observa en la Figura 3.23, para este caso en particular, la mejor opción para el número de bits de reserva es $\log_2(M) - 1$.

En cuanto al método CRR heurístico, su comportamiento difiere un poco en cuanto a los dos métodos nombrados anteriormente. Como se puede observar en la Figura 3.24, a excepción de cuando se tienen 2 bits disponibles por muestra, se puede afirmar que el mejor resultado de calidad es obtenido cuando no se asigna ningún bit de reserva, seguido nuevamente por $\log_2(M) - 1$. No obstante, es importante resaltar que entre estas dos mejores configuraciones la diferencia en los valores de calidad es de aproximadamente un $1\,\%$, lo cual es despreciable. Por el contrario, no asignar bits de reserva o asignar el valor de $\log_2(M) - 1$ sí tiene repercusiones significativas en el costo de procesamiento del bloque 3, debido a que aumenta el número de iteraciones necesarias para alcanzar un resultado.

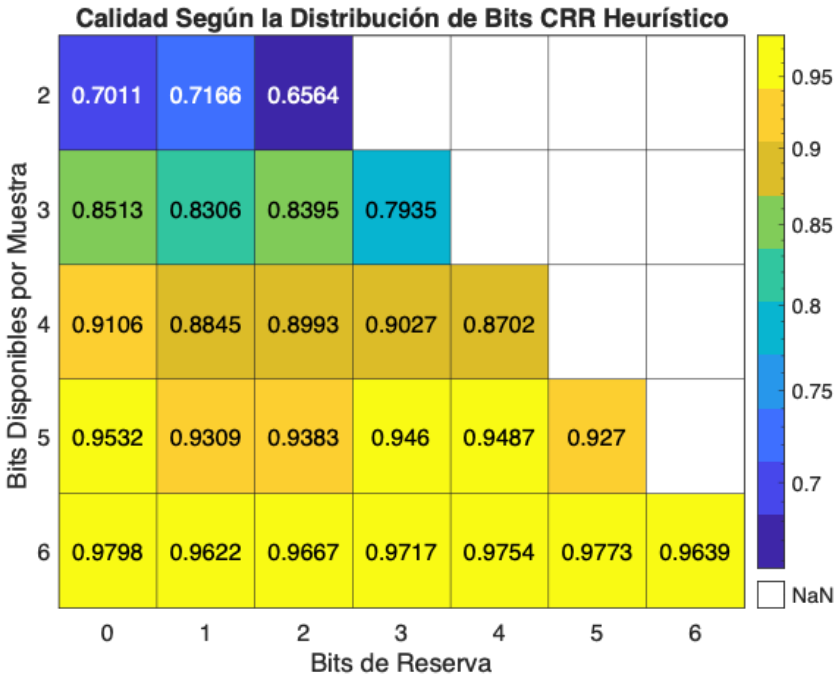


Figura 3.24: Calidad de la señal según bits reserva para método heurístico.

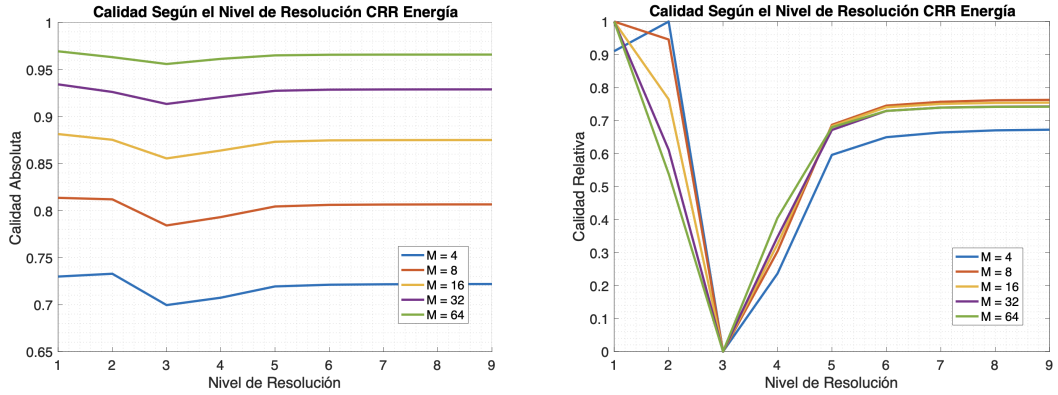
Por tanto, se llega a la conclusión de que no se justifica aumentar considerablemente el costo de procesamiento escogiendo la opción en donde no se asignan bits de reserva, sabiendo que el valor de $\log_2(M) - 1$ para los bits de reserva presenta valores muy favorables desde 2 bits disponibles por muestra.

3.3.3. Número de Niveles Resolución

Este parámetro influye directamente en la precisión y el detalle de la representación de la señal. El número de niveles de resolución, j, genera $j+1$ grupos de coeficientes diferentes, por lo que el valor de j es directamente proporcional con la complejidad del diseño del proceso de cuantificación en el dominio *Wavelet*. En general el número de niveles de resolución determina el grado de precisión con el que se puede "ver" el comportamiento de la señal, así, un valor elevado de j permite capturar detalles más finos de la señal.

Para determinar cuál es el nivel de resolución que mejor se ajusta al Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Esquema *Lifting*, se realizan pruebas con los tres métodos de CRR propuestos y se varía el nivel de resolución de 1 a 9. Cabe resaltar que para los tres métodos se ajusta el valor de bits de reserva, r, en $\log_2(M) - 1$. Adicionalmente, se consideran diferentes valores de niveles de cuantificación, esto es $M \in \{4,\ 8,\ 16,\ 32,\ 64\}$.

Los resultados de calidad obtenidos en cada uno de los tres métodos para cada uno de los valores de niveles de cuantificación evaluados muestran variaciones muy pequeñas con respecto al cambio en el número de niveles de resolución, por lo que para presentar estos resultados se muestra: a) la gráfica original de los valores de calidad, b) la gráfica de los valores de calidad normalizados para cada nivel de cuantificación. Adicionalmente, se construye una tabla con los valores de calidad máximo y mínimo, así como su desviación estándar.



(a) (b)

Figura 3.25: Calidad según el nivel de resolución para método CRR de energía (a) absoluta (b) relativa.

En la Figura 3.25 se muestran los resultados obtenidos con el CRR de energía. Al

analizar cuál es el mejor nivel de resolución para cada uno de los niveles de cuantificación (Figura 3.25b) se evidencia que para un $M = 4$ se obtiene el mejor valor de calidad con dos niveles de resolución, mientras que para los demás valores de M se obtienen mejores resultados de calidad con un nivel de resolución. Por otro lado, los resultados absolutos (Figura 3.25a) muestran que las variaciones en los valores de calidad disminuyen conforme aumenta el valor de M, como se puede corroborar a partir de la desviación estándar obtenida en cada caso, las cuales se encuentran en la Tabla 3.3.

Tabla 3.3: Valores de calidad y desviación estándar de método CRR de energía.

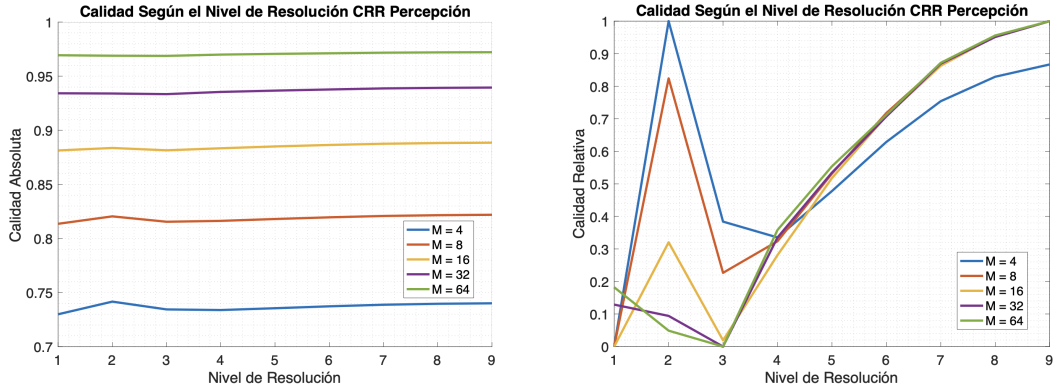
Nivel de Cuantificación	**Calidad Máxima**	**Calidad Mínima**	**Desviación Estándar**
4	0.7328	0.6995	0.0103
8	0.8135	0.7842	0.0093
16	0.8814	0.8555	0.0077
32	0.9342	0.9134	0.0060
64	0.9694	0.9559	0.0038

A partir de los resultados obtenidos se infiere que la elección del número de niveles de resolución no es relevante cuando se tiene un número elevado de niveles de cuantificación ($M = 32,\ 64$), puesto que se tienen suficientes bits por muestra para representar los valores de los coeficientes sin importar la forma en que éstos se agrupen. Por el contrario, en los casos en los que el número de niveles de cuantificación es pequeño ($M = 4$) es evidente que lo más recomendable es tener pocos niveles de resolución y que en estos casos una mala elección en este parámetro puede repercutir negativamente en la calidad de la señal de voz cuantificada.

Finalmente, se resalta que las curvas de calidad relativa no tienen un comportamiento lineal con respecto al nivel de resolución; por el contrario, todas las curvas tienen su valor mínimo en $j = 3$. Este comportamiento se explica dado que entre mayor sea el número de niveles de resolución, menor es el número de coeficientes en los grupos de más bajo orden (componentes de baja frecuencia), siendo estos grupos los que mayor energía tienen. De este modo, para que el método CRR de energía realice una distribución de recursos adecuada, es conveniente que se tengan pocos niveles de resolución con longitudes y porcentajes de energía cercanos o, por el contrario, que se tenga un número elevado de niveles de resolución con mayor contraste entre las longitudes y los porcentajes de energía de los grupos de coeficientes.

La Figura 3.26b, muestra la calidad relativa para el método CRR de percepción.

Al realizar el análisis de la calidad según la variación de M, se evidencia que para $M = 4$ se obtiene una mejor calidad para dos niveles de resolución, mientras que para los demás valores de M se obtiene la mejor calidad con nueve niveles de resolución; sin embargo, al analizar los valores de calidad absolutos (Figura 3.26a), se observa que para niveles de M elevados, la calidad no presenta grandes variaciones con respecto al nivel de resolución, lo que se puede ratificar con los bajos valores de desviación estándar consignados en la Tabla 3.4; esto quiere decir que, al igual que para el método de CRR de energía, entre más bajo sea M, la escogencia del nivel de resolución tiene mayor impacto en la calidad de la señal.



(a) (b)

Figura 3.26: Calidad según el nivel de resolución para método CRR de percepción (a) absoluta (b) relativa.

Tabla 3.4: Valores de calidad y desviación estándar de método CRR de percepción.

Nivel de Cuantificación	Calidad Máxima	Calidad Mínima	Desviación Estándar
4	0.7416	0.7299	0.0036
8	0.8219	0.8135	0.0029
16	0.8885	0.8813	0.0027
32	0.9394	0.9334	0.0023
64	0.9721	0.9687	0.0013

Por último, se puede observar que para este método no se evidencia un patrón estable ante la variación de los niveles de resolución. Para valores de $j \geq 4$ en

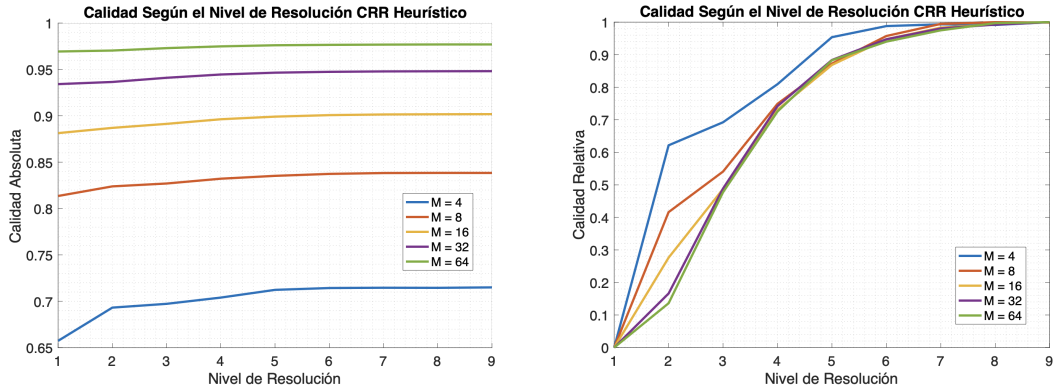
todos los casos se observa que la calidad aumenta conforme se incrementa el nivel de resolución; no obstante, para valores de $j < 4$ se tienen dos tipos de comportamientos según el valor de M. Para $M = \{32,\ 64\}$ se tiene que la calidad decae ligeramente al pasar de $j = 1$ a $j = 2$, y que a partir de $j = 3$ tiene un sutil incremento. Para $M = \{4,\ 8,\ 16\}$ idealmente se tendría el decaimiento para valores de $j < 4$; sin embargo, en estos 3 casos se tiene un pico en $j = 2$ cuyo valor es comparable con el obtenido con $j = 9$.

Es importante resaltar que la elección del número de niveles de resolución implica que se van a crear $j + 1$ grupos de coeficientes, los cuales abarcan diferentes bandas de frecuencia. De esta manera, existe la posibilidad de que algunas de estas divisiones del espectro coincidan con bandas de frecuencia importantes para garantizar la calidad de señales de voz según el sistema auditivo humano, mientras que en otros casos se agrupen bandas de frecuencia cruciales y bandas de frecuencia intrascendentes, por lo que se podría llegar a una distribución de recursos ineficiente.

Así, para un nivel de resolución, al tener sólo dos grupos de coeficientes de la misma longitud, no se hace uso de todo el potencial del método de asignación de CRR, y se distribuyen de manera ineficiente los bits, por otro lado para dos niveles de resolución se evidencia una mejora de la calidad puesto que se tiene mayor detalle con los grupos de coeficientes, lo que permite distribuir con mayor precisión los bits según su relevancia en la percepción, mejorando así la calidad; sin embargo, al aumentar a tres niveles de resolución, se evidencia una desmejora en la calidad, esto puede darse debido a que las frecuencias asignadas a cada grupo de coeficientes presentan discrepancias con respecto a la homogeneidad de la relevancia de sus muestras, es decir que los grupos de coeficientes abarcan bandas de frecuencias que tienen diferentes impactos en la señal, generando problemas con la caracterización de las muestras incluidas en cada grupo de coeficientes, lo que a su vez disminuye la calidad.

En cuanto al método CRR heurístico, se puede observar en la Figura 3.27b, que todas las curvas de calidad relativa según el nivel de resolución tienen un comportamiento creciente asintótico, lo cual permite inferir que la mejor calidad se presenta para nueve niveles de resolución. Esto se adjudica a la complejidad del método heurístico para encontrar una asignación de bits adecuada para cada grupo de coeficientes, que brinde una calidad “inmejorable” dentro de las posibilidades consideradas; esto permite que el nivel de resolución sea directamente proporcional a la calidad, ya que a mayor resolución se puede hacer una mejor caracterización, considerando que el método heurístico remueve bits de grupos de coeficientes que no los requieren y se los asigna a grupos de coeficientes en los que impactan en mayor magnitud en la calidad.

Analizando la Figura 3.27a, se evidencia que la calidad absoluta no presenta variaciones significativas en la calidad de la señal con respecto al número de niveles de resolución, a excepción de la curva $M = 4$, la cual presenta una variación de calidad más notoria cuando se pasa de $j = 1$ a $j = 2$.



(a) (b)

Figura 3.27: Calidad según el nivel de resolución para método CRR de heurístico (a) absoluta (b) relativa.

En la Tabla 3.5, se puede observar que, efectivamente, los valores de desviación estándar para todas las curvas son bastante bajos y el que mayor variabilidad tiene es el cuantificador con $M = 4$, debido a las variaciones notorias entre el primer y segundo nivel de resolución.

Tabla 3.5: Valores de calidad y desviación estándar de método CRR de heurístico.

Nivel de Cuantificación	Calidad Máxima	Calidad Mínima	Desviación Estándar
4	0.7150	0.6574	0.0188
8	0.8384	0.8136	0.0086
16	0.9018	0.8814	0.0075
32	0.9482	0.9342	0.0053
64	0.9770	0.9694	0.0029

Después de analizar los resultados de calidad según el nivel de resolución se identifica que para los tres métodos propuestos, tienen algunas convergencias en el comportamiento: a medida que aumenta M la desviación estándar disminuye, lo que indica que las puntuaciones de calidad tienden a agruparse cerca de la

media y esto sugiere que el cuantificador es menos sensible a las variaciones en el número de niveles de resolución, lo que puede interpretarse como una mayor estabilidad en la calidad percibida, dado que hay un mayor número de recursos disponibles, independientemente de las agrupaciones disponibles, cada muestra tiene un número adecuado de bits. Por esta misma razón se precisa que, independientemente del método CRR, para cuantificadores con un bajo número de niveles de cuantificación la elección del nivel de resolución es crítica si se desea obtener una mejor calidad en la señal cuantificada. En consecuencia, para la escogencia del número de niveles de resolución, debe primar el impacto en este parámetro en bajos niveles de cuantificación, por lo cual, se determina qué:

- Para niveles de cuantificación bajos el método de CRR de percepción, brinda mejores resultados de calidad con dos niveles de resolución. Por otro lado, aunque para niveles de cuantificación altos, el método de CRR de percepción presenta una mejoría al aumentar el nivel de resolución, el aumento en calidad conseguido con nueve niveles de resolución no se considera significativo. Debido a que esta mejoría en calidad no tiene trascendencia, pero el costo de procesamiento sí aumenta al procesar tantos niveles de resolución, se sugiere seguir usando dos niveles de resolución.
- Para el método de CRR de energía, se recomienda usar un nivel de descomposición para todas las variaciones de M, ya que brinda el mejor resultado para la mayoría de niveles de cuantificación y, adicionalmente, es la opción que presenta menor costo computacional.
- Para el método de CRR heurístico, a pesar de que tiene un comportamiento no lineal creciente, los niveles de resolución a partir de dos no generan incrementos significativos en la calidad, por lo que se recomienda trabajar con dos niveles de resolución para cualquier M.

En la Tabla 3.6 se muestra el resumen de las recomendaciones realizadas a partir del estudio de la calidad con respecto al nivel de resolución.

Tabla 3.6: Recomendaciones sobre j según el método de CRR.

Método CRR	**Nivel de Resolución**
Percepción	2
Energía	1
Heurístico	2

Finalmente, se comparan los tres métodos CRR propuestos para los diferentes niveles de cuantificación, con los parámetros recomendados previamente, se miden los valores del NMSE y la PESQ en porcentajes a partir de la señal cuantificada resultante, los cuales, posteriormente, son promediados. Los resultados de la Figura 3.28 evidencian que el método de percepción y el de energías tienen un comportamiento muy similar, mientras que el método heurístico tiene una desventaja notoria en niveles de cuantificación bajos. No obstante, a partir de ocho niveles de cuantificación el método heurístico obtiene una sutil ventaja frente al método de percepción.

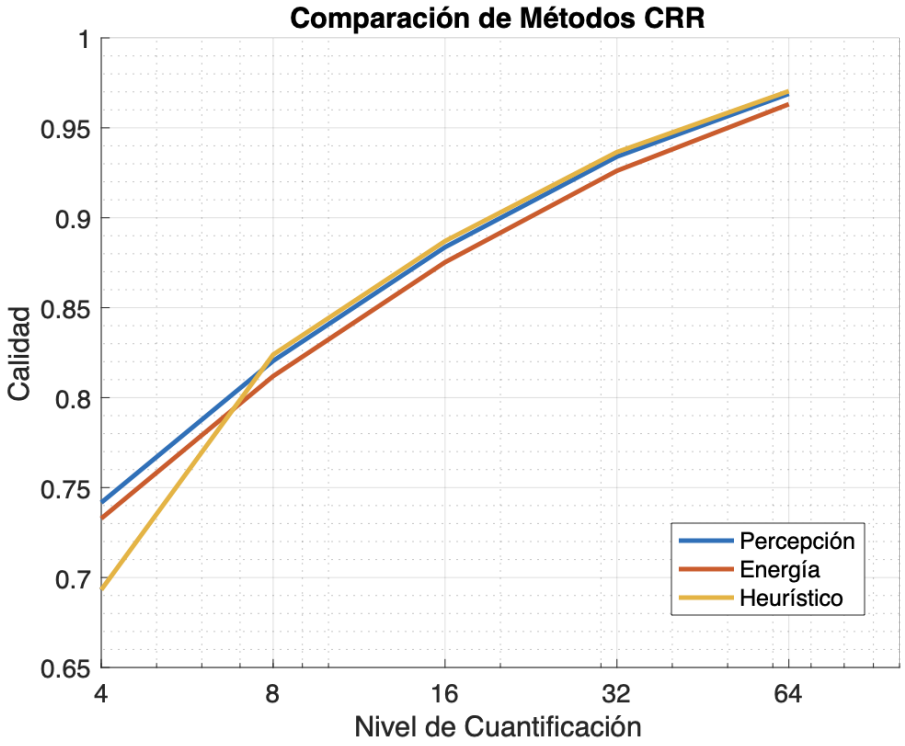


Figura 3.28: Comparación de métodos CRR.

Teniendo en cuenta que, este trabajo de grado busca tener buenos resultados aún con bajos niveles de cuantificación, se decide que el método que mayor beneficio presenta, al mantener su comportamiento y, por lo tanto, tener buenos resultados de calidad para bajos niveles de cuantificación, es el método de percepción.

Una vez escogido el método de percepción, se analiza la relevancia de las bandas de frecuencias abarcadas por cada coeficiente en el escenario escogido, es decir, para dos niveles de resolución. En la Figura 3.29 se muestran las bandas de

frecuencia dadas para cada uno de los grupos de coeficientes [10]. Adicionalmente, en la Tabla 3.7 se plasman los porcentajes de relevancia de dichas bandas, obtenidos a partir del promedio de los porcentajes dados por los 300 audios del repositorio y todas las familias *Wavelet* consideradas.

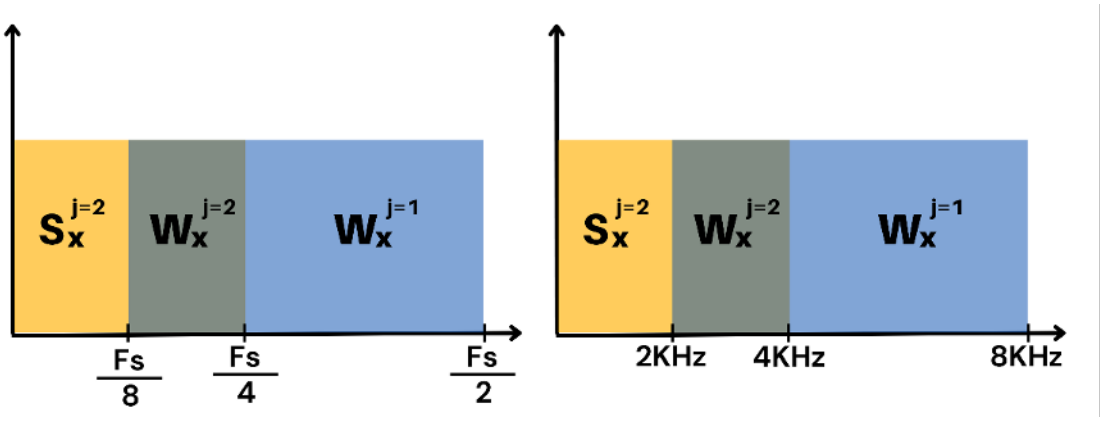


Figura 3.29: Bandas de frecuencia para coeficientes con dos niveles de resolución y $Fs = 16$ KHz.

Tabla 3.7: CRR para bandas de frecuencia para coeficientes con dos niveles de resolución y $Fs = 16$ KHz.

Coeficiente	Banda de frecuencia	CRR Percepción
${S_x}^{j=2}$	0KHz - 2KHz	0.856122
${W_x}^{j=2}$	2KHz - 4KHz	0.101011
${W_x}^{j=1}$	4KHz - 8KHz	0.042866

Los datos presentados en la Tabla 3.7 permiten inferir que las frecuencias en el rango de 0 KHz a 2 KHz desempeñan un papel significativo en la percepción auditiva humana, con un valor de CRR de 0.856. En contraste, las frecuencias entre 2 KHz y 4 KHz muestran una relevancia mucho menor, registrando un CRR de 0.101. Por último, la banda de frecuencias de 4 KHz a 8 KHz exhibe una relevancia inferior al $5\,\%$. Eso concuerda con el hecho de que algunos estándares

[10] Las divisiones del espectro que se muestran en la figura 3.29 son ideales, ya que las *Wavelets* al ser de duración finita en el tiempo no pueden tener una duración finita en el dominio de la frecuencia, como lo son las funciones rectangulares utilizadas en este caso. Analizando a las *Wavelets* como filtros pasa banda, se tiene que las bandas de transición de su espectro de magnitud dependen del tipo de familia y de los momentos de desvanecimiento de éstas; no obstante, de forma aproximada se muestran estas divisiones para evidenciar las bandas de frecuencia en las que más resuenan los diferentes grupos de coeficientes.

asumen el ancho de banda de la voz cómo 4 KHz [44], pero de acuerdo con lo explicado anteriormente con respecto a los bits de reserva, es fundamental mantener una representación mínima de todas las muestras de la señal de voz para evitar la introducción de distorsiones audibles, por tanto ese $5\,\%$ de relevancia de la banda de 4 KHz a 8 KHz contribuye a la inteligibilidad de la señal.

En síntesis, para este caso en particular de dos niveles de resolución, las bandas de frecuencias más bajas son más relevantes para el oído humano por tanto las distorsiones inducidas a las muestras con frecuencias en estas bandas, generan mayor distorsión en la señal percibida por el oyente.

3.4. Señalización

Es fundamental abordar el tema de la señalización, dado que es un aspecto importante en el proceso de transmisión y recepción de las señales de voz. En el ámbito de las telecomunicaciones, el *payload* se refiere a la información útil o los datos que se transmiten de una fuente a un destino específico. Es decir, representa el contenido real de la señal que se desea transmitir, en este caso, las señales de voz cuantificadas mediante el cuantificador de señales de voz en el dominio *Wavelet* utilizando el esquema *Lifting*.

La señalización, por otro lado, se refiere a los datos adicionales y la información de control que se agregan al *payload* para permitir su identificación, recuperación, enrutamiento, entre otros. Estos datos de señalización son necesarios para asegurar la correcta entrega y recepción de la información en el contexto de un sistema de telecomunicaciones.

Es fundamental tener en cuenta que la inclusión de la señalización aumenta el costo de envío, ya que introduce una sobrecarga en la transmisión. Por lo tanto, es necesario considerar de manera cuidadosa cómo se implementa la señalización en el sistema de comunicaciones para garantizar una transmisión eficiente y confiable de las señales de voz cuantificadas.

En esta sección, se estudia a detalle el costo de la señalización necesaria para que el receptor pueda recuperar la señal de voz transmitida, es decir, lo que compete al contexto del cuantificador de señales de voz en el dominio *Wavelet* utilizando el esquema *Lifting*.

Recordando que cada trama tiene $j+1$ cuantificadores, es decir que el número de cuantificadores es directamente proporcional con el número de coeficientes, implica que cada coeficiente tiene su propio costo de señalización, ς, como se observa en la Figura 3.30.

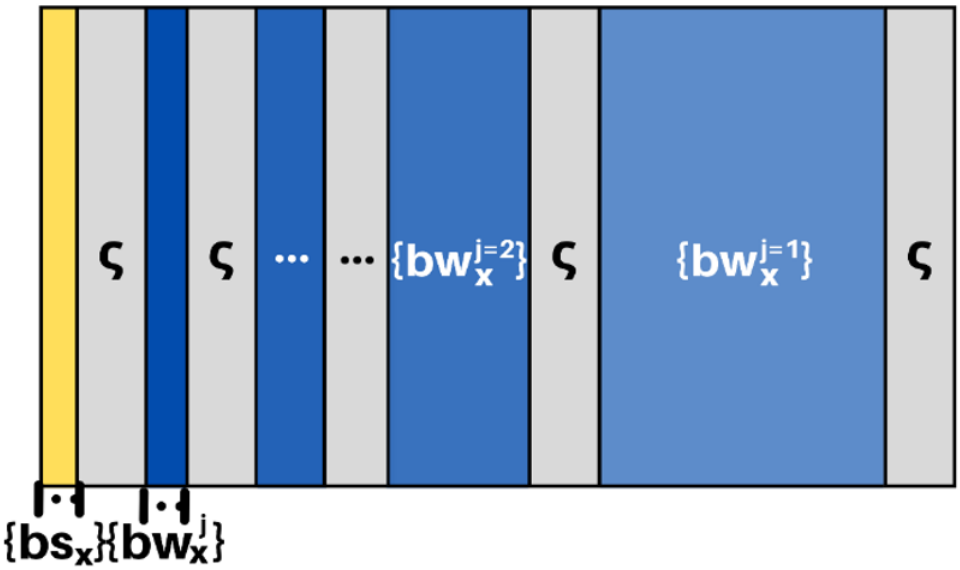


Figura 3.30: Señalización de trama.

El costo de señalización por coeficiente, ς, tiene en cuenta que se deben representar los valores de amplitud mínimo y máximo por grupo de coeficientes y, adicionalmente, el número de niveles de cuantificación. En la ecuación 3.18 se describe matemáticamente cómo se calcula el valor de ς.

$$\varsigma = 2 \cdot (B_{RD} + 1) + 4, \tag{3.18}$$

teniendo en cuenta que B_{RD} es el número de bits dados para representar el valor máximo y mínimo, para este caso se decide que $B_{RD} = 16$ bits, por tanto, se podrán representar hasta 2^{16} valores de máxima y mínima amplitud. Adicionalmente se tiene un bit más para representar la polaridad del número. Por otro lado, se tienen disponibles 4 bits para representar las potencias del nivel de cuantificación M, es decir, con 4 bits, el número decimal más grande que se puede representar es 15, por tanto, se pueden representar hasta 2^{15} valores de M, ya que M solo toma valores potencias de 2.

Con lo dicho anteriormente, se conoce que $\varsigma = 38$ bits.

Finamente, para conocer el valor de señalización total por trama, ς_i, se debe multiplicar el valor de ς por el número de coeficientes:

$$\varsigma_i = \varsigma \cdot (j + 1)\,. \tag{3.19}$$

Así, si se quiere encontrar el valor de señalización por la señal de voz, se debe multiplicar el valor de señalización por trama ς_i, por el número de tramas.

3.5. Plan de Pruebas

Una vez encontradas las variables óptimas en el diseño del cuantificador (método CRR, número de niveles de resolución, j, y bits de reserva, r), se plantea el plan de pruebas con el cual se evalúa cual cuantificador (Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Esquema *Lifting*, Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Algoritmo de Mallat, o Cuantificador de Señales de Voz en el Dominio del Tiempo) brinda mejores resultados de calidad.

3.5.1. Prueba 1

Procesar los 300 audios del repositorio teniendo en cuenta las diferentes familias escogidas, con el Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Esquema *Lifting*, discriminando los niveles de cuantificación.

Propósito: Obtener un valor promedio de calidad por familia en cada nivel de cuantificación para el Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Esquema *Lifting*, después de procesar todas las señales.

3.5.2. Prueba 2

Procesar los 300 audios del repositorio teniendo en cuenta las diferentes familias escogidas, con el Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Algoritmo de Mallat, discriminando los niveles de cuantificación.

Propósito: Obtener un valor promedio de calidad por familia en cada nivel de cuantificación para el Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Algoritmo de Mallat, después de procesar todas las señales.

3.5.3. Prueba 3

Procesar los 300 audios del repositorio, con el Cuantificador de Señales de Voz en el Dominio del tiempo discriminando los niveles de cuantificación.

Propósito: Sacar un valor promedio de calidad para cada nivel de cuantificación para el Cuantificador de Señales de Voz en el Dominio del Tiempo, después de procesar todas las señales.

3.5.4. Prueba 4

Escoger las familias que mejores resultados de calidad brindaron para las pruebas 1 y 2 para el Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Esquema *Lifting* y el Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Algoritmo de Mallat y realizar las pruebas subjetivas, con los tres cuantificadores en cuestión.

Propósito: Evaluar subjetivamente los tres cuantificadores y corroborar que las pruebas subjetivas tienen concordancia con las pruebas objetivas.

3.5.5. Prueba 5

Evaluar objetivamente la variación de la longitud de la trama para los tres cuantificadores implementados.

Propósito: Analizar el impacto de la variación de la longitud de la trama para cuantificar las señales de voz.

3.5.6. Prueba 6

Evaluar subjetivamente la variación de la longitud de la trama para el Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Esquema *Lifting*.

Propósito: Recomendar una longitud de la trama para cuantificar las señales con el Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Esquema *Lifting*.

CAPÍTULO 4

ANÁLISIS DE RESULTADOS

En este capítulo convergen los esfuerzos para evaluar y comparar rigurosamente los tres cuantificadores de señales de voz propuestos.

A través de las pruebas planteadas en el Capítulo 3, se espera obtener un panorama general de los beneficios y limitaciones de cada cuantificador en términos de calidad. Con lo cual se espera identificar, el cuantificador que mejor calidad brinda en función de variables como el número de niveles de cuantificación y la longitud de la trama y, de esta manera, dar una comprensión más profunda sobre su capacidad para preservar la calidad de las señales de voz en el proceso de cuantificación. Adicionalmente, la combinación de enfoques objetivos y subjetivos en la evaluación garantiza una visión completa de su rendimiento en la práctica.

Los resultados de estas pruebas no solo contribuyen a validar el enfoque y la eficacia de los cuantificadores desarrollados, sino que también permiten tomar decisiones informadas en el diseño de cuantificadores de señales de voz. Además, la comparación de las evaluaciones objetivas con las subjetivas proporciona una comprensión holística de cómo los métodos de evaluación se alinean en términos de la percepción humana.

4.1. Variación de Niveles de Cuantificación

4.1.1. Análisis de Pruebas Objetivas

Una vez procesados los audios del repositorio con los tres cuantificadores: el Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Esquema *Lifting*, el Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Algoritmo de Mallat y el Cuantificador de Señales de Voz en el Dominio del Tiempo; en la Figura 4.1 se presentan las curvas de calidad para cada cuantificador, las cuales se basan en el promedio de la calidad de los audios procesados al variar el número de niveles de cuantificación. En cuanto a los cuantificadores basados en *Lifting* y Mallat particularmente, se hace un promedio adicional, correspondiente a los resultados de las seis familias *Wavelet* nombradas en la Tabla 3.1. Para evidenciar el rango de calidad dado por las diferentes familias analizadas, se emplean las barras observadas en cada número de niveles de cuantificación, las cuales indican la mejor y la peor calidad obtenida con las familias *Wavelet* para ese número de niveles de cuantificación en particular.

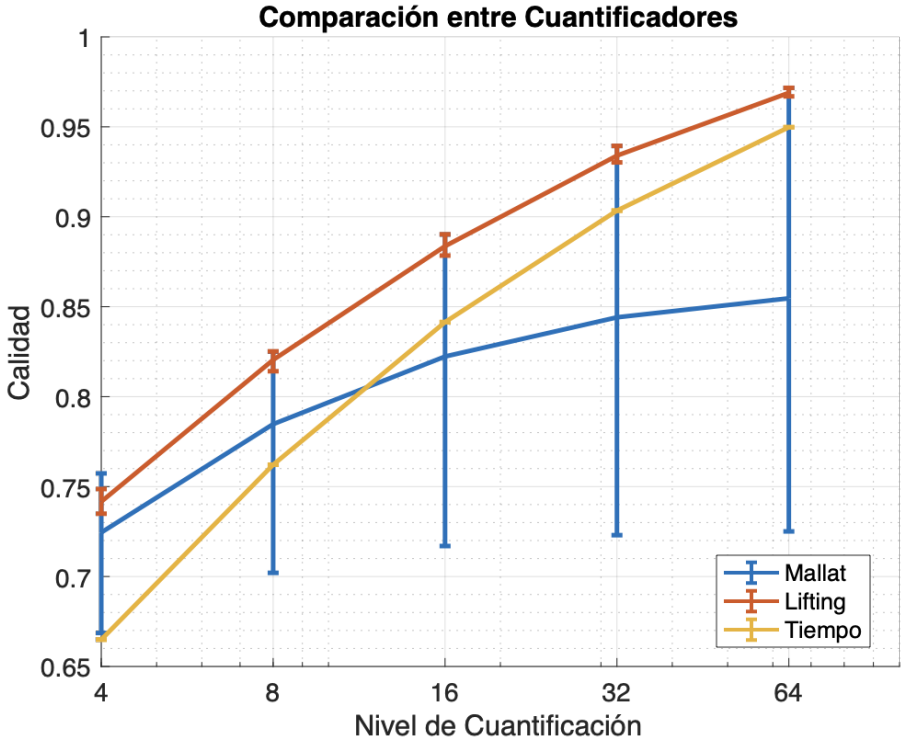


Figura 4.1: Comparación de calidad de los tres cuantificadores evaluados.

Dicho esto, se observa que el rango de valores de calidad obtenido por la variación de las familias *Wavelet* es mucho más amplio para los resultados del cuantificador basado en Mallat, esto se le atribuye al hecho de que la transformada inversa con el algoritmo de Mallat por sí misma no brinda una reconstrucción fidedigna de la señal, ya que, como se explicó anteriormente, entre más similares sean las longitudes del filtro y de la señal, el manejo de la respuesta transitoria de los filtros es más crítico, por tanto, en este caso es contraproducente utilizar familias con filtros de mayor longitud[11].

En el caso del cuantificador basado en *Lifting* se tiene que las diferentes familias *Wavelet* no generan grandes variaciones en la calidad obtenida, como reflejan las pequeñas barras verticales de color rojo en la Figura 4.1; sin embargo, las diferencias que existen, aunque sean pequeñas, se deben a que cada familia *Wavelet* tiene asociadas una función *Wavelet* y una función *Scaling*, por lo que su co-

[11] Con el algoritmo de Mallat sólo se tiene una reconstrucción fiel con filtros de longitud 2 y que, en este trabajo de grado, por efectos de comparación con el esquema *Lifting*, la máxima longitud que se está considerando es filtros con 18 coeficientes. No obstante, el algoritmo de Mallat permite generar un gran número de familias *Wavelet*, las cuales pueden tener asociados filtros de hasta 90 coeficientes. Por esta razón, para el caso particular del procesamiento de señales de voz a través de tramas de corta duración se restringe mucho el número de familias que se podrían utilizar.

rrelación con la señal de voz genera coeficientes con características específicas en cada caso. Una de estas características es la varianza o la dispersión que existe entre los valores de estos coeficientes, por lo que repercute en el rango dinámico de la información que se va a cuantificar, así, dado que M es un valor fijo, el tamaño del escalón aumenta y la precisión de la cuantificación disminuye, generando así, más distorsión en la información cuantificada.

Por los motivos explicados anteriormente la familia *Haar* obtiene mejores resultados, tanto para el cuantificador basado en *Lifting* como para el basado en Mallat, en vista de que es la familia que menor longitud de filtros y menos dispersión tiene en sus coeficientes.

En cuanto al cuantificador en el tiempo, presenta una desventaja notoria frente a los cuantificadores en el dominio transformado, ya que éste se basa en aplicar directamente el proceso de cuantificación uniforme a las muestras de la señal de voz en el dominio del tiempo. Aunque es una técnica sencilla y directa, tiende a enfrentar limitaciones significativas cuando se trata de señales más complejas, como las señales de voz. La principal razón detrás de los resultados inferiores del cuantificador en el tiempo se debe a la falta de capacidad para capturar eficazmente la estructura de la señal de voz y sus características específicas. Cuando se cuantifica en el dominio del tiempo, no se consideran las relaciones y patrones entre las muestras. En contraste, los cuantificadores basados en *Lifting* y Mallat, utilizan la transformada *Wavelet*, que descompone la señal en diferentes niveles de resolución y frecuencia, capturando así tanto las características de alta frecuencia como las de baja frecuencia de la señal. Esta capacidad de descomponer la señal en diferentes componentes esenciales permite que los cuantificadores en el dominio transformado (*Lifting* y Mallat) enfaticen la información relevante, la cual se aprovecha en el diseño del cuantificador, en este caso con los diferentes métodos de CRR propuestos.

Cabe resaltar que la Figura 4.1, también evidencia que a partir de dieciséis niveles de cuantificación la curva del cuantificador del tiempo, supera a la media del cuantificador de Mallat, ya que, como se explicó anteriormente, el proceso de síntesis (transformada inversa) agrega una distorsión adicional sobre la señal a la que introduce el cuantificador; sin embargo, el valor máximo del cuantificador basado en Mallat coincide con el valor máximo del cuantificador basado en *Lifting* y supera al cuantificador del tiempo, es por eso que se debe escoger adecuadamente la familia *Wavelet* para el caso del algoritmo de Mallat, de lo contrario, el procesamiento adicional asociado a pasar a un dominio transformado pierde el sentido.

Adicionalmente, se puede observar que la brecha de calidad entre los cuantificadores en el dominio transformado y el cuantificador en el tiempo va disminuyendo

a medida que los niveles de cuantificación aumentan, esto debido a que tener un mayor número de niveles de cuantificación, está ligado al número de bits que tiene cada muestra para ser representada, en ese sentido, se permite una mejor representación de todas las muestras de la señal y, por ende, deja ser tan relevante el hecho de repartir eficientemente los bits de la trama para cada grupo de coeficientes según su relevancia, ya que en este caso todas las muestras de la señal ya poseen una buena representación.

Por último, en la Tabla 4.1 se muestran a detalle los valores de calidad resultantes después de procesar las señales de voz con los tres cuantificadores evaluados. Cabe resaltar que para los cuantificadores en el dominio *Wavelet* se hace uso de la familia que mejor resultado dio para la mayoría de los casos, es decir *Haar*. En esta tabla también se resaltan de color verde los valores máximos de calidad obtenidos para cada nivel de cuantificación, con ello se evidencia que el cuantificador *Lifting* mantiene una ventaja casi insignificante frente al cuantificador de Mallat a excepción del caso de 64 niveles de cuantificación, donde los dos cuantificadores presentan la misma calidad.

Tabla 4.1: Calidad de las señales procesadas con los tres cuantificadores evaluados.

Cuantificador	**Niveles de Cuantificación**				
	4	8	16	32	64
Mallat	0.7416	0.8250	0.8901	0.9393	0.9716
Lifting	0.7418	0.8251	0.8902	0.9394	0.9716
Tiempo	0.6649	0.7621	0.8414	0.9034	0.9497

4.1.2. Análisis de Pruebas Subjetivas

Tras la ejecución de las pruebas subjetivas exhaustivas, delineadas en el Apéndice B, y tomando en cuenta que estas evaluaciones se efectúan utilizando la familia *Haar*, se derivan los resultados que reflejan las calificaciones del CMOS, es decir, las señales de voz procesadas con los cuantificadores se comparan con la señal original. Estos resultados se presentan en la Figura 4.2, donde se distinguen las barras azules representando al cuantificador basado en *Lifting*, las barras rojas correspondientes al cuantificador basado en Mallat, y las barras amarillas que simbolizan al cuantificador en el dominio del tiempo.

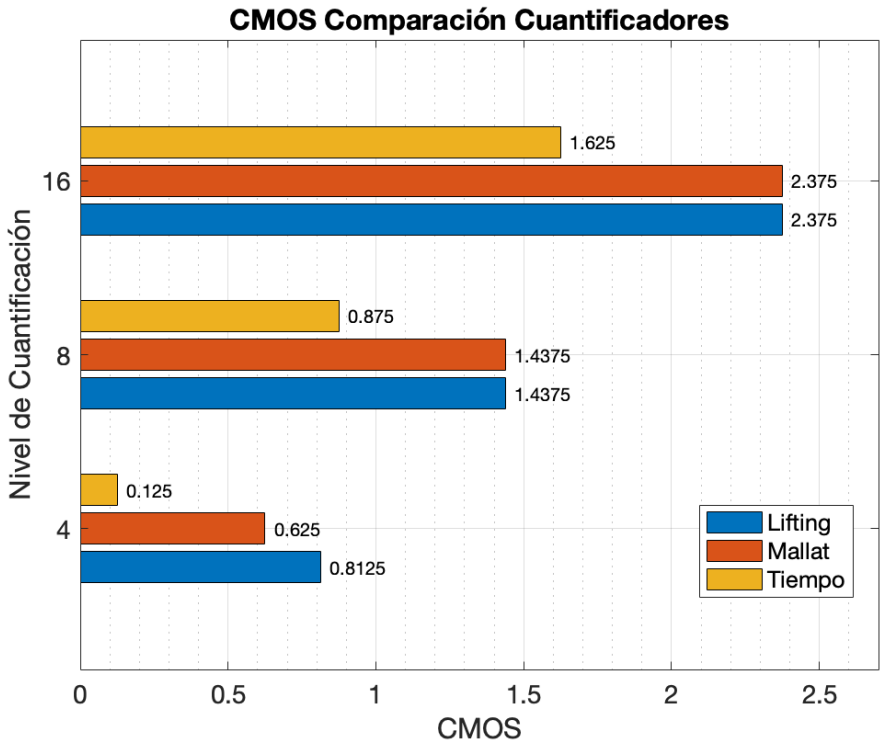


Figura 4.2: Comparación de calidad de los tres cuantificadores evaluados.

Para 16 niveles de cuantificación, tanto el cuantificador Mallat como el *Lifting* arrojan evaluaciones idénticas, lo cual se reafirma con 8 niveles de cuantificación. Este resultado robustece las pruebas cuantitativas, ya que los valores obtenidos para *Haar* en los cuantificadores de dominio transformado son comparativamente similares, como se visualiza en la Tabla 4.1.

En el caso de 4 niveles de cuantificación, a pesar de que los cuantificadores basados en Mallat y *Lifting* exhiben valores cuantitativos sumamente parecidos, los resultados cualitativos indican una leve ventaja del cuantificador basado en *Lifting* sobre el basado en Mallat. Esta pequeña disparidad podría deberse a efectos de la percepción auditiva, es decir, que la distorsión asociada a cada uno de los cuantificadores se percibe de manera diferente por los oyentes, lo cual es más notorio para niveles de cuantificación bajos; no obstante, no existe una diferencia significativa que permita determinar que *Lifting* tiene un mejor desempeño.

Además, es importante resaltar que los valores obtenidos para 16 niveles de cuantificación con los cuantificadores en el dominio transformado oscilan entre 2 (poco peor que el audio original) y 3 (casi igual que el audio original). Por otro lado, en el dominio del tiempo, se obtiene un resultado que varía entre 1 (peor

que el audio original) y 2 (poco peor que el audio original), lo que se interpreta como que en el dominio del tiempo la distorsión introducida siempre distancia a la señal de voz cuantificada con la señal de voz original, mientras que en el dominio *Wavelet* para algunas de las personas evaluadas las distorsiones no son notorias o no degradan la calidad de la voz de forma significativa.

Este patrón, en el cual las señales procesadas con el cuantificador en el tiempo obtienen resultados inferiores a nivel de calidad perceptual, en comparación a los otros cuantificadores, persiste para 8 y 4 niveles de cuantificación, con una tendencia descendente en la escala de calificación. Esto se alinea con la expectativa de que, al disminuir la cantidad de niveles de cuantificación, la precisión de las muestras cuantificadas disminuye, lo que en consecuencia afecta la calidad general de la señal.

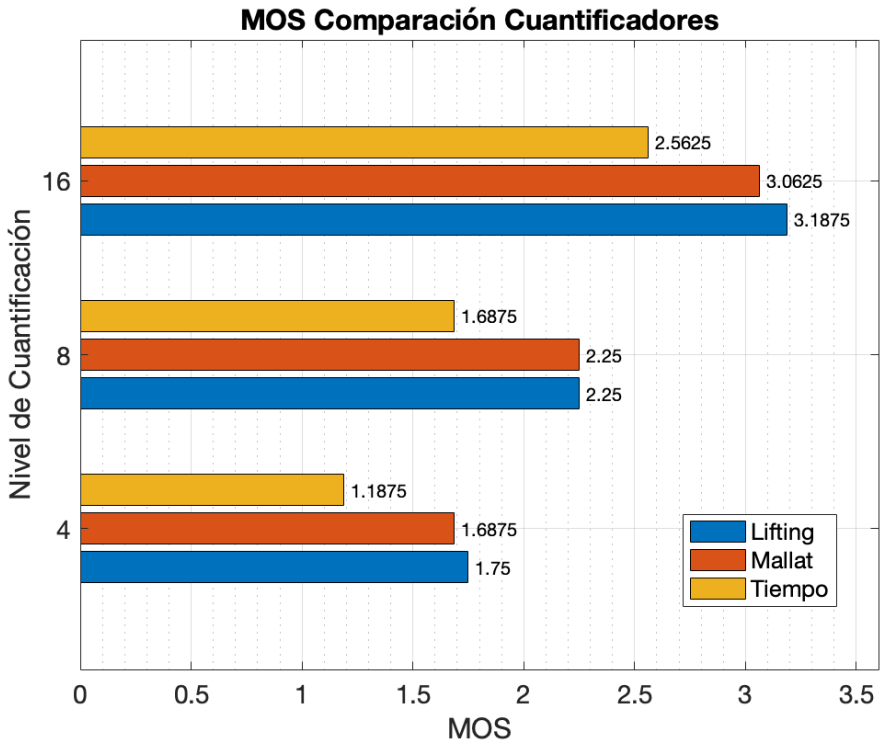


Figura 4.3: MOS de los cuantificadores evaluados.

En la Figura 4.3 se presentan los resultados obtenidos mediante la evaluación subjetiva utilizando la escala MOS, la cual refleja la percepción de calidad por parte de los oyentes para las señales procesadas con los tres cuantificadores implementados. Estos resultados ilustran la coherencia entre la distorsión percibida por los oyentes y los resultados cuantitativos. Para 16 y 4 niveles de cuantificación, se observa que la calidad percibida con el cuantificador basado en Mallat

es ligeramente inferior a la del cuantificador basado en *Lifting*. En el caso de 8 niveles de cuantificación, tanto Mallat como *Lifting* obtienen resultados idénticos, lo que sugiere que las diferencias entre estos dos procesos con estas variables no son significativas para los oyentes. Por otro lado, los resultados del cuantificador en el dominio del tiempo evidencian una desventaja en comparación con los otros cuantificadores, con una brecha más pronunciada en todos los casos.

Interpretando los resultados obtenidos en esta evaluación subjetiva, MOS, se puede concluir que el uso de los cuantificadores en el dominio transformado, en combinación con los métodos y algoritmos propuestos en este trabajo de grado, permite que para 16 niveles de cuantificación las señales procesadas tengan una calificación entre 3 (calidad de la señal justa) y 4 (calidad de la señal buena). Mientras que, las señales procesadas el cuantificador en el tiempo tengan una calificación entre 2 (calidad de la señal pobre) y 3 (calidad de la señal justa).

Para el caso de 8 niveles de cuantificación se observa que las señales procesadas por los cuantificadores basados en Mallat y *Lifting* tienen la misma calificación que está entre 2 (calidad de la señal pobre) y 3 (calidad de la señal justa) y las señales procesadas el cuantificador en el tiempo tienen una calificación entre 1 (calidad de la señal mala) y 2 (calidad de la señal pobre).

Finalmente, para 4 niveles de cuantificación el mejor resultado lo obtienen las señales procesadas por el cuantificador de *Lifting*, seguidamente las procesadas por el cuantificador de Mallat y por ultimo las procesadas por el cuantificador en el tiempo pero cabe resaltar que todas las señales son calificadas entre 1 (calidad de la señal mala) y 2 (calidad de la señal pobre), resultado coherente ya que al tener tan pocos niveles de cuantificación todas las señales poseen una distorsión muy molesta para los oyentes.

En general, se puede afirmar que los resultados subjetivos validan los hallazgos objetivos, ya que las diferencias entre los cuantificadores basados en *Lifting* y Mallat son mínimas en términos cuantitativos, y esto se ve reflejado también en una diferencia poco significativa en las calificaciones dadas por parte de los oyentes.

4.2. Variación de Tamaño de Trama

4.2.1. Análisis de Pruebas Objetivas

Una vez completadas las pruebas para analizar cómo la variación en la longitud de la trama afecta la calidad de la señal resultante, los resultados específicos para el cuantificador de *Lifting* se plasman en la Figura 4.4, los del cuantificador de Mallat se muestran en la Figura 4.6 y los del cuantificador en el tiempo en la Figura

4.7. El propósito de esta representación gráfica es visualizar el máximo valor de calidad obtenido en relación con el tamaño de la trama. En este sentido, se utiliza un código de colores para indicar las distintas longitudes de trama evaluadas y su correspondiente valor de calidad, de la siguiente manera: si la mejor calidad se logra con una longitud de trama de 64 ms, se representa mediante un tono azul oscuro; si la longitud de trama es 32 ms, el color utilizado es azul claro; para una longitud de trama de 16 ms, se emplea el color rojo y, finalmente, el amarillo denota 8 ms.

Cabe señalar que, en las gráficas, cada columna representa una variación en los niveles de cuantificación, y en el caso de las Figuras 4.4 y 4.6, cada fila corresponde a una familia *Wavelet* diferente. Es relevante destacar que estas familias *Wavelet* están organizadas de manera que la primera, de arriba hacia abajo, tiene asociados los filtros con menor longitud.

Analizando en detalle los resultados, en la Figura 4.4 se puede apreciar cómo la variación en las distintas longitudes de la trama influye en la calidad de los resultados. En esta figura se destaca que la longitud de trama que proporciona los mejores resultados de calidad es la de 8 ms. Esta observación lleva a la conclusión de que, para este cuantificador en particular, un menor tamaño de trama está asociado a un mejor rendimiento en el proceso de cuantificación, aspecto que se puede corroborar en la Tabla 4.2, en la cual se contrastan los valores de calidad para las duraciones de trama de 64 ms y 8 ms, en donde los mejores resultados se obtienen con 8 ms y se encuentran resaltados en color verde. A partir de estos resultados se evidencia que existe una brecha notable en las calidades obtenidas con las duraciones nombradas anteriormente, la cual se incrementa conforme disminuyen los niveles de cuantificación. Finalmente, se infiere que entre menor sea el número de niveles de cuantificación más conveniente es usar tramas de menor duración, porque la ganancia en calidad es mayor.

Tabla 4.2: Calidad con respecto a la variación de la trama.

Duración de Trama	**Niveles de Cuantificación**				
	4	8	16	32	64
8 ms	0.8294	0.9129	0.9626	0.9849	0.9940
64 ms	0.7418	0.8251	0.8902	0.9394	0.9716

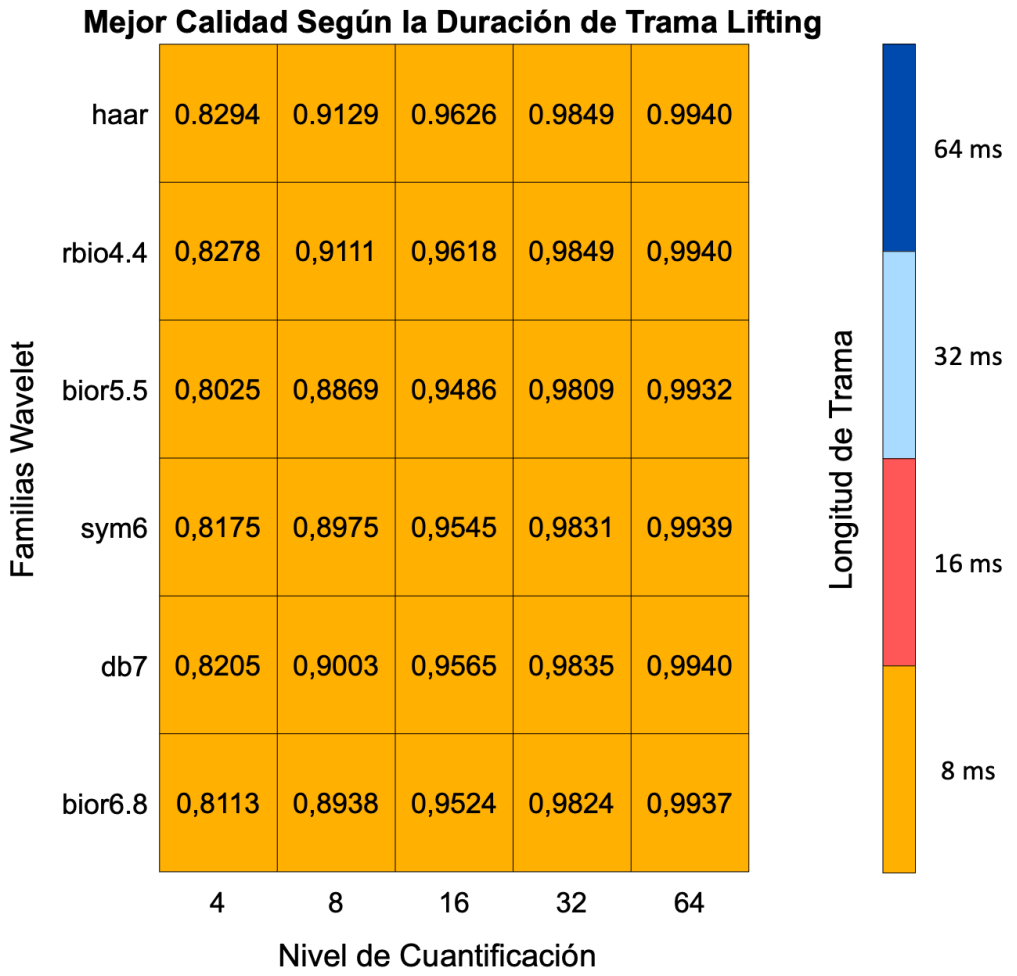


Figura 4.4: Calidad según la variación de longitud de la trama por familias *Wavelet* para el cuantificador *Lifting*.

Esta mejora en la calidad puede ser atribuida al hecho de que la dispersión de los coeficientes está directamente relacionada con la longitud de la trama. En otras palabras, a medida que aumenta el tamaño de la trama, también se observa un incremento en la dispersión de los coeficientes, aspecto que se evidencia en la Figura 4.5, en la cual se muestra el aumento de la dispersión en los coeficientes *Wavelet* y *Scaling* a medida que va incrementando el tamaño de la trama. Esta relación tiene un impacto significativo en los resultados de calidad brindada por el cuantificador, ya que el crecimiento de la dispersión conlleva que el rango dinámico se amplíe, lo que a su vez provoca un aumento en el tamaño del escalón utilizado en el proceso de cuantificación. Este aumento en el tamaño del escalón conlleva una disminución en la precisión de los niveles de cuantificación, afectando negativamente la calidad de la señal cuantificada.

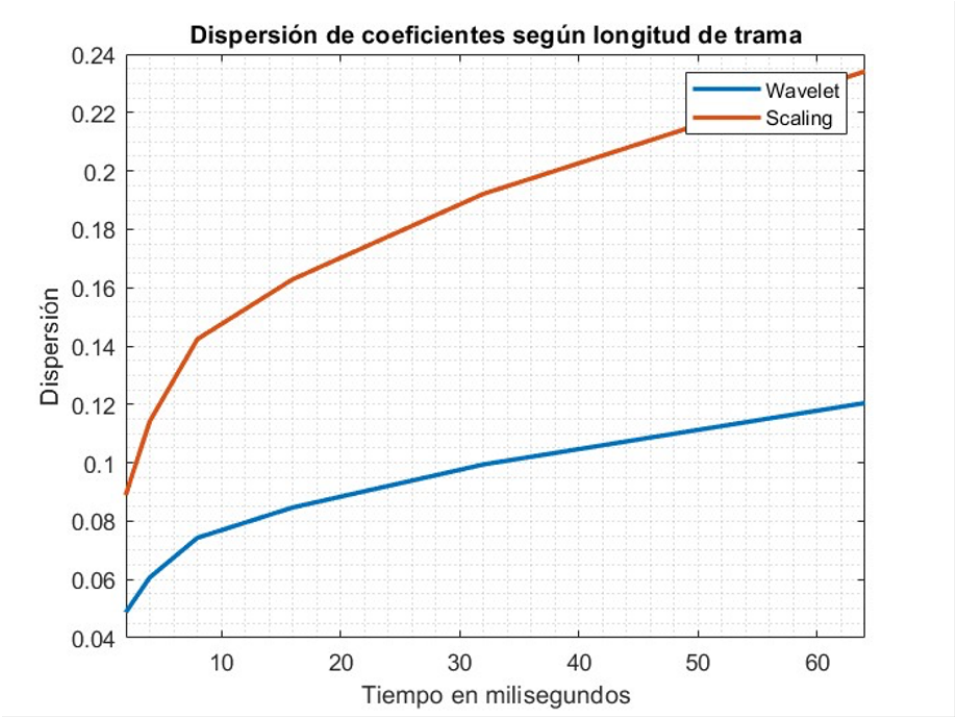


Figura 4.5: Dispersión de coeficientes según longitud de trama.

Al analizar los resultados de calidad relacionados con las variaciones en la longitud de la trama correspondientes para las diferentes familias *Wavelet* para el cuantificador Mallat, como se muestra en la Figura 4.6, se evidencia un comportamiento diferente en comparación con el cuantificador *Lifting*. Aquí, la óptima calidad no siempre se logra empleando la trama más corta. Detallando el comportamiento de este cuantificador en función de la longitud de la trama, se verifica que solo en el contexto de la familia *Haar* se experimenta una mejora en la calidad al reducir la longitud de la trama, y esto es consistente en todos los niveles de cuantificación.

Esta tendencia puede atribuirse a que la familia *Haar* posee una longitud de filtro pequeña (2 muestras). Por tanto, la disminución en la longitud de la trama no desfavorece la calidad de la señal, ya que el tamaño del filtro no es comparable con el tamaño mínimo de la trama, en este caso 8 ms.

Seguida de la familia *Haar*, se encuentra la familia *rbio4.4* con la segunda menor longitud de filtros entre las familias analizadas (10 muestras). En este escenario, se constata que para 4 niveles de cuantificación se logra la óptima calidad con una trama de 8 ms; sin embargo, al aumentar a 8 niveles de cuantificación, se obtiene la mejor calidad con 16 ms, y así sucesivamente, al aumentar el número de niveles de cuantificación se obtiene una mejor calidad con una longitud de

trama mayor.

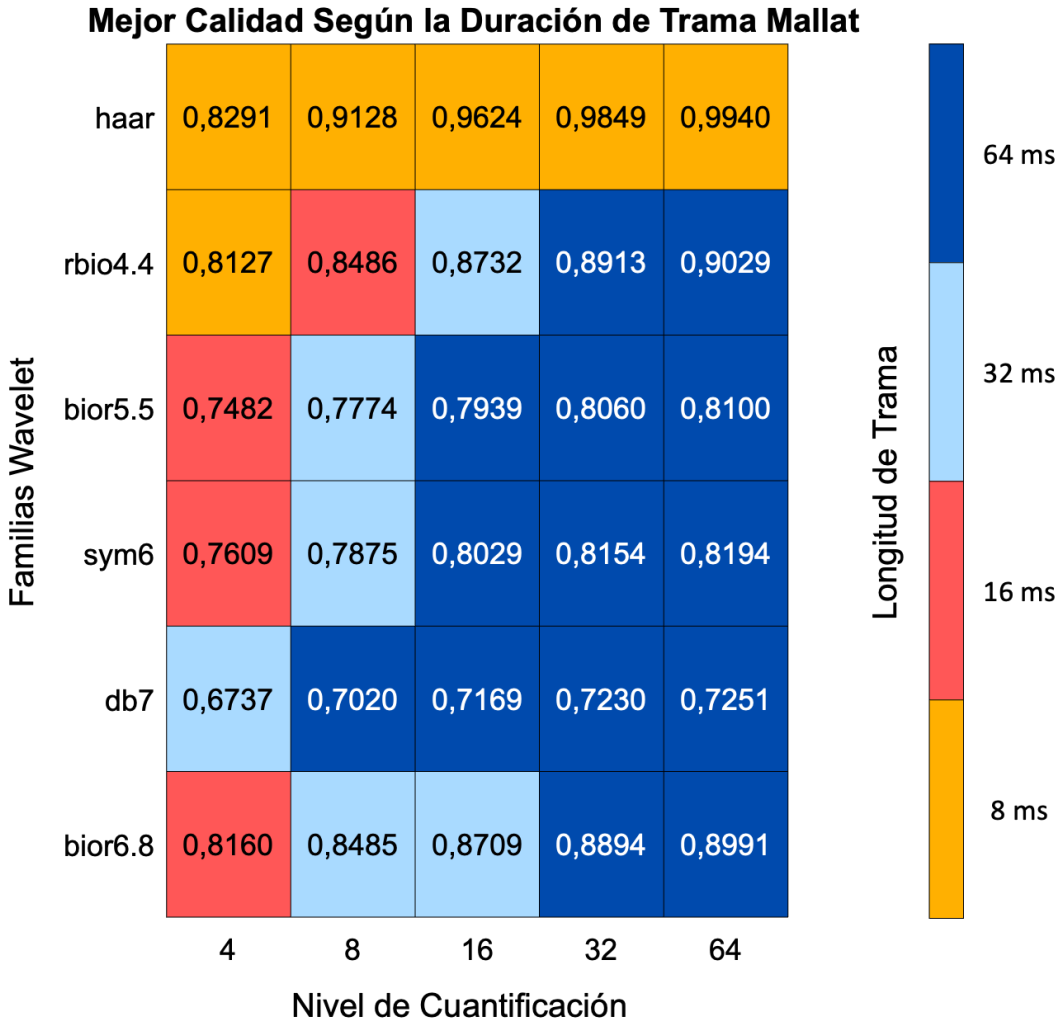


Figura 4.6: Calidad según la variación de longitud de la trama por familias *Wavelet* para el cuantificador Mallat.

Posteriormente se encuentran las familias *bior5.5* y la *sym6* con 12 muestras cada una. Su comportamiento para cuatro niveles de cuantificación indica que la mejor longitud de trama es la de 16 ms, y a medida que va aumentando el nivel de cuantificación, la mejor calidad se ve reflejada para longitudes de trama mayores, hasta llegar a 64 ms que es la máxima longitud de trama.

Para la familia *db7*, la cual tiene una longitud de filtro de 14 muestras, se puede observar que en 4 niveles de cuantificación se obtiene la mejor calidad con una

trama de 32 ms y desde 8 niveles de cuantificación en adelante la mejor calidad está dada por la trama de 64 ms.

En cuanto a la familia *bior6.8*, que es la que mayor longitud de filtro tiene, con 18 muestras, se observa que para 4 niveles de cuantificación vuelve a obtener la mejor calidad con la trama de 16 ms; para 8 y 16 niveles se obtiene que 32 ms es la mejor opción para la calidad y para niveles más altos de cuantificación resulta que el tamaño de trama más apropiado es 64 ms.

Este patrón sugiere que, a medida que la longitud del filtro comienza a crecer, hay otro factor relevante para la obtención de una buena calidad y es el número de niveles de cuantificación. En concordancia con lo mencionado, se infiere que entre mayor sea la longitud de los filtros de la familia utilizada, menor es el beneficio reflejado en la calidad, al seguir disminuyendo el tamaño de la trama, ya que se induce mayor distorsión debido a que el filtro resulta más equiparable con la dimensión de la trama. Adicionalmente, el aumento del número de niveles de cuantificación se traduce en una mejora significativa de la calidad, lo cual permite que se dé un mejor resultado de calidad sin necesidad de reducir la longitud de trama.

Adicionalmente, es importante resaltar que, en el caso específico de la familia *bior6.8*, se observa una discrepancia en el patrón analizado. Esto sugiere que entran en consideración otras características, intrínsecas de la familia *Wavelet*, como su varianza y la forma en la que se correlaciona con las señales procesadas. Estos factores pueden generar que la relación entre la longitud de trama y el tamaño de sus filtros no sigan el mismo comportamiento observado en otras familias *Wavelet* referente a la calidad obtenida. Es decir, en casos particulares, como con la familia *bior6.8*, ciertas peculiaridades de la familia pueden influir en cómo se interrelacionan estos aspectos y en los resultados de calidad que se obtienen. Esto permite resaltar un último aspecto y es el hecho de que esta familia *Wavelet* obtiene la segunda mejor ponderación de calidad después de la familia *Haar*.

Finalmente, en la Figura 4.7 se muestran los resultados del cuantificador en el tiempo según las variaciones de los niveles de cuantificación y se evidencia que los mejores resultados de calidad se dan para la menor longitud de trama, aspecto que se le atribuye a la relación directamente proporcional entre la longitud de la trama y la dispersión de los valores de amplitud de las muestras.

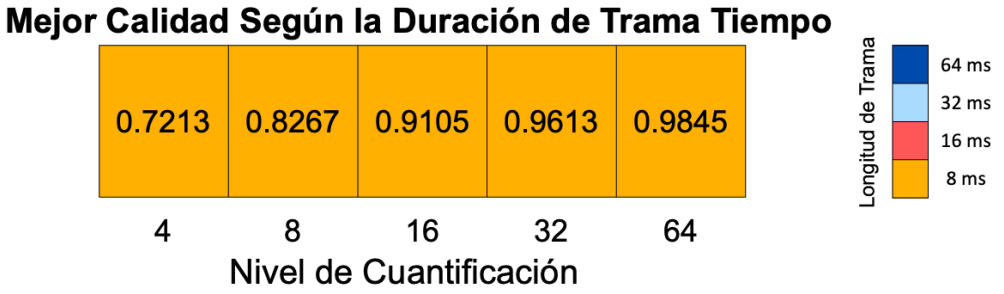


Figura 4.7: Calidad según la variación de longitud de la trama por familias *Wavelet* para el cuantificador en el tiempo.

4.2.2. Análisis de Pruebas Subjetivas

En el proceso de evaluación de la calidad de los audios procesados bajo diferentes duraciones de trama, se aplican dos pruebas específicas, descritas en detalle en el Apéndice B.

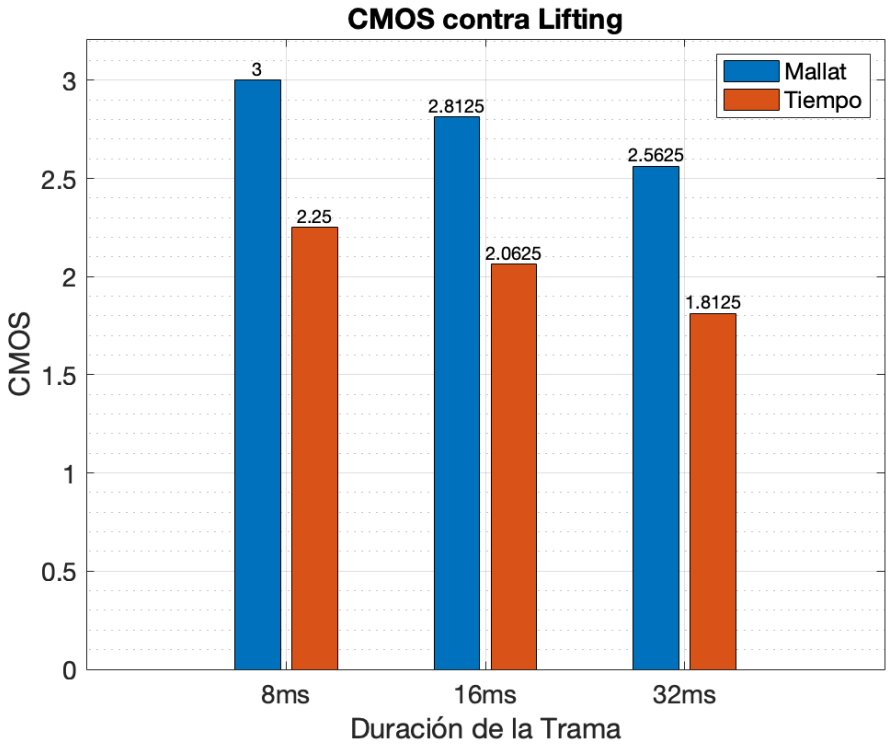


Figura 4.8: Calificación CMOS con el cuantificador *Lifting* de referencia y diferentes duraciones de trama.

Inicialmente se busca determinar si, al variar la longitud de la trama, se perciben diferencias entre los resultados obtenidos con el cuantificador basado en *Lifting* y los otros dos cuantificadores (Mallat y tiempo), para este análisis comparativo se utiliza la calificación CMOS. Los resultados obtenidos en estas pruebas se muestran en la Figura 4.8.

Al analizar la Figura 4.8 se destaca que el cuantificador basado en Mallat, proporciona resultados que oscilan alrededor del valor 3 en la escala CMOS para todas las variaciones de la longitud de trama estudiadas. Esto sugiere que las señales procesadas con el cuantificador basado en Mallat exhiben una similitud considerable con las procesadas utilizando el cuantificador *Lifting*. En contraposición, se observa que el cuantificador en el dominio del tiempo genera calificaciones cercanas a 2 para todas las duraciones de trama evaluadas, este patrón indica que las señales procesadas por el cuantificador en el dominio del tiempo presentan una calidad ligeramente inferior a las obtenidas mediante el cuantificador *Lifting*.

Adicionalmente, a partir de los resultados de la Figura 4.8 también se puede inferir que, a pesar de las variaciones en la duración de la trama, la brecha de calidad entre los distintos cuantificadores se mantiene prácticamente constante (las diferencias no son significativas). Estos resultados, alineados con las pruebas anteriores, consolidan la idea de que la longitud de la trama influye en la calidad, pero la diferencia entre los cuantificadores permanece consistente en proporción a las variaciones de la trama.

Finalmente, con el propósito de validar los resultados objetivos que muestran un aumento en la calidad de la señal reconstruida para duraciones de trama menores, se realiza una prueba subjetiva utilizando la rúbrica MOS, la cual tiene como objetivo principal explorar las variaciones en la duración de la trama entre 8 ms y 64 ms, y cómo estas variaciones influyen en la calidad que los oyentes atribuyen a las señales procesadas. Los resultados se obtienen utilizando el cuantificador basado en Lifting y se muestran en la Figura 4.9.

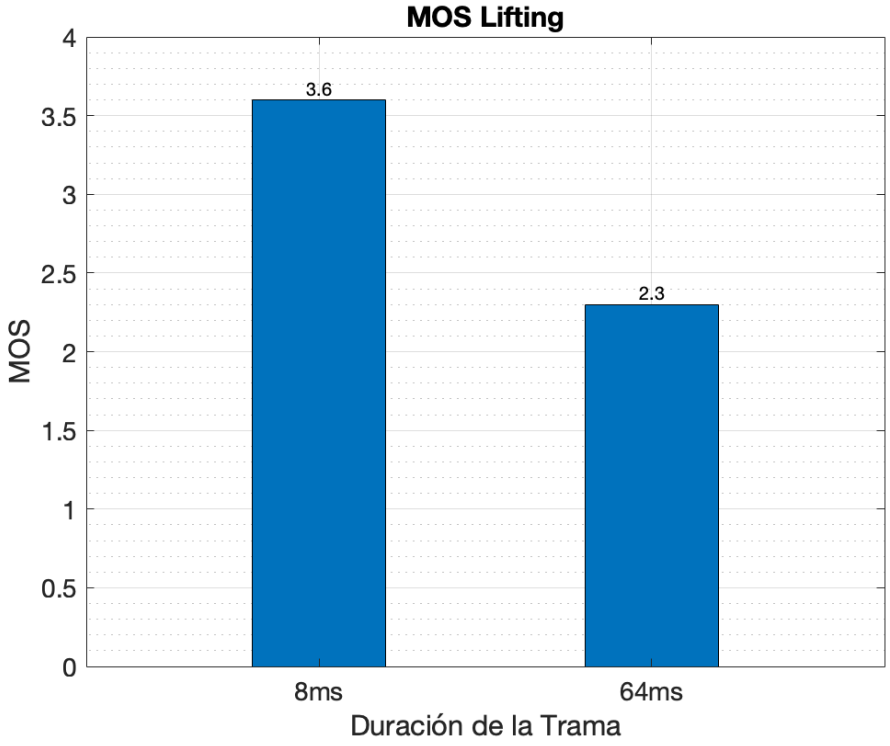


Figura 4.9: Calificación MOS con el cuantificador *Lifting* y diferentes duraciones de trama.

En la visualización de los resultados se evidencia una distinción clara en la calidad que los participantes perciben al modificar la longitud de la trama. Notablemente, la longitud de 8 ms conlleva una percepción de mayor calidad en comparación con la longitud de 64 ms. Estos resultados subjetivos validan los descubrimientos objetivos que se reflejan en la Tabla 4.2, reforzando la conclusión de que una duración de trama más breve se traduce en una calidad auditiva superior.

CAPÍTULO 5

CONCLUSIONES Y TRABAJOS FUTUROS

5.1. Conclusiones

La elección del método de transformación de dominio puede tener un impacto significativo en la calidad de la señal cuantificada. La aplicación del esquema de *Lifting* permite la transformación al dominio *Wavelet* sin inducir distorsión adicional a la señal, por tanto, la recuperación de la señal original se realiza de manera fiel. Por el contrario, se evidencia que el algoritmo de Mallat, en su proceso de análisis y síntesis, introduce cierto grado de distorsión en la señal, dependiendo de la familia *Wavelet* escogida y la longitud de la señal a analizar, ya que si las longitudes de los filtros y la señal son comparables se genera un escenario más propenso a la distorsión. Adicionalmente, esta distorsión se amplifica a medida que se incrementan los niveles de resolución, impactando negativamente en la fidelidad de la señal recuperada, debido a que la longitud de la señal analizada disminuye a medida que aumenta el nivel de resolución.

La caracterización proporcionada por la transformada *Wavelet*, por medio de los distintos coeficientes *Wavelet* y *Scaling*, resulta crucial en un proceso de cuantificación no uniforme, como el planteado en este trabajo de grado. Sin embargo, aunque aumentar el número de grupos de coeficientes, a través de los niveles de resolución, puede conducir a una distribución más desigual de la información de la señal, no necesariamente es lo más eficaz, ya que la ganancia en términos de calidad que aporta un número de niveles de resolución superiores a dos no es significativa, mientras que sí se tiene un costo computacional adicional asociado a incrementar la descomposición que se realiza sobre la señal.

La asignación de bits de cuantificación a través de los métodos de CRR emerge como un factor crucial, con un impacto de gran magnitud en la calidad resultante de la señal cuantificada. Este proceso es esencial, ya que constituye la manera en que se aprovecha la importancia intrínseca de cada uno de los grupos de coeficientes resultantes de la transformada desde diversas perspectivas. Por tanto, la elección de un método CRR es un factor relevante en el diseño de los cuantificadores en el dominio *Wavelet*, debido a que este método no solo incide en la calidad de la señal resultante, sino también en el rendimiento general del cuantificador, ya que la complejidad del método escogido puede desencadenar en un mayor costo computacional. No obstante, aunque el método CRR permite estimar la importancia de cada grupo de coeficientes en la reconstrucción de la señal, se encontró que, con el fin de tener una buena calidad, se deben designar unos bits de reserva para representar todos los coeficientes del dominio *Wavelet* con algún grado de precisión, por lo que el método CRR se encarga de repartir sólo una porción de los bits disponibles. Este resultado muestra que, para la frecuencia de muestreo considerada en las señales de voz de este trabajo de grado, es importante la información de la señal contenida en todos los grupos de coeficientes, aunque algunos grupos de coeficientes toleran un mayor grado de distorsión que

otros.

Existe una influencia de la longitud de la trama en la calidad de los resultados de cuantificación, a medida que se reduce la longitud de la trama, se obtienen mejoras sustanciales en la calidad de la señal cuantificada. La razón detrás de esta tendencia radica en la relación entre la longitud de la trama y la dispersión de los valores de amplitud a cuantificar. Una menor longitud de trama resulta en una reducción de la dispersión de los valores de amplitud, lo que a su vez conlleva a un menor rango dinámico y a una menor distorsión al realizar la aproximación a los niveles de cuantificación. Es importante resaltar que la compensación de esta mejora es una mayor exigencia de procesamiento, debido a que el cuantificador debe tratar un mayor número de señales sin que esto derive en un mayor retardo, por otro lado, se tiene que la relación entre carga útil y señalización disminuye, puesto que para una misma cantidad de información de señalización se reduce en cada trama la cantidad de información útil.

El cuantificador basado en el esquema *Lifting* presenta una gran estabilidad en la calidad de la señal al realizar las diferentes variaciones de los parámetros del cuantificador, cualidades que resaltan la robustez del cuantificador *Lifting* ante la diversidad de enfoques *Wavelet*, otorgándole una ventaja frente al cuantificador Mallat. Adicionalmente, para el cuantificador de *Lifting* su proceso se basa en operaciones matriciales que involucran multiplicaciones, lo que tiende a ser más eficiente, en términos computacionales, en comparación con las convoluciones sucesivas utilizadas en el cuantificador Mallat.

En general, tanto en los resultados objetivos, como en los resultados subjetivos, los cuantificadores en el dominio transformado permiten obtener un mejor desempeño que un cuantificador uniforme en el dominio del tiempo. Adicionalmente, a partir de las pruebas realizadas en este trabajo de grado se concluye que en términos generales es más conveniente utilizar: el esquema *Lifting* en lugar del algoritmo de Mallat para realizar los procesos de análisis y síntesis en el dominio *Wavelet*, el método CRR de percepción por su equilibrio entre calidad y costo de procesamiento, familias *Wavelet* de bajo orden como la *Haar*, y longitudes de trama dependientes del número de niveles de cuantificación disponibles, i.e., a menor número de niveles de cuantificación menor longitud de la trama.

5.2. Trabajos Futuros

El presente trabajo de grado se enfocó en la evaluación el desempeño del Cuantificador de Señales de Voz en el Dominio *Wavelet* con el Esquema *Lifting* y la comparación de éste con el Cuantificador de Señales de Voz en el Dominio *Wavelet* con el Algoritmo de Mallat y con el Cuantificador de Señales de Voz en

dominio del Tiempo. Para ello se diseñó un algoritmo que comprende diferentes parámetros y métodos, cómo lo son los métodos de CRR y la escogencia de las familias *Wavelet*, entre otros. En el desarrollo del modelo, se observó que estos parámetros requieren un análisis extenso, digno de una investigación específica para cada uno de ellos, es por esto que se proponen los siguientes trabajos futuros:

- Analizar a profundidad las características intrínsecas de cada familia *Wavelet* y su impacto en la calidad del cuantificador en el dominio *Wavelet* basado en *Lifting*.
- Analizar nuevos métodos de CRR, que permitan potenciar las características de las señales en el dominio *Wavelet*, y generar así un cuantificador en el dominio transformado con mayor eficiencia.
- Considerar la implementación de la transformada *Wavelet* entera a través del esquema *Lifting* y evaluar comparativamente el efecto de la distorsión del cuantificador sobre la señal de voz reconstruida.
- Evaluar el impacto de incluir una etapa previa de clasificación de tramas de habla y silencio sobre el Cuantificador de Señales de Voz en el Dominio *Wavelet* con el Esquema *Lifting*.
- Evaluar la calidad de los resultados obtenidos utilizando otros cuantificadores, como los cuantificadores no lineales en el dominio temporal o aquellos basados en la transformada de coseno, en comparación con el cuantificador implementado en este trabajo de grado basado en el esquema *Lifting*.

REFERENCIAS

[1] A. Jensen and A. l. Cour-Harbo, *Ripples in Mathematics: The Discrete Wavelet Transform*. Springer Science & Business Media, Jun. 2001, google-Books-ID: hMvhjWxb0_MC.

[2] "Digital Signal Processing System Design - 2nd Edition." [Online]. Available: https://www.elsevier.com/books/digital-signal-processing-system-design/kehtarnavaz/978-0-12-374490-6

[3] "Conversión de la señal analógica en digital. | ICTV02 .- La señal de radiodifusión. Primera parte: TV terrestre." [Online]. Available: https://ikastaroak.ulhi.net/edu/es/IEA/ICTV/ICTV02/es_IEA_ICTV02_Contenidos/website_522_conversin_de_la_seal_analgica_en_digital.html

[4] P. Colarusso, L. H. Kidder, I. W. Levin, and E. N. Lewis, "Raman and Infrared Microspectroscopy," in *Encyclopedia of Spectroscopy and Spectrometry*, J. C. Lindon, Ed. Oxford: Elsevier, 1999, pp. 1945–1954. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B0122266803004026

[5] L. Tan and J. Jiang, "Chapter 11 - Multirate Digital Signal Processing, Oversampling of Analog-to-Digital Conversion, and Undersampling of Bandpass Signals," in *Digital Signal Processing (Third Edition)*, 3rd ed., L. Tan and J. Jiang, Eds. Academic Press, 2019, pp. 529–590. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128150719000117

[6] U. Zölzer, *Digital Audio Signal Processing*. John Wiley & Sons, Mar. 2022, google-Books-ID: ccV6EAAAQBAJ.

[7] V. K. Garg and Y.-C. Wang, "2 - Digital Communication System Concepts," in *The Electrical Engineering Handbook*, W.-K. Chen, Ed. Burlington: Academic Press, Jan. 2005, pp. 957–964. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780121709600500700

[8] R. G. Gallager, *Principles of Digital Communication*. Cambridge University Press, Feb. 2008, google-Books-ID: 5W0aYFU02igC.

[9] M. A. U. Khan and M. J. T. Smith, “5.3 - Fundamentals of Vector Quantization,” in *Handbook of Image and Video Processing (Second Edition)*, ser. Communications, Networking and Multimedia, A. Bovik, Ed. Burlington: Academic Press, Jan. 2005, pp. 673–688. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780121197926501030

[10] J. Bellamy, *Digital telephony*, 3rd ed., ser. Wiley series in telecommunications and signal processing. New York: Wiley, 2000.

[11] L. Hanzo, F. C. A. Somerville, and J. Woodard, *Voice and Audio Compression for Wireless Communications*. John Wiley & Sons, Jun. 2008, google-Books-ID: 9VoHhMWufcQC.

[12] J. B. Alonso, M. A. Ferrer, J. de León, and C. M. Travieso, “CUANTIFICACIÓN DE LA CALIDAD DE LA VOZ PARA SU EVALUACIÓN CLÍNICA POR MEDIO DEL ANÁLISIS ACÚSTICO,” *IV Jornadas en Tecnología del Habla*, p. 6, 2006.

[13] M. M. Silva Zambrano, “Cuantificación de Señales de Voz Utilizando Wavelets,” Tesis de Maestría, Universidad del Cauca. Facultad de Ingeniería Electrónica y Telecomunicaciones. Maestría en Electrónica y Telecomunicaciones, Popayán, Cauca, 2022.

[14] P. C. Loizou, “Speech Quality Assessment,” in *Multimedia Analysis, Processing and Communications*, ser. Studies in Computational Intelligence, W. Lin, D. Tao, J. Kacprzyk, Z. Li, E. Izquierdo, and H. Wang, Eds. Berlin, Heidelberg: Springer, 2011, pp. 623–654. [Online]. Available: https://doi.org/10.1007/978-3-642-19551-8_23

[15] “ITU-R Recommendation BS.562-3 Subjective assessment of sound quality,” 1990.

[16] “ITU-T Recommendation P.830 Subjective performance assessment of telephone-band and wideband digital codecs,” 1996.

[17] E. H. Rothauser, W. D. Chapman, N. G. H. R. Silbiger, H. L. Hecker, G. E. Urbanek, K. S. S. AI, and M. Weinstock, “IEEE Recommended Practice for Speech Quality Measurements,” *IEEE Transactions on Audio and Electroacoustics*, vol. 17, no. 3, pp. 225–246, Sep. 1969, conference Name: IEEE Transactions on Audio and Electroacoustics.

[18] “ITU-T Recommendation P.10/G.100 Vocabulary for performance, quality of service and quality of experience,” 2017.

[19] “ITU-T Recommendation P.800 Methods for subjective determination of transmission quality,” 1996.

[20] W. Voiers, “Diagnostic acceptability measure for speech communication systems,” in *ICASSP ’77. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, May 1977, pp. 204–207.

[21] “ITU-T Recommendation P.835 Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm,” 2003.

[22] Z. Wang and A. C. Bovik, “Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures,” *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, Jan. 2009, conference Name: IEEE Signal Processing Magazine.

[23] W.-S. Lai, C.-J. Tseng, and J.-J. Ding, “Improved structural similarity measurement for vocal signals,” in *2013 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2013, pp. 301–304, iSSN: 2158-1525.

[24] S. Kandadai, J. Hardin, and C. Creusere, “Audio quality assessment using the mean structural similarity measure,” in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, Mar. 2008, pp. 221–224.

[25] M. Karjalainen, “A new auditory model for the evaluation of sound quality of audio systems,” in *ICASSP ’85. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 10, Apr. 1985, pp. 608–611.

[26] G. Chen and V. Parsa, “Loudness pattern-based speech quality evaluation using Bayesian modeling and Markov chain Monte Carlo methods,” *The Journal of the Acoustical Society of America*, vol. 121, no. 2, pp. EL77–EL83, Feb. 2007, publisher: Acoustical Society of America. [Online]. Available: https://asa.scitation.org/doi/10.1121/1.2430765

[27] A. Rix, J. Beerends, M. Hollier, and A. Hekstra, “Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs,” in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*, vol. 2, May 2001, pp. 749–752 vol.2, iSSN: 1520-6149.

[28] “Transformada de Fourier de tiempo corto - MATLAB stft - MathWorks América Latina.” [Online]. Available: https://la.mathworks.com/help/signal/ref/stft.html

[29] N. Kehtarnavaz, “CHAPTER 7 - Frequency Domain Processing,” in *Digital Signal Processing System Design (Second Edition)*, N. Kehtarnavaz, Ed. Burlington: Academic Press, Jan. 2008, pp. 175–196. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780123744906000076

[30] S.-Y. Huang and Z. Bai, “Wavelets, Advanced,” in *Encyclopedia of Physical Science and Technology (Third Edition)*, R. A. Meyers, Ed. New York: Academic Press, Jan. 2003, pp. 753–771. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B012227410500939X

[31] C. Burrus, R. Gopinath, and H. Guo, “Introduction to Wavelets and Wavelet Transform—A Primer,” *Recherche*, vol. 67, Jan. 1998.

[32] “Continuous Wavelet Transform and Scale-Based Analysis - MATLAB & Simulink - MathWorks América Latina.” [Online]. Available: https://la.mathworks.com/help/wavelet/gs/continuous-wavelet-transform-and-scale-based-analysis.html

[33] C. E. Heil and D. F. Walnut, “Continuous and Discrete Wavelet Transforms,” *SIAM Review*, vol. 31, no. 4, pp. 628–666, Dec. 1989, publisher: Society for Industrial and Applied Mathematics. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/1031129

[34] I. Daubechies, *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, Jan. 1992. [Online]. Available: http://epubs.siam.org/doi/book/10.1137/1.9781611970104

[35] S. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, Jul. 1989, conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[36] W. Sweldens, “The Lifting Scheme: A Custom-Design Construction of Biorthogonal Wavelets,” *Applied and Computational Harmonic Analysis*, vol. 3, no. 2, pp. 186–200, Apr. 1996. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1063520396900159

[37] T. Acharya and C. Chakrabarti, “A Survey on Lifting-based Discrete Wavelet Transform Architectures,” *Journal of VLSI signal processing systems for signal, image and video technology*, vol. 42, no. 3, pp. 321–339, Mar. 2006. [Online]. Available: https://doi.org/10.1007/s11266-006-4191-3

[38] A. M. Kadhim, N. Saad, S. Alsaad, and A. Mjeed, “Speech Steganography System Using Lifting Wavelet Transform,” *International Information Institute (Tokyo). Information*, Apr. 2019.

[39] A. Jensen and A. la Cour-Harbo, “The Discrete Wavelet Transform via Lifting,” in *Ripples in Mathematics: The Discrete Wavelet Transform*, A. Jensen and A. la Cour-Harbo, Eds. Berlin, Heidelberg: Springer, 2001, pp. 11–24. [Online]. Available: https://doi.org/10.1007/978-3-642-56702-5_3

[40] “G.711 : Pulse code modulation (PCM) of voice frequencies.” [Online]. Available: https://www.itu.int/rec/T-REC-G.711-198811-I/

[41] Bin Jiang and J. Yang, “Preferred frame length for the short-time magnitude spectrum on speech intelligibility and speech quality,” *2011 8th International Conference on Information, Communications & Signal Processing*, pp. 1–3, Dec. 2011, conference Name: 2011 8th International Conference on Information, Communications & Signal Processing (ICICS 2011) ISBN: 9781457700316 9781457700293 9781457700309 Place: Singapore Publisher: IEEE. [Online]. Available: http://ieeexplore.ieee.org/document/6174266/

[42] J. M. Ramirez, H. A. Romo, and M. M. Silva Zambrano, *Telecomunicaciones Digitales*, 1st ed. Editorial uc, 2020.

[43] “P.862 : Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs.” [Online]. Available: https://www.itu.int/rec/T-REC-P.862/

[44] “ITU-T Recommendation G.711 General aspects of digital transmission systems,” 1990.

[45] D. Kovačić and E. Balaban, “Voice gender perception by cochlear implantees,” *The Journal of the Acoustical Society of America*, vol. 126, no. 2, pp. 762–775, Aug. 2009. [Online]. Available: https://pubs.aip.org/jasa/article/126/2/762/903974/Voice-gender-perception-by-cochlear-implanteesa

[46] J. T. Eichhorn, R. D. Kent, D. Austin, and H. K. Vorperian, “Effects of Aging on Vocal Fundamental Frequency and Vowel Formants in Men and Women,” *Journal of voice : official journal of the Voice Foundation*, vol. 32, no. 5, pp. 644.e1–644.e9, Sep. 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5832520/

[47] “What Is the M4A Format? | M4A vs. MP3 vs. WAV.” [Online]. Available: https://cloudinary.com/guides/video-formats/what-is-the-m4a-format-understanding-the-difference-between-m4a-mp3-and-wav

[48] “ITU-R Recommendation BS.1284-2 General methods for the subjective assessment of sound quality,” 2019.

Apéndice A

REPOSITORIO DE SEÑALES DE VOZ

En este apéndice se explica el desarrollo de la construcción del repositorio de señales de voz, el cual tiene en cuenta las siguientes consideraciones:

1. **Número de personas grabadas.** Es importante tener suficientes personas grabadas para garantizar que la base de datos sea representativa y no sesgada hacia un grupo particular.

2. **Contenido de las grabaciones.** Se tiene en cuenta el tipo de palabras o frases que se van a grabar para la base de datos, seleccionando un guión que incluya palabras y frases que cubran diferentes sonidos y acentos del habla, así como diferentes niveles de dificultad. Además, el guión es lo suficientemente amplio para capturar la variabilidad del habla.

3. **Edades de las personas grabadas.** La edad de las personas grabadas puede tener un impacto significativo en el rango de frecuencias que se graban, por ello se tiene en cuenta que los niños y los adultos mayores tienden a tener un rango de frecuencias más limitado que los adultos jóvenes. Por ello se tiene en cuenta este aspecto para seleccionar a las personas que van a participar en la grabación y asegurarse de que se incluyan personas de diferentes edades en la base de datos.

4. **Formato de grabación.** Es fundamental asegurarse de que la calidad de grabación sea lo suficientemente alta para que las señales de voz sean claras y fáciles de entender. Por ello se hace uso de un micrófono de calidad y se escoge el formato más adecuado para los audios.

5. **Tamaño del repositorio de señales de voz.** El tamaño de la base de datos es importante para poder cubrir la variabilidad del habla. Por ello se considera que la base de datos contenga varios audios por cada persona seleccionada.

6. **Consentimiento de los participantes.** Es importante obtener el consentimiento de los participantes antes de grabar su voz y asegurarse de que comprendan cómo se utilizará su voz en el repositorio.

A.1. Número de Personas Grabadas

El número de personas que deben ser grabadas para tener un buen número de muestras de voz depende de varios factores, como la variabilidad en las características de la voz que se desea analizar y la complejidad del modelo de reconocimiento de voz que se utilizará.

Como regla general, se recomienda que haya al menos 10 grabaciones de voz por persona para tener suficiente variabilidad en las características de la voz.

Además, se recomienda que se graben al menos 30 personas para tener un conjunto de datos lo suficientemente grande para entrenar un modelo de reconocimiento de voz con resultados fiables. Teniendo en cuenta el sesgo que existe entre las frecuencias de voz de los hombres y las mujeres [45], se decide grabar a 15 mujeres y 15 hombres para tener un grupo de muestras homogéneo en cuanto al sexo de los participantes.

A.2. Contenido de las Grabaciones

Las grabaciones para el repositorio de señales de voz pueden ser realizadas de las siguientes maneras:

- **Texto libre.** Se le pide a la persona que hable de lo que quiera durante unos minutos.
- **Frases comunes.** Se le pide a la persona que repita una lista de frases comunes en su idioma nativo.
- **Preguntas y respuestas.** Se realizan preguntas a la persona y se solicita que responda en frases completas.
- **Números y letras.** Se le pide a la persona que diga una serie de números o letras en orden.
- **Escritura creativa.** Se le pide a la persona que lea un párrafo de un libro o que escriba algo en un papel y luego lo lea en voz alta.

Para el repositorio construido en este trabajo de grado se escoge la opción de frases comunes, ya que esto brinda una variedad de muestras de voz naturales y proporciona muestras de voz más emocionales o expresivas. Adicionalmente es la opción que permite tener las mismas frases para todas las personas, y de esta manera obtener un repositorio más homogéneo y organizado.

A.2.1. Frases del Guión

Después de decidir el uso de frases comunes para la construcción del repositorio de señales de voz, se llevó a cabo un proceso de selección de frases representativas en español. Se escogieron cuidadosamente 10 frases cortas que abarcaban diferentes palabras y estructuras lingüísticas para capturar la diversidad del habla cotidiana.

Es importante destacar que se procuró que las frases fueran equilibradas en términos de dificultad y frecuencia de palabras. Se evitaron frases demasiado

complejas o especializadas, así como frases muy repetitivas o demasiado simples. Esto permitió obtener una muestra amplia y variada de palabras y expresiones comunes del lenguaje.

A continuación, la lista de las frases escogidas:

- El éxito no es el resultado de la suerte, sino de la combinación de talento, trabajo duro y perseverancia.
- Se evidencia la dedicación y compromiso en cada página del trabajo de grado.
- La felicidad no es algo que se encuentra, es algo que se construye día a día.
- No es arrogancia, es confianza. Y la confianza es clave para el éxito.
- No hay duda de que su proyecto es una contribución valiosa al conocimiento científico.
- La libertad no es hacer lo que queremos, sino tener la capacidad de elegir lo que es mejor para nosotros y para los demás.
- Esta tesis será un referente en el área de estudio por su rigurosidad y profundidad.
- Soy el mejor en todo lo que hago, incluso en admitir mis errores.
- La única forma de hacer un gran trabajo es amar lo que haces.
- Siempre sé lo que estoy haciendo, aunque a veces los demás no lo entiendan.

A.3. Edades de las Personas Grabadas

La selección de las edades de los participantes también es un aspecto importante a considerar, ya que la voz humana presenta variaciones en el espectro de frecuencias a lo largo de diferentes etapas de la vida [46], por lo que al incluir participantes de diferentes edades, se garantiza que se capturen todas las frecuencias relevantes para el análisis de señales de voz. Por ello, en este trabajo de grado, se decide seleccionar participantes que abarcaban un rango de edades entre los 9 y 70 años, con el propósito de obtener un espectro completo de las frecuencias utilizadas en las aplicaciones de voz.

Además, al seleccionar participantes dentro de este rango de edades, se evaden de cierta forma las limitaciones significativas en la lectura de las frases. Los

participantes de estas edades tienden a tener una capacidad de lectura y pronunciación adecuada, lo que contribuye a obtener grabaciones de voz claras y comprensibles.

Así mismo, la inclusión de participantes en un amplio rango de edades refleja la diversidad de la población y mejora la representatividad de las muestras de voz utilizadas en el estudio. Esto permite obtener resultados más robustos y generalizables, ya que se consideran las variaciones vocales y lingüísticas propias de diferentes etapas de la vida.

A.4. Formato de Grabación

La selección del formato de grabación tiene gran importancia crucial en el procesamiento de señales de voz, ya que el formato utilizado puede afectar significativamente la calidad de la grabación y el costo de procesamiento posterior. En esta sección, se consideran los factores clave para elegir el formato de grabación, para garantizar una alta calidad de señal y una manipulación y análisis de datos eficientes.

Al elegir un formato de grabación, se debe tener presente el equilibrio óptimo entre la calidad de audio y el tamaño del archivo resultante, es por eso qué se toma la decisión de usar el formato M4A ya que este formato presenta las siguientes características [47] :

- **Alta calidad de audio.** Aunque M4A utiliza compresión con pérdida, el uso del códec AAC permite una alta calidad de audio incluso a tasas de bits más bajas. Es decir que se puede obtener una calidad de audio satisfactoria con archivos de tamaño más pequeño.
- **Tamaño de archivo más pequeño.** Gracias a la compresión con pérdida eficiente del códec AAC, los archivos M4A tienen un tamaño más reducido en comparación con los archivos WAV sin comprimir. Esto permite mejorar el costo de procesamiento, aspecto relevante para este trabajo de grado ya que se realizarán pruebas con grandes volúmenes de datos de señales de voz y se requiere un procesamiento más rápido.
- **Compatibilidad amplia.** El formato M4A es compatible con una amplia gama de dispositivos y reproductores de audio, lo que garantiza que se puedan reproducir los archivos en diferentes plataformas y dispositivos sin problemas.

- Este formato soporta mono, estéreo y canales de sonido envolvente; a 8, 16 o 24 bits por muestra [1].

A.5. Tamaño del Repositorio de Señales de Voz

Teniendo en cuenta los aspectos mencionados anteriormente sobre el formato de grabación y considerando la importancia de obtener una muestra representativa y diversa, el repositorio de señales de voz está compuesto por un total de 300 audios.

Incluye la participación de 30 personas en total, con una distribución equitativa de sexo. Seleccionando 15 mujeres y 15 hombres para asegurar una representación equilibrada y evitar sesgos relacionados con el género en el análisis de las señales de voz. Además, considera la diversidad en los rangos de edad de los participantes para capturar las variaciones inherentes al desarrollo vocal a lo largo de la vida.

Cada persona seleccionada para el estudio contribuye con 10 grabaciones de voz en el formato M4A, lo que da como resultado un total de 300 audios en el repositorio.

A.6. Consentimiento de los Participantes

Es importante que los participantes estén plenamente informados y que brinden su consentimiento de manera voluntaria y consciente para participar en la investigación. Este proceso incluye proporcionar a los participantes una explicación detallada de los objetivos del estudio, los procedimientos de grabación de voz, el uso y la confidencialidad de los datos, así como los posibles riesgos y beneficios asociados con su participación.

A continuación, se presenta el consentimiento informado presentado a los participantes.

[1] En todos estos casos el número de posibles valores de amplitud, es superior a los niveles de cuantificación considerados en este trabajo de grado, por lo que no afecta al funcionamiento del cuantificador propuesto.

FORMATO DE CONSENTIMIENTO INFORMADO PARA UN PROYECTO DE INVESTIGACIÓN

El Trabajo de Grado titulado **Cuantificación de Señales de Voz en el Dominio *Wavelet* Utilizando Esquema *Lifting***, cuyo objetivo es **Analizar el desempeño de un cuantificador de señales de voz en el dominio *Wavelet* utilizando el esquema *Lifting***, el cual se desarrollará por los estudiantes Lina Virginia Muñoz Garcés y Jhon Fredy Romero Núñez, pertenecientes al programa de Ingeniería Electrónica y Telecomunicaciones de la Facultad de Ingeniería Electrónica y Telecomunicaciones de la Universidad del Cauca, bajo la dirección de MSc. María Manuela Silva Zambrano adscrita al Departamento de Telecomunicaciones, se realizará como requisito para optar al título de Ingeniera en Electrónica y Telecomunicaciones.

1. **Participantes:** Personas en el rango de edad de 9 a 70 años.

2. **Propósito de la investigación:** Evaluar la calidad de la señal de voz cuantificada y medir la eficiencia del algoritmo de cuantificación en términos de preservación de la información y reducción de la distorsión. Mediante este análisis, se busca mejorar la comprensión de las propiedades y beneficios del esquema *Lifting* en la cuantificación de señales de voz, con el fin de contribuir al desarrollo de técnicas más efectivas en el procesamiento de señales de voz. Tipo de intervención de la investigación: Participará en la grabación de 10 audios para obtener una señal con sus frecuencias de voz.

3. **Selección de las participantes:** Usted está cordialmente invitado a formar parte de este proyecto de investigación por estar en un rango de personas entre 9 y 70 años.

4. **Participación voluntaria:** Usted puede elegir si quiere ser parte del proyecto de investigación. Si no quiere ser parte del estudio su trabajo continuará y nada cambiará. Aún si usted acepta ser parte del estudio ahora, usted se puede arrepentir luego y dejar de participar.

5. **Procedimientos:** Si Usted acepta participar en el estudio, se le realizara: la grabación de 10 audios de unas frases previamente plasmadas en un guión. El proceso tendrá una duración aproximada de (8) minutos.

6. **Riesgos y molestias:** Usted debe colaborar con su asistencia en la entrevista en profundidad. Esta será realizada por un estudiante del programa de Ingeniería Electrónica y Telecomunicaciones de la Universidad del Cauca.

7. **Beneficios:** Luego de finalizar el proyecto de investigación, los hallazgos pueden ayudar para el desarrollo de nuevos algoritmos para cuantificación de señales de voz que a su vez generarían mejoras en la calidad de la señal final obtenida en los servicios de comunicaciones.

8. **Incentivos:** Usted no recibirá ningún incentivo de índole económico (no se le dará dinero) por participar en el estudio.

9. **Confidencialidad:** La información derivada de la investigación será manejada de manera exclusiva por nuestro grupo de investigación. La información será guardada de manera segura por parte del equipo. Nadie fuera de nuestro equipo de investigación verá la información sobre usted.

   Cuando expliquemos la investigación a otras personas, no usaremos su nombre o nada que permita que otras personas conozcan su identidad. La información será guardada, cumpliendo los criterios de confidencialidad y respeto. Cabe resaltar que todos los datos personales utilizados en este proyecto no serán utilizados en otras investigaciones. Se construirá una base de datos, a la cual solo tendrá acceso el grupo de investigación supervisado por la Ing. María Manuela Silva Zambrano, C.C. 1.061'767.739, de la Facultad de Ingeniería Electrónica de la Universidad del Cauca, Campus de Tulcán. Celular: +57 316 495 8150.

10. **Divulgación de resultados:** Una vez toda la información haya sido analizada, escribiremos sobre nuestros resultados sin mencionar los datos de los participantes.

11. **Derecho a rehusar o a retirarse:** Como se dijo anteriormente, su participación es voluntaria. Usted puede retirarse del estudio en cualquier momento si así lo desea. En ese caso, su información será eliminada. Esta investigación contiene los elementos éticos que la ley y la doctrina exigen (código *Helsinki* - código *Nuremberg* resolución 008430 de 1993) que rigen la ética en la investigación científica en Colombia. Se garantiza total confidencialidad con los datos recolectados.

12. **Información de contacto:** Si tiene preguntas, las puede hacer ahora o posteriormente. Se le dará una copia escrita de este consentimiento. Si tiene preguntas adicionales, por favor contáctenos a través de la MSc. María Manuela Silva Zambrano. Celular: +57 316 495 8150.

13. **Certificado de consentimiento:** Entiendo que se me va a realizar una serie de grabaciones. Entiendo que no existe ningún riesgo. Sé que no recibiré dinero, sino el beneficio que los resultados de investigación que ayuden el desarrollo de nuevos algoritmos para cuantificación de señales de voz que a su vez generarían mejoras en la calidad de la señal final obtenida en los servicios de comunicaciones. Se me ha dado el nombre y dirección de un investigador que puede ser contactado fácilmente.

He leído o me ha sido leída la información precedente. He tenido la oportunidad de hacer preguntas. Estoy satisfecho/a con las respuestas a todas mis preguntas.

Doy consentimiento voluntario para hacer parte en este estudio. También puedo retirarme en cualquier momento.

Nombre legible del participante
Cédula.

He leído exactamente o he sido testigo de la lectura correcta del consentimiento al participante potencial, y éste ha tenido la posibilidad de hacer preguntas. Confirmo que el/la participante ha dado consentimiento libremente.

Lina Virginia Muñoz Garcés
Estudiante. CC:1.061'820.073
Código: 100618010435

Jhon Fredy Romero Núñez
Estudiante. CC:1.002'970.732
Código: 100617021479

María Manuela Silva Zambrano
Directora.
CC:1.061'767.739.

En constancia de aceptación, se firma el Acta por los que en ella intervienen, a los __ días del mes de _____ del 20__ y se da una copia de este consentimiento informado cada participante.

Apéndice B

MEDIDAS DE DISTORSIÓN SUBJETIVAS

Para evaluar y comprender la calidad de los cuantificadores de señales de voz, las medidas subjetivas juegan un papel crucial al proporcionar una visión directa y auténtica de cómo las personas perciben y valoran la calidad auditiva. En este apéndice se detallan los aspectos relevantes para llevar cabo las pruebas subjetivas que abordan los diversos cuantificadores empleados en el contexto de este estudio.

A continuación, se detallan los pasos realizados para llevar a cabo las pruebas subjetivas, los cuales comprenden desde los escenarios diseñados para las pruebas hasta la rúbrica de calificación que completarán los participantes, con ello, se busca capturar una variedad de perspectivas que enriquezcan la evaluación de la calidad perceptual. Con lo cual, se espera obtener resultados que respalden las pruebas objetivas previamente realizadas.

B.1. Cuantificadores Evaluados

Las pruebas se realizan con base en la evaluación de tres tipos de cuantificadores. Cada uno de estos cuantificadores posee sus propias características y enfoques para efectuar la cuantificación de señales de voz, lo que puede influir en la calidad auditiva percibida por los oyentes.

B.1.1. Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando Esquema *Lifting*

Este cuantificador utiliza los parámetros seleccionados en la sección 3.3 los cuales se evidencian nuevamente en la Tabla B.1. Como método para asignar los CRR se utiliza el método de percepción, los bits de reserva para cada muestra serán $\log_2(M) - 1$, se utilizan dos niveles de resolución y adicionalmente se escoge la familia *Wavelet Haar*, la cual obtuvo los mejores resultados de calidad en las pruebas objetivas.

Tabla B.1: Parámetros usando para el Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando Esquema *Lifting* en las pruebas subjetivas.

Parámetro	**Valor**
Método para asignar CRR	Percepción
Niveles de resolución, j	2
Bits de reserva, r	$\log_2(M) - 1$
Familia *Wavelet*	*Haar*

B.1.2. Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Algoritmo Mallat

Este cuantificador utiliza los mismos parámetros usados con el Esquema *Lifting* (Tabla B.1), con el fin de poder obtener una comparación justa y que la única diferencia radique en la forma en la que se realizan la DWT y la IDWT.

B.1.3. Cuantificador en el Dominio del Tiempo

Este cuantificador realiza una cuantificación uniforme en el dominio del tiempo, sin realizar los procesos adicionales de asignación de bits.

B.2. Mediciones Subjetivas

En concordancia con lo expuesto en la sección 3.2.5 de este trabajo, se utilizan dos importantes medidas subjetivas: la Puntuación de Media de Opinión (MOS) y la Puntuación Media de Opinión de Comparación (CMOS). Estas medidas no se basan en métricas técnicas o algoritmos de procesamiento, sino en la experiencia auditiva de los participantes, los cuales evalúan las señales cuantificadas y asignan puntuaciones que reflejan su opinión sobre la calidad percibida.

En el proceso de mediciones subjetivas, los oyentes escuchan las señales cuantificadas y proporcionan su opinión sobre la calidad de manera cuantitativa, con una escala predefinida. Esta metodología permite evaluar de manera más completa el desempeño de los cuantificadores, considerando no sólo las propiedades técnicas, sino también la experiencia auditiva real.

B.3. Condiciones de Pruebas

Para realizar las pruebas subjetivas se tiene en cuenta las siguientes condiciones, con base a las sugerencias dadas por la ITU en la recomendación ITU-R BS.1284-2 [48]:

- Todos los participantes de la prueba son estudiantes de últimos semestres o egresados de ingeniería electrónica y telecomunicaciones, por tanto, se consideran expertos en el tema.
- El número de oyentes total de la prueba es dieciséis ya que recomendación sugiere un mínimo de oyentes expertos por lo general de diez personas.
- Se intercambiarse de forma aleatoria el orden de lo audios escuchados para cada participante para evitar sesgos debido al orden.

- Se utiliza unos auriculares de alta calidad para que los participantes escuchen las señales de voz.
- Se aísla cada uno de los participantes para hacer la prueba.
- Se mantienen las mismas condiciones de prueba para todos los participantes.

B.4. Escenarios de Pruebas

Para llevar a cabo las pruebas subjetivas mencionadas previamente, se desarrollan dos conjuntos de pruebas, uno para cada tipo de medida (CMOS y MOS).

B.4.1. Primera Prueba

La primera fase de esta prueba se realiza con base en la medida CMOS, con la cual se compara la calidad del audio procesado en cada uno de los cuantificadores con respecto a la del audio original. En la Tabla B.3 se describen detalladamente los escenarios que se consideran para esta prueba.

La segunda fase de esta prueba se realiza con base en la medida MOS, con la cual se mide la calidad del audio que percibe cada uno de los participantes después de procesar la señal, para los tres cuantificadores evaluados. En la Tabla B.2 se describen detalladamente los escenarios que se desarrollan para esta prueba.

La prueba se realiza de tal forma que los participantes escuchan un audio original y un audio cuantificado por nivel de cuantificación y procedan a evaluar el MOS y el CMOS. En la Figura B.1 se observa un ejemplo de una iteración de la prueba, para la cual se tienen los siguientes pasos: primer paso, reproducción del audio original; segundo paso, reproducción del audio procesado con el Cuantificador de Voz en el Dominio *Wavelet* Utilizando el Esquema *Lifting*, con cuatro niveles de cuantificación; tercer paso, evaluación del CMOS a partir de la comparación del primero estimulo con el segundo estímulo; cuarto paso, evaluación del MOS para el audio cuantificado.

Tabla B.2: Escenarios de pruebas para medición de MOS.

E.	DESCRIPCIÓN	M	CUANTIFICADOR	EVALUACIÓN
1	Se califica la calidad de la señal procesada por el Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Esquema *Lifting*	4, 8, 16	Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Esquema *Lifting*	La calidad la señal con la rúbrica de MOS
2	Se califica la calidad de la señal procesada por el Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Algoritmo de Mallat	4, 8, 16	Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Algoritmo de Mallat	La calidad la señal con la rúbrica de MOS
3	Se califica la calidad de la señal procesada por el Cuantificador Uniforme en el Dominio del Tiempo	4, 8, 16	Cuantificador Uniforme en el Dominio del Tiempo	La calidad la señal con la rúbrica de MOS

Tabla B.3: Escenarios de pruebas para medición de CMOS.

E.	DESCRIPCIÓN	M	PRIMER ESTÍMULO	SEGUNDO ESTÍMULO	EVALUACIÓN
1	Se realiza la comparación entre el Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Esquema *Lifting* y la señal original	4, 8, 16	Señal original	Señal procesada por el Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Esquema *Lifting*	La calidad del segundo estimulo comparado con el primero con la rúbrica de CMOS
2	Se realiza la comparación entre el Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Algoritmo de Mallat y la señal original	4, 8, 16	Señal original	Señal procesada por el Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Algoritmo de Mallat	La calidad del segundo estimulo comparado con el primero con la rúbrica de CMOS
3	Se realiza la comparación entre el Cuantificador Uniforme en el Dominio del Tiempo y la señal original	4, 8, 16	Señal original	Señal procesada por el Cuantificador Uniforme en el Dominio del Tiempo	La calidad del segundo estimulo comparado con el primero con la rúbrica de CMOS

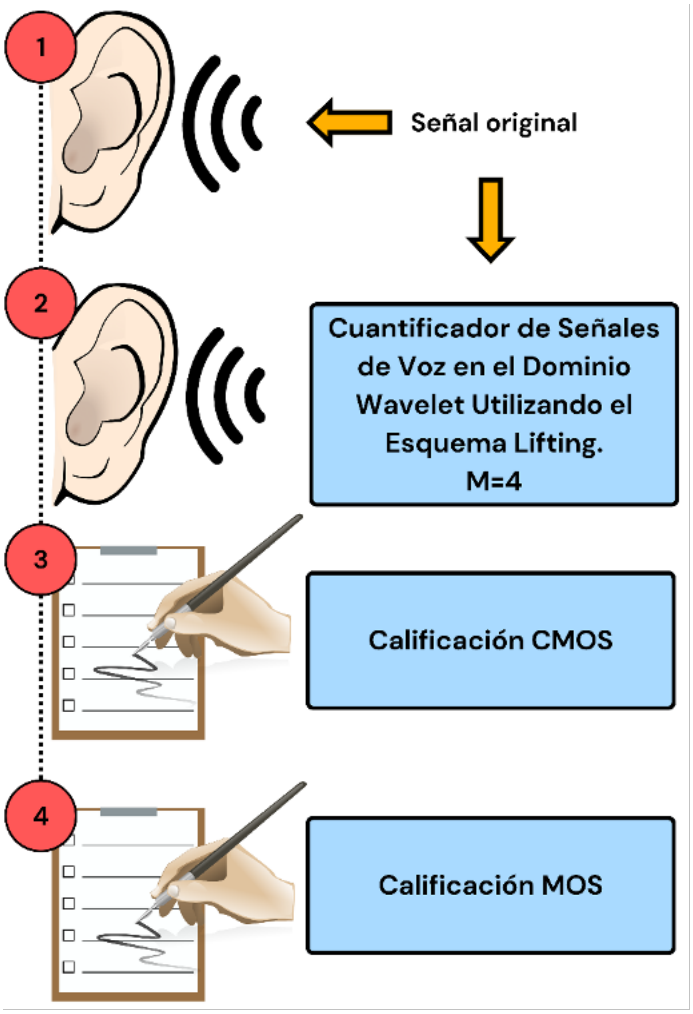


Figura B.1: Iteración de primera prueba de mediciones subjetivas.

B.4.2. Segunda Prueba

La segunda prueba consiste en la medición del CMOS entre los diversos cuantificadores, considerando la variación en la longitud de la trama.

Para llevar a cabo estas pruebas, se realiza una comparación de la calidad de la señal procesada utilizando diferentes enfoques: el Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Esquema *Lifting* y el Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Algoritmo de Mallat. Se varía la duración de la trama en intervalos de 8 ms, 16 ms y 32 ms, y también se evalúa el Cuantificador de Señales de Voz en el Dominio *Wavelet* Utilizando el Esquema *Lifting* en comparación con el Cuantificador en el dominio del tiempo. Las mismas variaciones en el tamaño de la trama se aplican en este último caso.

Cabe resaltar que las pruebas se realizan con cuatro niveles de cuantificación, ya que es escenario más crítico. Esta prueba tiene como fin determinar si existen diferencias significativas entre los cuantificadores evaluados al variar el tamaño de las tramas.

B.4.3. Tercera Prueba

La tercera prueba consiste en la medición de la calidad de la señal por medio de MOS para diferentes variaciones del tamaño de la trama (8 ms y 64 ms) con el cuantificador de Lifting, cabe destacar que, se utiliza el escenario más crítico para los niveles de cuantificación, es decir cuatro niveles. Esta prueba tiene como fin, determinar cuál es el tamaño de trama que mayor calidad brinda ante la percepción de los oyentes

B.5. Rubricas de Evaluación

Para que los participantes evalúen los diferentes audios planteados para los escenarios nombrados es la sección B.4. se tiene una rubrica de calificación para cada tipo de medida. Para ellos se hace uso de las rúbricas evidenciadas en la Tabla 1.1 (B.4) y la Tabla 1.2 (B.5) del Capítulo 1 de la monografía para calificar el CMOS y el MOS respectivamente.

Tabla B.4: Calificaciones de comparación categórica usados en CMOS.

Calificación	**La calidad del segundo estimulo comparado con el primero es:**
6	Mucho mejor
5	Mejor
4	Poco mejor
3	Casi igual
2	Poco peor
1	Peor
0	Mucho peor

Tabla B.5: Escala de calificación MOS.

Calificación	**Calidad de la señal**	**Nivel de distorsión**
5	Excelente	Imperceptible
4	Buena	Apenas perceptible, pero no molesta
3	Justa	Perceptible y un poco molesta
2	Pobre	Molesta, pero no detestable
1	Mala	Muy molesta y detestable