ALGORITMO DE ESTEGANOGRAFÍA DE AUDIO BASADO EN CODIFICACIÓN DE FASE

Universidad
del Cauca

Sofia Dussan Narvaez
Adriana Sofia Henao Ordoñez

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

ALGORITMO DE ESTEGANOGRAFÍA DE AUDIO BASADO EN CODIFICACIÓN DE FASE

Universidad
del Cauca

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

Sofia Dussan Narvaez
Adriana Sofia Henao Ordoñez

Directora: MSc. María Manuela Silva Zambrano
Codirector: PhD(c). Siler Amador Donado

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

TABLA DE CONTENIDO

LISTA DE FIGURAS

LISTA DE TABLAS

LISTA DE ACRÓNIMOS

ADC Conversión Analógico a Digital, *Analog to Digital Converter*.

BER Tasa de Error de Bit, *Bit Error Rate*.

CMOS Comparación de la Puntuación Media de Opinión, *Comparison Mean Opinion Score*.

DAC Conversión Digital a Analógico, *Digital to Analog Converter*.
DFT Transformada Discreta de Fourier, *Discrete Fourier Transform*.

FFT Transformada Rápida de Fourier, *Fast Fourier Transform*.
FT Transformada de Fourier, *Fourier Transform*.

HAS Sistema Auditivo Humano, *Human Auditory System*.

IFFT Transformada Rápida Inversa de Fourier, *Inverse Fast Fourier Transform*.

LSB Bit Menos Significativo, *Least Significant Bit*.

M-NRMSE Error Cuadrático Medio Normalizado Promedio, *Mean Normalized Root Mean Square Error*.
MOS Puntuación de Opinión Media, *Mean Opinion Score*.
MOV Valores de Resultados del Modelo, *Model Output Value*.
MSB Bit Más Significativo, *Most Significant Bit*.

ODG Grado de Diferencia Objetiva, *Objective Difference Grade*.

PCM Modulación por Impulsos Codificados, *Pulse Code Modulation*.
PEAQ Evaluación Perceptual de Calidad de Audio, *Perceptual Evaluation of Audio Quality*.
PSD Densidad Espectral de Potencia, *Power Spectral Density*.

SNR Relación Señal a Ruido, *Signal to Noise Ratio*.
SSIM Indice de Similitud Estructural, *Structural Similarity Index*.

INTRODUCCIÓN

La cantidad de información digital a nivel mundial se ha incrementado considerablemente con la integración de diversas redes y usuarios a Internet, generando un interés creciente en áreas como la seguridad digital. Por ello, surgen disciplinas como la criptografía y la esteganografía: la criptografía busca hacer que un mensaje sea indescifrable para un observador que no posee la clave adecuada; sin embargo, éste reconoce que hay información secreta presente en dicho mensaje, llevándolo a tratar de extraerla. Por su lado, la esteganografía se fundamenta en el desconocimiento por parte del observador sobre el envío de información confidencial, la cual se encuentra oculta en un medio portada que es transportado a través de un canal de comunicación.

La esteganografía es una disciplina diversa y versátil que permite la implementación de distintos tipos de archivos como medio portador, incluyendo vídeos, imágenes y, de particular interés en este caso, audios. Sin embargo, debe tenerse en cuenta la naturaleza propia de estos archivos y la técnica a implementar para llevar a cabo un procesamiento adecuado para la inserción de los datos secretos. La aplicación de la esteganografía en archivos de audio resulta muy conveniente, ya que hay una gran diversidad de señales de audio digital y su propagación a través de las redes de comunicación es universal y frecuente, lo que hace que se consideren excelentes transportadoras de información oculta, además, la redundancia de ciertos archivos de audio hace que la información secreta pase totalmente inadvertida debido a las limitaciones del Sistema Auditivo Humano (HAS, *Human Auditory System*), si es que un observador no deseado intercepta el archivo portador conocido como archivo *stego*.

Para llevar a cabo un proceso estenográfico ideal debe tenerse en cuenta sus tres propiedades fundamentales: la robustez, que se encuentra relacionada con la supervivencia del mensaje secreto ante los procesamientos propios de una red o externos a la misma, la imperceptibilidad, la cual hace referencia a la similitud perceptiva entre el medio portada antes de la incrustación y el archivo *stego* como resultado de la inserción de información y, la capacidad de incrustación, que se entiende como la cantidad de información secreta que se guarda en el archivo portada.

Dado que mantener un equilibrio entre las tres propiedades de la esteganografía puede resultar en un proceso bastante complejo, se han explorado diversas técnicas tanto en el dominio temporal como en el dominio transformado para la aplicación de esta disciplina. Aunque algunas son más comunes que otras, los algoritmos implementados en el dominio transformado ofrecen una mayor robustez en cuanto a la inaccesibilidad del mensaje secreto. Adicionalmente, la imperceptibilidad del archivo *stego* en un canal de comuni-

caciones y la cantidad de información secreta que transporta varía dependiendo de la técnica esteganográfica aplicada; tal es el caso de la técnica de codificación de fase en la cual se realizan cambios de fase a un determinado segmento de audio que representan cierta cantidad de información secreta, bajo la premisa de que las modificaciones sobre ciertas componentes de frecuencia pasan desapercibidas por el HAS. Cabe resaltar que la obtención del espectro de fase de la señal de audio portada se realiza mediante la aplicación de una transformada, que en éste caso es la Transformada de Fourier (FT, *Fourier transform*).

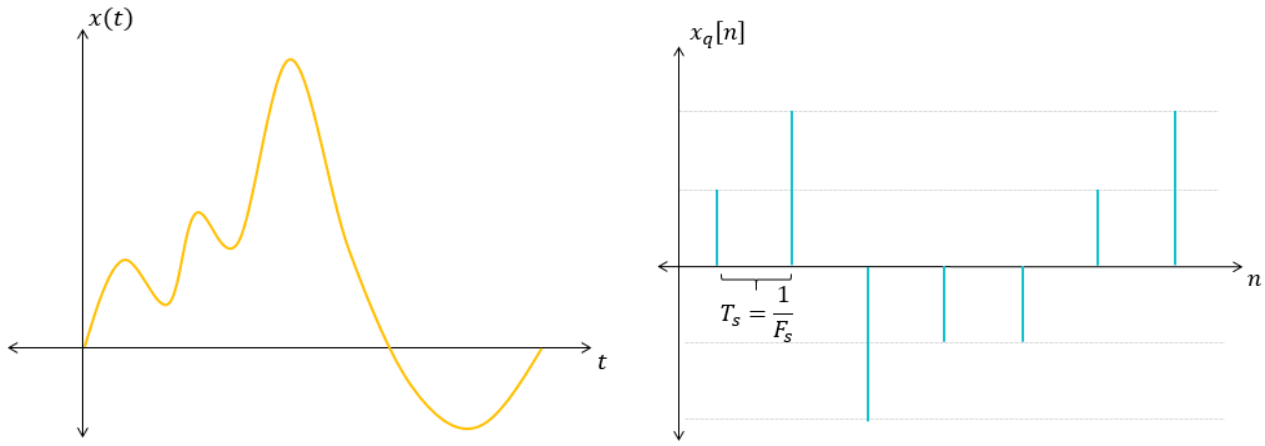
Bajo este contexto, en el presente trabajo de grado se propone un algoritmo de esteganografía de audio basado en codificación de fase y se evalúa su desempeño, en términos del grado de similitud entre la señal de audio portada y el audio *stego* resultante tras la inserción, con relación a la imperceptibilidad y la capacidad de incrustación de información según medidas objetivas y subjetivas, realizando una comparación con un método representante de la esteganografía de audio en el dominio temporal.

CAPÍTULO 1

GENERALIDADES

1.1. Señales Analógicas y Digitales

Una señal analógica es una representación de un fenómeno físico que, a lo largo del tiempo, toma valores continuos, esto es: una señal continua en amplitud y tiempo; mientras que una señal digital es aquella que se encuentra compuesta por un conjunto de muestras con amplitud y tiempo discretos; generalmente estas señales se representan por medio de una secuencia binaria. En la Figura 1.1 se ejemplifican estos tipos de señales.



(a) *Señal analógica* (b) *Señal digital*

Figura 1.1: *Señales analógica y digital.*

Por su parte, con el desarrollo de los microprocesadores y procesadores digitales es posible enlazar dichas variables analógicas con los procesos digitales a través de los llamados: Conversores Analógico a Digital (ADC, *Analog to Digital Converter*) y los Conversores Digital a Analógico (DAC, *Digital to Analog Converter*); cuyo objetivo básico es transformar una señal eléctrica analógica en una representación digital y, de la misma forma, transformar una señal digital en su equivalente analógico [2].

1.1.1. Conversión analógico a digital

Los ADC están compuestos por tres etapas: muestreo, cuantificación y codificación de fuente, las cuales se ilustran en la Figura 1.2.

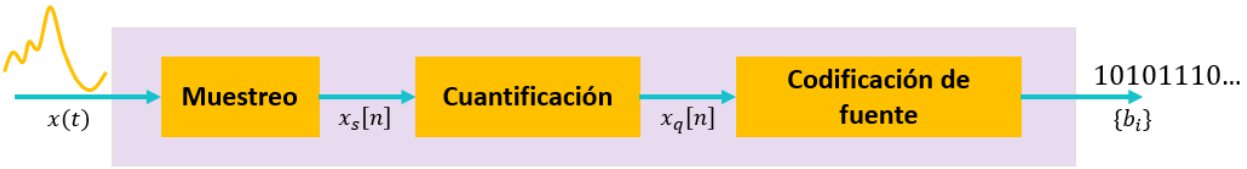


Figura 1.2: *conversión analógico a digital [2].*

Este proceso inicia con la etapa de muestreo, en donde se convierte la señal original en una secuencia de muestras $x_s[n]$, tomadas periódicamente con una frecuencia tal, que se garantice la reconstrucción de la señal original a partir de sus muestras, es decir, para asegurar la correcta conversión se debe considerar la frecuencia de muestreo $F_s = \frac{1}{T_s}$, para lo cual el Teorema de muestreo Nyquist-Shannon establece que ésta debe ser de mínimo dos veces el ancho de banda de la señal muestreada $(F_s = 2Bw)$, de lo contrario se da paso al fenómeno de distorsión conocido como *aliasing* [2].

La cuantificación, por su parte, es la etapa en donde se clasifican las amplitudes de dichas muestras dentro de un conjunto finito de posibles niveles de cuantificación (M niveles) que hacen parte de un alfabeto, como se observa en la Figura 1.3, obteniendo así una discretización de la forma de onda tanto en tiempo como en amplitud. Cabe resaltar que es en este bloque donde se pierde la mayor cantidad de información, debido a las aproximaciones que se deben considerar para realizar la discretización, generando un error de cuantificación (ε), que hace alusión a la diferencia entre la muestra original y la muestra cuantificada; sin embargo, aunque este error siempre está presente, la distorsión se puede mitigar al expandir el alfabeto de cuantificación, reduciendo la distancia entre cada uno de los niveles, lo que a su vez exige una mayor capacidad de procesamiento y almacenamiento en los dispositivos digitales.

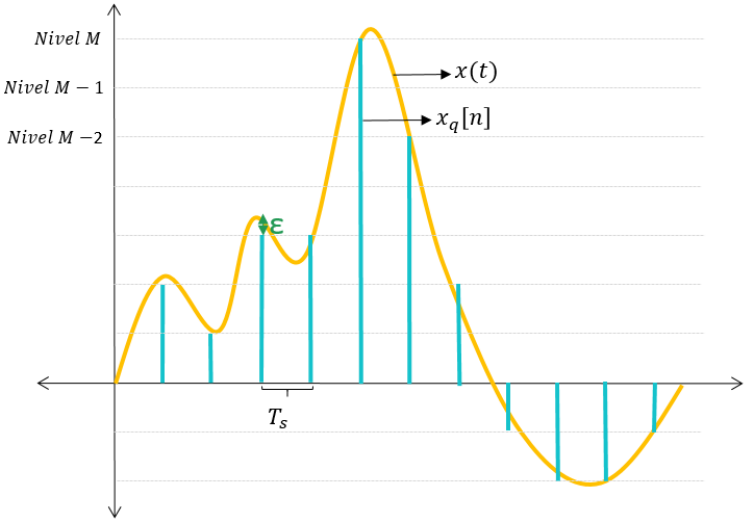


Figura 1.3: *Etapa de cuantificación en el ADC.*

Finalmente, en la etapa de codificación de fuente se asigna una palabra código binaria cuya longitud puede ser fija o variable; según sea la probabilidad de ocurrencia de cada uno de los niveles de cuantificación previamente establecidos, permitiendo representar la señal original a partir de una secuencia de bits [2].

1.1.2. Conversión digital a analógico

Al someter una secuencia de bits a la conversión digital a analógico, es necesario conocer la codificación de fuente, el alfabeto de cuantificación y la frecuencia de muestreo implementada en la ADC.

Los bits se agrupan según la longitud de la codificación de fuente usada en la ADC. Cada agrupación representa un nivel de amplitud (perteneciente al alfabeto de cuantificación), que se establece en un instante de tiempo dado por el periodo de muestreo (T_s). Finalmente, se implementa un filtro pasa bajo, conocido como filtro interpolador [2], el cual permite suavizar la señal resultante. Este proceso se presenta en la Figura 1.4.

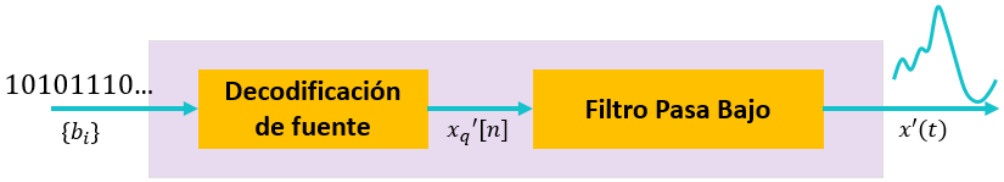


Figura 1.4: *Conversión digital a analógico.*

En términos prácticos, la salida del DAC no es una representación exacta de la señal analógica original o deseada, debido a las aproximaciones realizadas en la etapa de cuantificación. En la Figura 1.5 se ejemplifican las posibles diferencias entre la señal original y la señal reconstruida.

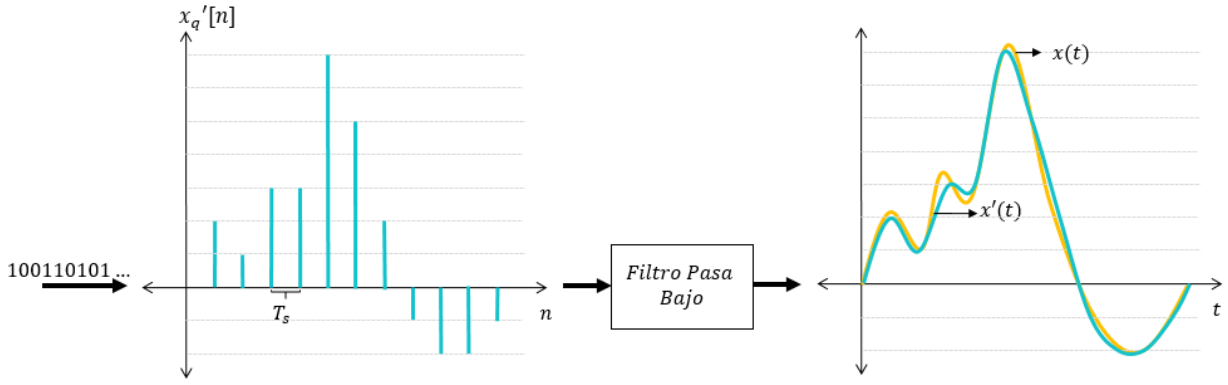


Figura 1.5: *Distorsión por cuantificación. Muestras reconstruidas ($x_q[n]$), señal original($x(t)$) y reconstruida ($x'(t)$).*

1.1.3. Análisis espectral

Las señales pueden someterse a un análisis en un nuevo dominio mediante el uso de una transformación, gracias a ello se obtiene otra forma de representar la información de la señal garantizando, idealmente, que el proceso es invertible. En este caso, el dominio de la frecuencia es el deseado; para ello se recurre a la FT, la cual es una transformación bidireccional. La FT utiliza como funciones base señales sinusoidales, las cuales se pueden representar como exponenciales complejas, por lo que se tiene que:

1. Transformada de Fourier: permite obtener una señal en el dominio de la frecuencia, $\tilde{x}(f)$, a partir de su representación temporal, $x(t)$.

$$\tilde{x}(f) = F\{x(t)\} = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}\,\mathrm{d}t. \tag{1.1}$$

2. Transformada Inversa de Fourier: permite obtener la señal temporal, $x(t)$, a partir de su representación en el dominio de la frecuencia, $\tilde{x}(f)$.

$$x(t) = F^{-1}\{\tilde{x}(f)\} = \int_{-\infty}^{\infty} \tilde{x}(f)e^{j2\pi ft}\,\mathrm{d}f. \tag{1.2}$$

Partiendo de la Ecuación 1.1 se tiene que la señal en el dominio de la frecuencia es compleja, por lo que gráficamente esta señal se puede analizar por medio de sus espectros de magnitud y fase, expuestos en la Ecuación 1.3. El espectro de magnitud es el más usado, dado que brinda información sobre las componentes de frecuencia de mayor influencia, la distribución de la energía o la potencia de la señal, el ancho de banda, entre otros.

$$F\{x(t)\} = |\tilde{x}(f)|e^{j\varphi_x} = Re\{\tilde{x}(f)\} + jIm\{\tilde{x}(f)\}. \tag{1.3}$$

Si se tiene una señal real en el dominio del tiempo, como las señales de audio, por propiedades de la FT, el espectro de magnitud se caracteriza por ser de simetría par y el espectro de fase por ser de simetría impar [3], tal y como se observa en la Figura 1.6.

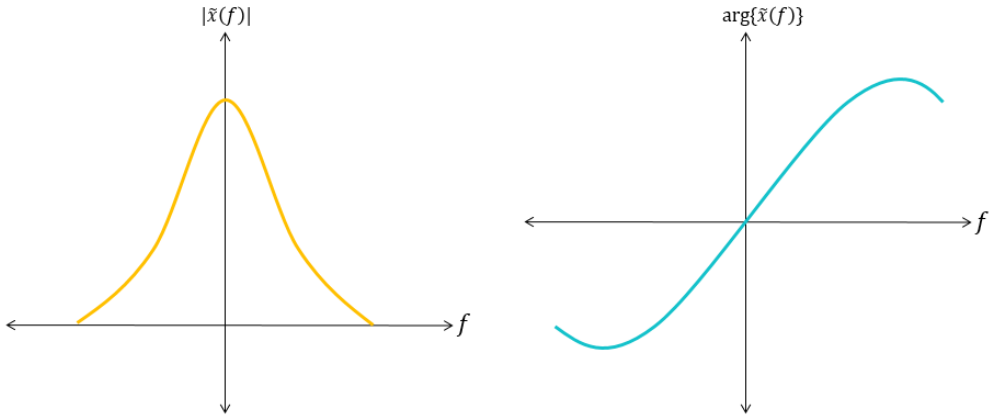


Figura 1.6: *Espectro de magnitud y fase.*

La FT es una función continua que refleja las infinitas componentes de una señal en el dominio de la frecuencia; esta cualidad hace que su aplicación en un entorno computacional, en términos de procesamiento, sea ineficiente, es por ello que para realizar la conversión se recurre a la Transformada Rápida de Fourier (FFT, *Fast Fourier Transform*) y la Transformada Rápida Inversa de Fourier (IFFT, *Inverse Fast Fourier Transform*) que son una versión computacional eficiente de la Transformada Discreta de Fourier (DFT, *Discrete Fourier Transform*).

1. Transformada Discreta de Fourier: permite obtener la representación en la frecuencia, $\tilde{x}[k]$, a partir de la señal discreta en el tiempo $x[n]$. De la Ecuación 1.4 se infiere que al aplicar la DFT se genera una señal en la frecuencia discreta y periódica.

$$\tilde{x}[k] = F\{x[n]\} = \sum_{n=0}^{N-1} x[n] e^{\frac{-j2\pi nk}{N}}. \tag{1.4}$$

2. Transformada Discreta Inversa de Fourier: permite obtener la señal discreta del tiempo, $x[n]$, a partir de la señal discreta en la frecuencia, $\tilde{x}[k]$.

$$x[n] = F^{-1}\{\tilde{x}[k]\} = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{x}[k] e^{\frac{j2\pi nk}{N}}. \tag{1.5}$$

Realizando una asociación con algunos de los procesos desarrollados dentro de un ADC:

- Para la etapa de muestreo, se obtienen muestras de una señal analógica a una frecuencia de F_s, dando como resultado una señal discreta en el tiempo($x_s\,[n]$). Al aplicar la FFT sobre dicha señal se obtiene un espectro periódico, con réplicas espectrales cada F_s, esto es, el espectro de la señal discreta se encuentra condicionado por el periodo entre cada una de las muestras temporales; es por ello que, entre mayor sea la frecuencia de muestreo bajo la cual se ha obtenido la señal discreta, las réplicas espectrales se encuentran más alejadas entre sí, tal y como se presenta en las Figuras 1.7 y 1.8, donde se ha muestreado y aplicado la FFT a una señal de audio con una F_s de $5.51KHz$ y $22.05KHz$, respectivamente. Ésta también es una representación gráfica de la aplicabilidad del teorema de muestreo de Nyquist-Shannon.

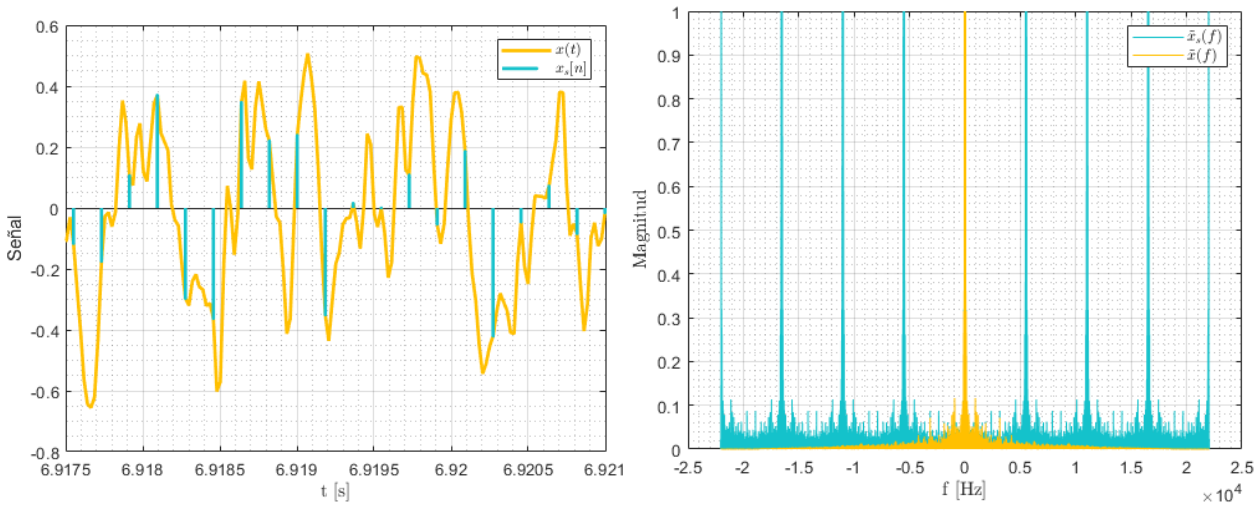


Figura 1.7: *Espectro de magnitud del proceso de muestreo. Frecuencia F_s de $5.51KHz$.*

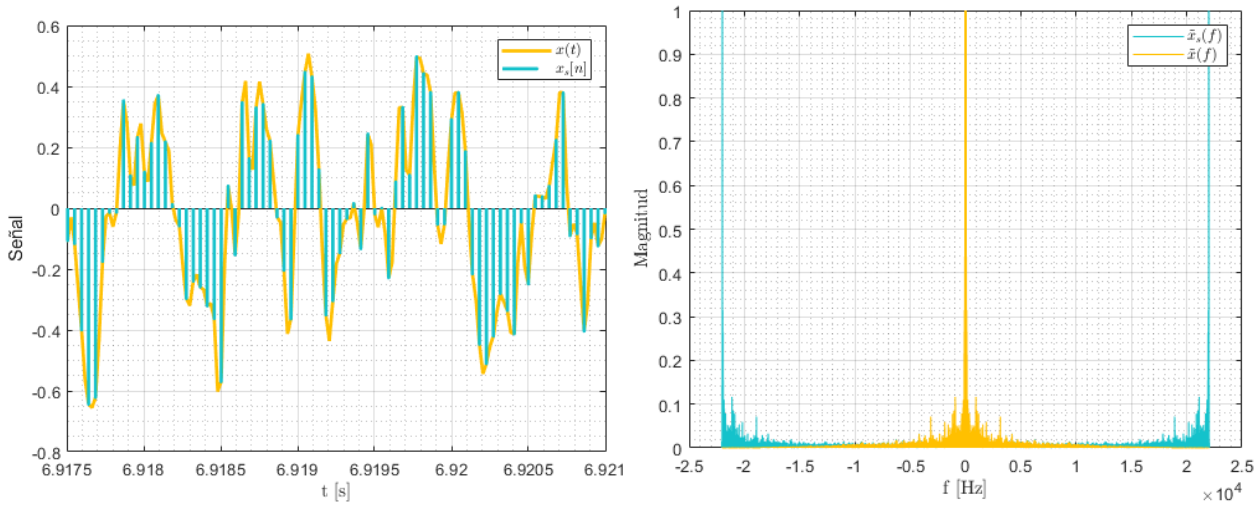


Figura 1.8: *Espectro de magnitud del proceso de muestreo. Frecuencia F_s de $22.05KHz$.*

- Para la etapa de cuantificación, la discretización en magnitud de las muestras temporales de la señal, como se ha mencionado, generan un error de cuantificación, que en el dominio de la frecuencia se puede ver reflejado como una modificación del espectro de la señal, dependiendo del tamaño del alfabeto de cuantificación esta modificación puede ser más o menos notoria. En las Figuras 1.9 y 1.10 se muestran los espectros de la señal antes y después del proceso de cuantificación para un alfabeto dado por una resolución de 2 bits (4 niveles de cuantificación) y 4 bits (16 niveles de cuantificación), respectivamente.

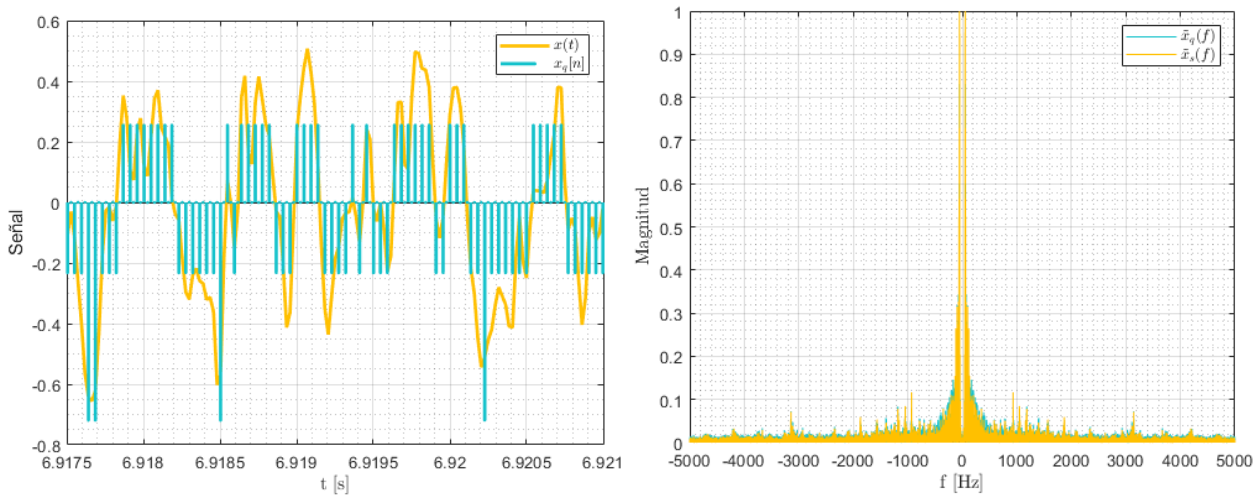


Figura 1.9: *Espectro de magnitud del proceso de cuantificación. 4 niveles de cuantificación.*

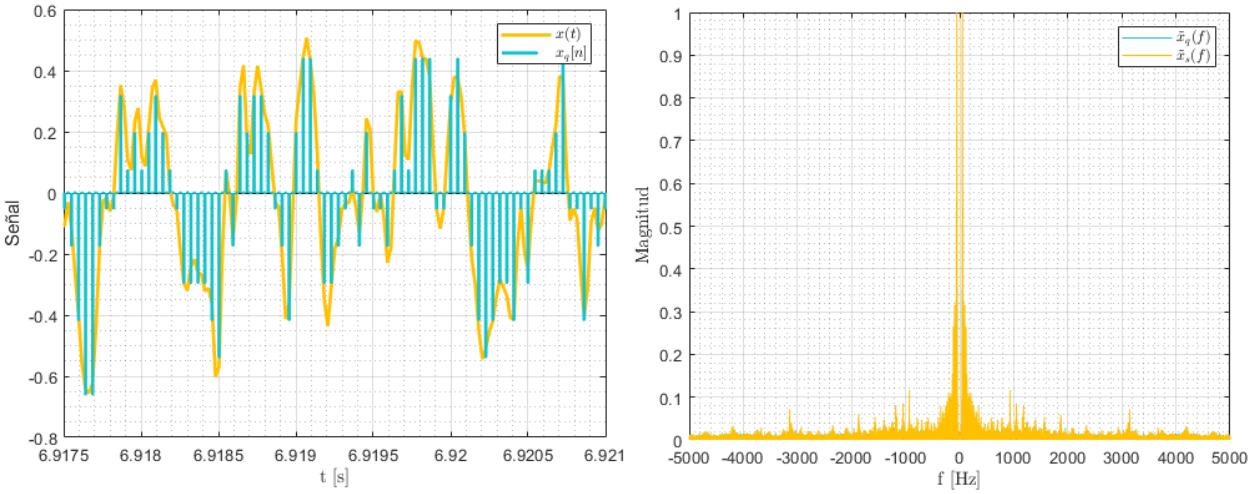


Figura 1.10: *Espectro de magnitud del proceso de cuantificación. 16 niveles de cuantificación.*

Como se ha previsto, entre mayor es el tamaño del alfabeto de cuantificación, menos perceptibles son las alteraciones que provoca el error de cuantificación sobre el espectro de la señal; sin embargo, esto no implica que los espectros, antes y después de la cuantificación, sean idénticos.

1.1.4. Espectro de fase

En la Ecuación 1.3 se observa la representación polar de una señal. Dicha representación muestra que cada componente tiene asociada una fase inicial (φ), que en el plano

complejo, se puede entender como la posición de partida de cada fasor, como se ejemplifica para el caso de 3 componentes en la Figura 1.11.

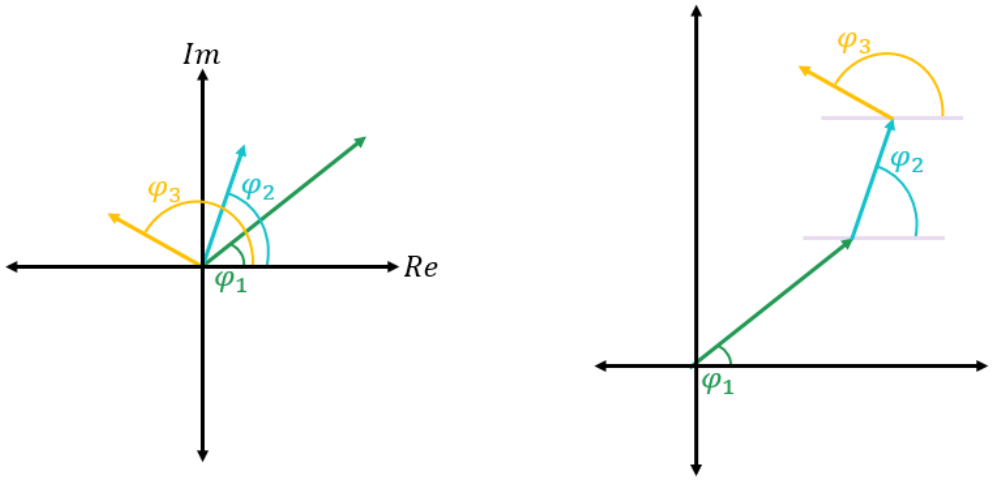


Figura 1.11: *Fase inicial de un conjunto de sinusoides.*

Una sinusoide puede representarse mediante la fórmula de Euler; tal es el caso de la función coseno expuesta en la Ecuación 1.6.

$$\cos(2\pi f_o t + \varphi) = \frac{1}{2}[e^{j(2\pi f_o t+\varphi)} + e^{j[-(2\pi f_o t+\varphi)]}]. \tag{1.6}$$

Es notorio que la sinusoide se puede representar por dos fasores de magnitud de 0.5 y fases opuestas; en este último aspecto se resalta que, al asumir una fase inicial φ se origina una simetría especular (o simetría de reflexión) entre ambos fasores, respecto al eje real del plano complejo. El transcurso del tiempo incentiva el movimiento de ambos fasores en direcciones contrarias, lo cual se puede asociar a un cambio de fase sincronizado, pero de valores opuestos, entre ambos fasores.

Para reconstruir fielmente la sinusoide se necesita de ambos fasores donde, acatando la Ecuación 1.6, cada uno debe tener un argumento de signo opuesto al otro, de lo cual se deduce la necesidad de una simetría impar en el espectro de fase asociado a un tono puro.

Por su lado, al calcular el espectro de una señal a partir de la Ecuación 1.1, se obtiene una magnitud y fase, esto es, la señal puede ser compleja en el dominio de la frecuencia. Matemáticamente se tiene que:

$$F\{x(t)\} = |\tilde{x}(f)|\, e^{-j2\pi\varphi_x(f)} = |\tilde{x}(f)|\cos(-2\pi\varphi_x(f)) + j\,|\tilde{x}(f)|\sin(-2\pi\varphi_x(f)).$$

A partir de esto, la expresión del argumento de la señal compleja (fase) se puede calcular de la forma:

$$arg\{\tilde{x}(f)\} = \arctan\left[\frac{|\tilde{x}(f)|\sin(-2\pi\varphi_x(f))}{|\tilde{x}(f)|\cos(-2\pi\varphi_x(f))}\right] = -2\pi\varphi_x(f),$$

no obstante; en la Ecuación 1.6 se muestra que cada tono tiene que tener un mismo valor de fase, pero con signo opuesto, por lo que la función de las fases se puede definir como:

$$\varphi_x(f) = -\varphi_x(-f),$$

la cual corresponde a la definición de una señal impar y dado que $arg\{\tilde{x}(f)\}$ es una versión escalonada de la función de fases, entonces se puede afirmar que el espectro de fase de una señal real debe tener simetría impar.

1.2. Señales de Audio

1.2.1. Sonido, HAS y audio

El sonido es considerado como el movimiento de moléculas que ejercen fuerzas unas contra otras en un medio elástico; para el caso del ser humano, dicho medio es el aire. El movimiento molecular genera cambios de presión, según la cantidad de moléculas acumuladas en un área del espacio en un instante de tiempo, con lo cual, físicamente, "el sonido es una onda de presión que viaja a través del tiempo y el espacio" [4]. Su velocidad depende de la temperatura y la humedad, dado que estos factores alteran el medio de propagación. Bajo condiciones estándar de presión y humedad, a una temperatura de 20℃, su velocidad es de $343,4m/s$ [5]

El ser humano tiene la capacidad de captar dichas variaciones de presión a través de su HAS, expuesto en la Figura 1.12, el cual se conforma por el oído externo, el oído medio y el oído interno. Así, cuando se escucha un sonido, éste se transmite como una onda y llega al oído externo; las ondas pasan a través del canal auditivo, un pasaje delgado, que conduce al tímpano, quien recibe la onda en forma de choque. Las vibraciones generadas se envían a los huesecillos del oído medio (estribo, yunque, martillo), quienes son los encargados de amplificar dichas vibraciones y enviarlas a la cóclea, provocando un efecto de ondulación que da como resultado la formación de una onda viajera a lo largo de la membrana basilar; de esta forma las células sensoriales presentes en la parte superior de la membrana basilar, llamadas células ciliadas, reconocen las ondas sonoras. Finalmente, las señales de estas oscilaciones son transmitidas por los nervios auditivos al cerebro, y es él quien puede distinguir el tono según la región de la membrana basilar que esté oscilando [6].

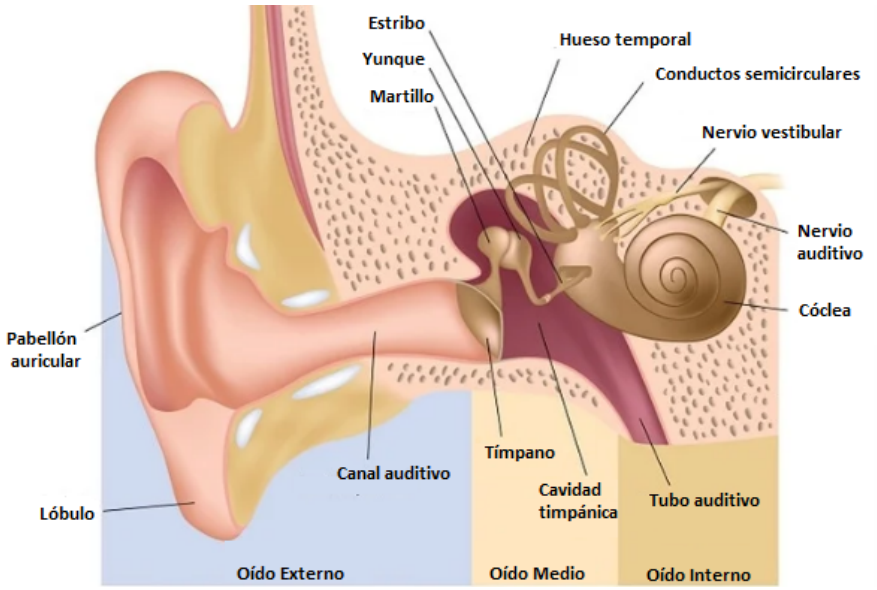


Figura 1.12: *Sistema auditivo humano. Tomada de [6].*

Por su parte, según la Real Academia Española (RAE) [7], el audio hace referencia a la grabación, reproducción y transmisión del sonido. Así, una señal de audio está compuesta por valores de amplitud, que pueden ser continuos o discretos, los cuales, para su transmisión o propagación, se transforman en variaciones de presión que pueden ser captadas por el HAS (señales acústicas). En este sentido, las señales de audio pueden ser de tipo analógico o digital.

1.2.2. Señal de audio analógica

Una señal de audio analógica puede verse como un conjunto de intensidades producidas por el sonido que varían en función del tiempo en forma continua, las cuales, para su manipulación digital, deben ser captadas para su representación eléctrica y sometidas a la ADC; es por ello que se usa el micrófono como dispositivo transductor electroacústico, capaz de representar aquellas variaciones mecánicas en eléctricas [8].

Ahora, el oído humano es capaz de apreciar frecuencias entre de $20Hz$ a $20KHz$; por otra parte, los micrófonos logran la transformación de sonidos en el rango de $10Hz$ a $20KHz$ [9], aunque esto varía según el fabricante, a señales eléctricas también de tipo analógico.

1.2.3. Señal de audio digital

Con el fin de adaptar la señal de audio analógica a un entorno computacional, se implementa un ADC para obtener una aproximación del audio en términos digitales; el audio es representado por un conjunto de datos binarios que facilitan su procesamiento y

almacenamiento. Asimismo, esta conversión puede garantizar una calidad aceptable en comparación a la que brinda el audio analógico, puesto que en la representación digital, si se cumple con el teorema de Nyquist-Shannon, sólo se introduce distorsión durante el proceso de cuantificación.

Con lo anterior, se reconoce que una señal de audio es una representación eléctrica del sonido; sin embargo, dicho sonido puede provenir del habla humana o de un conjunto de elementos, generalmente, instrumentos, por lo que se puede tener una nueva categorización: señales de voz o de música.

1.2.4. Señales de voz

En el espectro de una señal de voz es posible reconocer la frecuencia fundamental y los armónicos de ésta. Dicha frecuencia fundamental permite diferenciar a la persona que está hablando, dado que su valor depende de las características anatómicas de cada individuo.

> "La laringe humana es capaz de producir una amplia gama de frecuencias (rango vocal), que varían en función de la edad y del sexo. Los valores normales de frecuencia fundamental son de unos $125Hz$ para el hombre, $250Hz$ para la mujer y $350Hz$ en la infancia [10]".

Los armónicos de la señales de voz se distribuyen a lo largo del espectro, además de otras componentes resultantes de la intensidad, timbre, vocalización y extensión del habla; como ejemplo de ello, en la Figura 1.13 se observa el espectro de una señal de voz, donde la frecuencia fundamental del emisor es de $247Hz$. Sin embargo, el espectro "tiende a decaer de manera significativa por encima de los $4KHz$ para los sonidos sonoros, y por encima de los $8KHz$ para los sonidos sordos" [10].

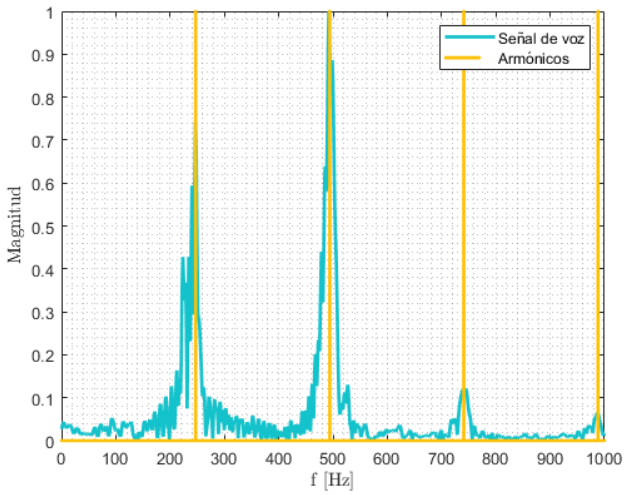


Figura 1.13: *Armónicos de una señal de voz.*

En el ámbito de las telecomunicaciones, el ancho de banda disponible en un sistema para la transmisión de información es limitado, por lo que se debe restringir el rango de frecuencias que componen a las señales de voz. A lo largo de la evolución de la telefonía, se han definido diferentes valores para este ancho de banda; inicialmente para una señal de voz de calidad aceptable (poca distorsión) se consideraba una frecuencia máxima de $3.4KHz$, algunas veces denominada como banda estrecha [11]. Garantizando bandas de transición, el espectro se considera hasta los $4KHz$, con lo que la frecuencia de muestreo mínima a utilizar, dado por el teorema de Nyquist-Shannon, es de $8KHz$. Tomando como referencia el codificador G.711 [12], que usa una resolución de 8 bits, el flujo de datos resultante de la señal de voz es de $64Kbps$.

A lo largo de los años, los requerimientos han presentado variaciones, dando paso a codificadores como, e.g., el G.723.1 que cuenta con un resolución de 16 bits; diseñado bajo principios de codificación de predicción lineal y dos velocidades binarias de $5,3Kbps$ y $6,3Kbps$ para la transmisión de señales de voz o audio como un servicio multimedia [13]; y el codificador G.726 que cuenta con resoluciones de 2, 3, 4 y 5 bits que generan flujos de datos variables ($16Kbps$, $24Kbps$, $32Kbp$s y $40Kbps$) para una frecuencia de muestreo de $8KHz$ en transmisiones de voz [14].

1.2.5. Señales de música

Las señales de música están compuestas por una gran variedad de elementos que incluyen instrumentos, voces y efectos de sonido. Cada uno de estos factores aporta su propio espectro de frecuencia que, al combinarse, generan un espectro de frecuencia mucho más amplio y complejo comparado con el de una señal de voz.

La representación digital de este tipo de señales inició con el formato de los CD, los cuales codifican información utilizando una resolución de 16 bits y una frecuencia de muestreo de $44.1KHz$, debido a que el rango de frecuencias audible por el HAS va hasta los $20KHz$. A partir de este precedente se han propuesto algoritmos que aprovechan el incremento en el número de niveles de cuantificación, la frecuencia de muestreo, o, incluso, ambos parámetros a la vez, para que la versión digital de una señal analógica sea más exacta, con lo cual surge el audio de alta definición, en donde los formatos de audio cuentan con una resolución de 24 bits y $96KHz$ [15].

La diferencia más evidente entre estas dos categorías radica en el ancho de banda usado para su representación, tal y como se muestra en la Figura 1.14.

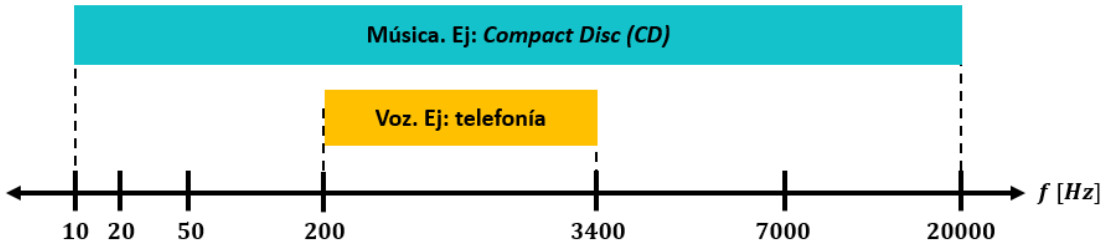


Figura 1.14: *Ancho de banda de audio de voz y música. Adaptado de [5].*

1.2.6. Formatos de archivo de audio digital

Un formato de audio es un contenedor multimedia que guarda una grabación de audio digital; cada uno de éstos se diferencia por medio de una extensión específica del archivo. Cabe resaltar que existe una diversidad de formatos de audio que cuentan con características propias y se clasifican de acuerdo a la compresión del audio, es decir, formatos de audio sin comprimir y formatos de audio con compresión[1], que a su vez pueden o no introducir pérdidas.

La Modulación por Impulsos Codificados (PCM, *Pulse Code Modulation*) es la técnica que soporta los inicios de la conversión de señales de audio analógico a digital. Ésta se encuentra dentro de los formatos sin compresión y trabaja con una frecuencia de muestreo de $44.1KHz$, a una tasa de bits de $1411Kbps$ [16], por lo cual los archivos de audio resultantes son de buena calidad. Dentro de esta categoría se encuentran los formatos: *WAV*, *AIFF*, *SU*, *AU* y *RAW*; sin embargo, cabe mencionar que este tipo de archivos sin comprimir son de gran tamaño, cerca de 10 Mega Bytes por cada minuto de audio.

Por otro parte, dentro de los archivos comprimidos con pérdidas se encuentran los formatos *MP3*, *AAC*, *Ogg* y *WMA*; en los cuales el sistema de codificación comprime los datos suprimiendo parte de ellos mediante las tasas de bits, e.g.: la compresión de *MP3* utiliza tasas de bits de $320Kbps$, $128Kbps$ o $96Kbps$ intentando minimizar la cantidad de datos que mantiene el archivo, reduciendo a su vez el peso y por tanto su calidad. Sin embargo, en esta reducción sólo se pierden canales mínimamente audibles por el HAS y, por ende, se mantiene gran parte de su fidelidad, es por ello que este tipo de formatos comprimidos con pérdidas suelen ser los más usados, ya que minimizan almacenamiento y cuentan con una calidad suficiente para garantizar una buena reproducción [17]. De forma comparativa, este tipo de archivos comprimidos suelen ser hasta 10 veces más

[1]Para la clasificación de los archivos de audio se considera sin compresión cuando no se realizan procesos adicionales a los del ADC básico para reducir su tamaño; la compresión con pérdidas se presenta cuando se realizan procesos complementarios para reducir el tamaño del archivo, generando pérdida de información; finalmente, los archivos de compresión sin pérdidas son aquellos en los que se realizan procesos adicionales para reducir el tamaño del archivo, sin que esto conlleve una pérdida relevante de información. Por lo general, los archivos de compresión con pérdidas logran tasas de compresión más altas en comparación con la compresión sin pérdidas.

pequeños que los formatos de archivo sin comprimir [18].

Finalmente, se encuentran los formatos de audio comprimidos sin pérdidas: *FLAC*, *MPEG-4* y *TTA*; en los cuales se eliminan la mayoría de silencios del archivo llevándolos a ser casi nulos, reservando más espacio para los demás componentes del audio, logrando así una reducción en el tamaño del archivo sin pérdida alguna de información, alcanzando tamaños entre cinco y diez veces más que los formatos comprimidos con pérdidas [17].

En la Figura 1.15 se muestra una comparación del peso entre los tres tipos de formatos de audio existentes, tomando en consideración 16 bits de resolución, a una frecuencia de muestreo de $44.1KHz$.

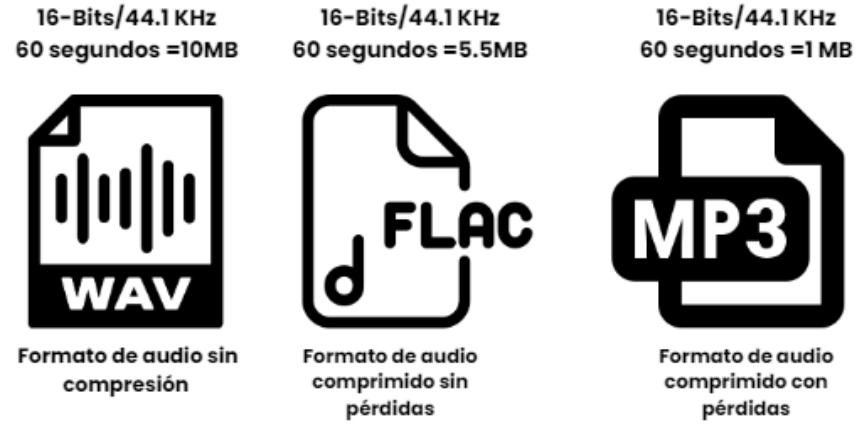


Figura 1.15: *Características de los formatos de archivo de audio. Adaptado de [17]*.

CAPÍTULO 2

ESTEGANOGRAFÍA

2.1. Definición de Esteganografía

El término esteganografía proveniente del griego *steganos* que significa encubierto y *graphos* que significa escritura, por lo que hace referencia al estudio y aplicación de técnicas de procesamiento digital dedicadas al ocultamiento de información confidencial utilizando archivos portada, ya sean audio o imágenes, de manera tal que ésta no pueda ser detectada por alguien diferente a quien transmite y a quien recibe el mensaje [19]. Dentro del proceso esteganográfico se resaltan ciertos elementos que se mencionan a continuación:

- **Archivo de portada**: Es la entidad que se emplea para ocultar un mensaje confidencial de un observador indeseado que pretenda interferir, de manera inadecuada, en una comunicación.
- **Información secreta**: Es el mensaje confidencial que se desea ocultar.
- **Archivo *stego***: Corresponde al archivo de portada al cual se le inserta la información secreta a partir de una técnica esteganográfica, con el fin de ocultar el mensaje para evitar su detección por parte de un observador o intruso que desee interceptar la información confidencial.

2.2. Propiedades de la Esteganografía

Para que el proceso de incrustación de datos digitales en una técnica esteganográfica se considere ideal es necesario cumplir con: maximizar el número de datos ocultos, minimizar la probabilidad de detección del mensaje por destinos no autorizados y garantizar su supervivencia tras el procesamiento digital [20]; surgiendo así tres características fundamentales, las cuales se muestran en la Figura 2.1 y se definen en mayor detalle en los siguientes items.

- **Robustez**: se encuentra relacionada con la supervivencia del mensaje secreto ante los procesamientos propios de una red o externos a la misma, del archivo *stego*, dentro de las cuales se encuentra: adición de ruido, re-muestreo, compresión, codificación, entre otras; y al nivel de inaccesibilidad que se puede establecer si un observador indeseado sospecha de la existencia de información oculta en el archivo *stego*.

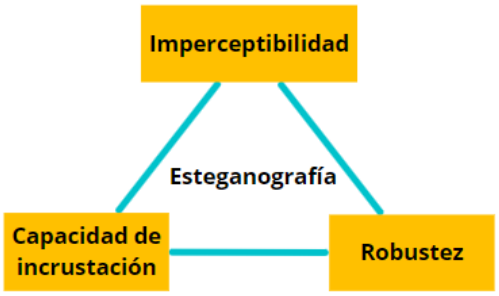


Figura 2.1: *Características ideales de un proceso esteganográfico.*

- **Imperceptibilidad**: esta propiedad de la esteganografía hace referencia a la similitud perceptiva entre la señal original antes del proceso de incrustación y la señal final que contiene la información confidencial, es decir, que tan imperceptibles son las modificaciones que contiene el archivo *stego*.

- **Capacidad de incrustación**: se entiende como la cantidad de información del mensaje secreto que se guarda dentro del archivo portada, para lo cual se comparan los tamaños de ambos archivos. Idealmente, se busca que las diferencias entre el archivo portada y el archivo *stego* no sean perceptibles; sin embargo, entre mayor sea la información incrustada, las dimensiones de los archivos tienden a ser similares, lo que podría hacer evidente estas modificaciones.

2.3. Esteganografía de Audio

Las señales de audio digital se consideran excelentes transportadoras de información oculta debido a la alta correlación entre muestras consecutivas y a que, para el HAS, algunas alteraciones sobre las señales de audio pasan inadvertidas. Es por esto que la esteganografía de audio se basa en el uso de un archivo de audio portada para el incrustamiento de información confidencial, la cual, en este caso, también es un archivo de audio. En la Figura 2.2 se muestra el proceso de esteganografía llevado a cabo sobre una señal de audio, a partir de un diagrama de bloques.

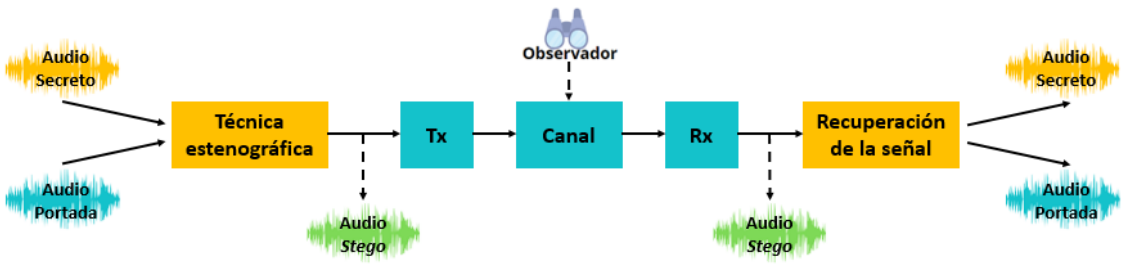


Figura 2.2: *Proceso esteganográfico basado en audio.*

El objetivo fundamental de la esteganografía es cumplir con las características mencionadas en la Figura 2.1; sin embargo, las técnicas de mayor capacidad de incrustación son las menos robustas o imperceptibles, así mismo, las técnicas más robustas tienden a ser las de menor capacidad de incrustación. Es por esto que, en el intento de encontrar armonía entre las tres propiedades, varios autores han propuesto diversas técnicas esteganográficas basadas en las cualidades físicas de las señales, así como también, en los diferentes procesos a los que éstas se ven sometidas antes de su transmisión. Tomando en consideración esta premisa, se presentan técnicas en:

1. El dominio temporal, tales como: *low bit encoding*, en el cual se sustituyen los Bits Menos Significativos (LSB, *Least Significant Bit*) del audio portada por bits que representan el mensaje oculto; *echo hiding*, en el que se adicionan componentes de eco valiéndose de su retardo, amplitud inicial y decaimiento para la representación del mensaje; y *hiding in silence intervals* en donde se modifica el número de muestras que representan los intervalos de silencio para llevar a cabo el ocultamiento de los datos [19].

2. En el dominio transformado, tales como: *spread spectrum*, en el cual se inserta la información secreta en el espectro de la señal portada, replicándolo e implementando un código que sólo reconoce el transmisor y el receptor del mensaje; *discrete wavelet transform* donde se realiza la transformación del audio portada al dominio de las *wavelets* y se ejecuta el proceso de incrustación; *tones insertion*, en donde se insertan tonos que representan el mensaje cerca a las componentes de mayor potencia del audio portada, generando el solapamiento de dichos tonos; *amplitude coding*, que modifica la magnitud de algunas componentes buscando ocasionar la mínima distorsión posible; *cepstral domain*, en el cual el audio portada se lleva al dominio Cepstral, sometiendolo a la FT, tomando el logaritmo de su espectro e implementando su transformada inversa, para luego incrustar la información secreta y, concluida la inserción, se realiza el proceso opuesto; y *phase coding*, la cual se fundamenta en el procesamiento digital de la fase de las componentes del audio portada para representar el mensaje, sobresaliendo por ser uno de los métodos de codificación más eficaces; en términos de la relación señal a ruido percibida [19].

En cuanto al lugar de implementación de algunas de las técnicas de esteganografía anteriormente indicadas, surge una clasificación en términos del dominio codificado: *in-encoder or post-encoder* [19], donde se realiza el ocultamiento de información y extracción del mensaje, antes o después del codificador del transmisor y decodificador del receptor, respectivamente. Cada una de las técnicas enumeradas se diferencian, además del proceso de implementación, en sus proporciones de robustez, su capacidad de incrustación y su imperceptibilidad. De ellas se destaca la codificación de fase (*phase coding*) en el dominio de la frecuencia, sobre la cual los autores en [21] resaltan la alta robustez e imperceptibilidad variable que ha logrado este método, condicionando una baja capacidad de incrustación.

2.4. Técnica de Codificación de Fase

La codificación de fase consiste en realizar cambios de fase a un determinado segmento de audio, bajo la premisa de que las modificaciones sobre las componentes del espectro de fase pasan desapercibidas por el HAS. En esta técnica se incrustan los bits del mensaje como desplazamientos de fase de las componentes de la señal original (audio portada), con el objetivo de obtener una codificación inaudible; sin embargo, la variación entre las fases no debe ser demasiado acentuada, dado que si los cambios son muy grandes estas distorsiones pueden ser perceptibles por el HAS dejando al descubierto el audio *stego*, especialmente cuando se emplean componentes de baja frecuencia, donde el HAS es más sensible a las alteraciones de la señal [22]. Ahora bien, esta técnica instaura un alfabeto finito que indica los valores de fase con los que se debe modificar la fase de las componentes de frecuencia del audio portada para representar los bits de información encubierta "1" y "0", en este caso, $\frac{-\pi}{2}$ y $\frac{\pi}{2}$ respectivamente [23]. Además, se establece una fase de referencia para lograr extraer el mensaje secreto en el receptor sin ningún tipo de distorsión.

No obstante, usar dos alteraciones de fase conlleva a una baja capacidad de incrustación, dado que se representa un único bit por medio de la modificación de fase de cada componente; por otro lado, depender de una fase de referencia implica que, en caso de que ocurran inconvenientes internos en el receptor, se pierda esta fase y no sea posible recuperar fielmente el audio secreto.

Algunos autores han propuesto variantes o mejoras de la codificación de fase tradicional, como la incrustación de datos en los bloques temporales del audio portada mediante la modificación total o parcial del espectro de fase de la señal, bajo la premisa de que el HAS no capta adecuadamente pequeñas y medianas variaciones de fase, especialmente en componentes de alta frecuencia del espectro audible humano [22]. Asimismo, se aplican ciertos criterios de adaptabilidad en el alfabeto usado para la representación de la información secreta como el presentado en [21], donde, según la longitud del audio portada y las dimensiones de la información secreta, se aplica un nivel adecuado de codificación de fase, esto es, se establece un alfabeto de desfases, en el que cada elemento de dicho alfabeto representa un número determinado de bits que componen el mensaje, e.g., los bits 101 se representan mediante un desfase de $\frac{-5\pi}{8}$. También surgen variantes en cuanto a la fase de referencia, que puede ser establecida en el primer bloque temporal de la señal o a lo largo de todo el audio portada; para este último caso, los datos son representados en términos de la diferencia de fase entre dos componentes de frecuencia adyacentes [24].

Aunque los autores aplican la codificación de fase bajo distintas consideraciones, éstos concuerdan con los pasos de ejecución expuestos en la Figura 2.3 y listados a continuación:

a) Segmentar el archivo de audio en bloques temporales del orden de los milisegundos.

b) Aplicar la FFT a cada bloque de forma individual.

c) Modificar la fase de las componentes de frecuencia de la señal dependiendo del criterio seleccionado para la representación del mensaje secreto.

d) Implementar la IFFT a cada bloque.

e) Ensamblar los bloques y así obtener el audio *stego*.

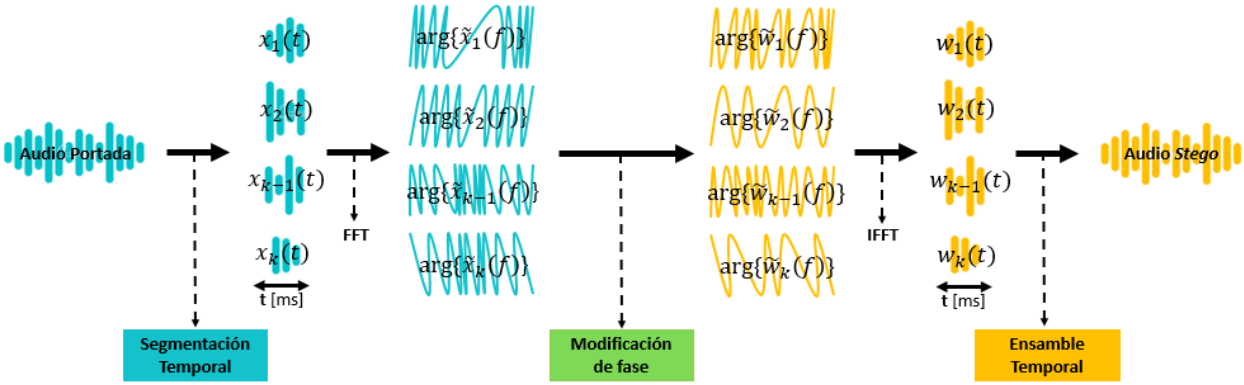


Figura 2.3: *Pasos generales de ejecución de la técnica de codificación de fase.*

2.5. Métricas

La efectividad de una técnica esteganográfica basada en audio puede evaluarse a partir de una amplia variedad de métricas, éstas pueden ser de carácter cuantitativo (objetivas) o cualitativo (subjetivas); es por ello que para el desarrollo de este trabajo de grado se implementan medidas objetivas y subjetivas, permitiendo realizar comparaciones entre las variaciones efectuadas sobre la propuesta inicial del algoritmo y la percepción del HAS, para obtener así una evaluación más completa. En esta sección, se presentan en detalle las métricas de evaluación implementadas; en el Apéndice A se describen algunas consideraciones asociadas a su elección.

- **BER**: La Tasa de Error de Bit (BER, *Bit Error Ratio*) se define como el número de bits recibidos de forma incorrecta al extraer el mensaje confidencial respecto al total de bit codificados; su rango de valores está dado entre 0 y 0.5, puesto que los valores mayores a este se consideran imágenes especulares de valores inferiores a 0.5, es decir, un valor de BER igual al 0.8 es equivalente a una BER de 0.2, y un valor de BER igual a 1 se considera perfecto, ya que los datos recibidos son simplemente

el inverso de los datos transmitidos. Asimismo, cabe mencionar que 0 indica nula presencia de errores y 0.5 la máxima incertidumbre sobre el mensaje. Esta métrica permite analizar la robustez del audio *stego* frente a posibles distorsiones en el canal de comunicación, así, entre menor sea el valor de la BER, mayor robustez presenta el audio *stego*. La expresión matemática para calcular la BER se muestra en la Ecuación 2.1 [25].

$$BER = \frac{\textit{Bits erróneos}}{\textit{Bits totales}}. \tag{2.1}$$

- **SNR**: La Relación Señal a Ruido (SNR, *Signal to Noise Ratio*), se define como la proporción existente entre la potencia de la señal que se transmite y la potencia del ruido que la altera, evaluando así, en este caso, la diferencia que existe entre el audio portada y el audio *stego* y, por tanto, la imperceptibilidad; su rango se puede medir en decibelios (*dB*) y toma valores positivos cuando la potencia de la señal es mayor que la del ruido, o negativos cuando la potencia del ruido es mayor a la potencia de la señal. Esta medida indica la cantidad de distorsión generada por los datos que conforman el mensaje confidencial incrustado en el audio *stego*; de esta manera, en cuanto mayor es el valor de la SNR, mayor imperceptibilidad tiene el audio *stego*. La expresión para calcular la SNR se muestra en la Ecuación 2.2, donde $x[n]$ representa la señal de audio portada y $w[n]$ la señal de audio *stego* [26].

$$SNR = \log_{10}\left[\frac{\sum x^2[n]}{\sum[x[n] - w[n]]^2}\right]. \tag{2.2}$$

- **SSIM**: El Índice de Similitud Estructural (SSIM, *Structural Similarity Index*), es una técnica perceptual que se basa en modelos de percepción humana, con la cual es posible evaluar la diferencia entre una señal original (audio portada) y su versión modificada (audio *stego*), a partir de tres medidas: luminosidad, contraste y estructura. La luminosidad hace referencia a una comparación de los valores medios de la señal; el contraste es la diferencia entre los valores más altos y más bajos de la señal; la estructura es la forma en que los valores de amplitud de la señal cambian a lo largo del tiempo, calculada a partir del coeficiente de correlación entre la señal original y la señal modificada, de esta manera los tres componentes se combinan para obtener una medida general de similutud entre las dos señales de audio [27]. El rango de valores de SSIM está establecido entre $[-1,1]$, en donde 1 indica que las dos señales son idénticas y -1 significa que las dos señales son totalmente opuestas en términos de estructura.

- **PEAQ**: La Evaluación Perceptual de la Calidad de Audio (PEAQ, *Perceptual Evaluation of Audio Quality*) es un método objetivo de evaluación para la calidad de audio publicado por la ITU-R BS.1387 en 2001 [28], con el propósito de brindar un método objetivo más adecuado que los tradicionales (medidas de SNR y distorsión) para medir la calidad de audio percibida de sistemas que emplean procesamiento analógico o digital de la señal [29].

  La PEAQ toma una señal de referencia y su posible versión degradada, las compara mediante un modelo de oído periférico y una etapa de extracción de características [28]. A continuación, calcula algunos índices de degradación objetivos relacionados con la calidad percibida de la señal, los cuales toman el nombre de Valores de Salida del Modelo (MOV, Model Output Values). La PEAQ combina múltiples MOV y se asignan a una escala de calidad objetiva denominada Grado de Diferencia Objetiva (ODG, *Objective Difference Grade*) que establece una descripción al deterioro de la calidad indicado en la Tabla 2.1 [28]. Cabe resaltar que la PEAQ se fundamenta en las recomendaciones *UIT-R BS.1116* para pruebas subjetivas [30].

  Tabla 2.1: Calificaciones del deterioro de la calidad usadas en PEAQ-ODG.

Calificación	**El deterioro de la calidad es**
0	Imperceptible
-1	Perceptible pero no molesto
-2	Ligeramente molesto
-3	Irritante
-4	Muy molesto

  Bajo la recomendación de la ITU, algunos autores han implementado el método para su uso y el de la comunidad; tal es el caso de Kabal quien en 2002 interpreta la norma y en 2004 habilita un repositorio público con su implementación en los lenguajes de python y MATLAB [31]. Por último, en 2023 se aprueba la recomendación *ITU-R BS.1387-2* que brinda mejoras para la precisión y optimización del método [32], aunque su reciente publicación conlleva a la espera de su aplicación por parte del sector de comunicaciones.

- **MOS**: La Puntuación de Opinión Media (MOS, *Mean Opinion Score*), es una métrica de evaluación subjetiva de calidad de audio, basada en la media aritmética de las calificaciones individuales realizadas por un grupo de personas; la puntuación brindada por una persona se encuentra en una escala de 1 a 5 (ver Tabla 2.2).

Tabla 2.2: Puntuación usada en MOS.

Calificación	La calidad del audio es
5	Excelente
4	Buena
3	Aceptable
2	Pobre
1	Mala

De está forma cada evaluador valora distintos audios y se toma el promedio de sus calificaciones, por tanto, si se tienen n evaluaciones de un mismo audio se tiene que [33]:

$$MOS = \frac{1}{n}\sum_{i=1}^{n} OS_i, \tag{2.3}$$

donde OS_i representa la opinión con las que se ha valorado el audio.

- **CMOS**: La Puntuación Media de Opinión de comparación (CMOS, *Comparison Mean Opinion Score*), es una métrica de evaluación subjetiva para evaluar la calidad de audio en función de la percepción de los oyentes. Esta métrica se basa en la realización de juicios comparativos entre dos señales de audio, con el objetivo de determinar cuál de ellas presenta mayor calidad y en qué medida, para ello se establecen dos preguntas: *¿Cuál de las dos señales tiene una mejor calidad?* y *¿En qué medida es mejor?* [1]. En la Tabla 2.3 se presenta la calificaciones de comparación categórica usadas en CMOS.

Tabla 2.3: Calificaciones de comparación usadas en CMOS. Adaptado de [1].

Calificación	La calidad del segundo estímulo comparado con el primero es
5	Mucho mejor
4	Mejor
3	Aproximadamente igual
2	Peor
1	Mucho peor

CAPÍTULO 3

DISEÑO DEL ALGORITMO

3.1. Metodología

El presente trabajo de grado sigue una adaptación de la metodología propuesta por Hevner en 2007 [34], en la cual se establecen tres ciclos de investigación (ver Figura 3.1): relevancia, diseño y rigor; los cuales permiten enmarcar la investigación en un contexto específico, utilizando la base de conocimiento de la disciplina, según sea necesario. Basándose en ello, en este capítulo se detalla todo lo relacionado con el ciclo de diseño a partir de los requerimientos y las necesidades del proyecto.

Figura 3.1: *Metodología del trabajo de grado.*

3.2. Requerimientos

Los principales requerimientos[2] que permiten cumplir con los objetivos del proyecto se enuncian a continuación.

3.2.1. Requerimientos funcionales

- El algoritmo debe aceptar audios con formato de compresión *WAV* para usarlos como audios portada y archivos en formato *WAV* o *MP3* para audios secretos.

[2]Los requerimientos se clasifican en funcionales y no funcionales; los requerimientos funcionales hacen referencia a las funcionalidades que deberá contar o acciones que podrá realizar un producto final. Por otra parte, los requerimientos no funcionales corresponden a requisitos que debe cumplir el producto final en cuanto a características intrínsecas como la fiabilidad, tiempo, seguridad, entre otros [35].

- El algoritmo debe permitir la incrustación de una señal de audio sobre otra señal de audio.

- El producto del algoritmo (audio *stego*) debe permitir variar la capacidad de incrustación de información, donde la cantidad máxima depende del método utilizado.

- El algoritmo debe permitir la adecuada extracción de la información secreta en recepción mediante la alineación de los parámetros entre el transmisor y el receptor.

3.2.2. Requerimientos no funcionales

- Las señales de información a utilizar deben permitir su representación como una secuencia binaria.

- El algoritmo debe garantizar la correcta transformación de las señales de audio al dominio de la frecuencia y su proceso inverso.

- La frecuencia de muestreo, Fs, del archivo de audio *stego* generado debe ser igual a la Fs del archivo de audio portada.

- Se debe validar el tamaño del audio secreto con relación al tamaño del audio portada, en función del método de incrustación.

- Los procesos de codificación y decodificación de fuente deben estar sincronizados según los niveles del cuantificador utilizado en el transmisor.

- La implementación del algoritmo y las pruebas para evaluar su desempeño tienen lugar en el entorno de simulación MATLAB.

3.3. Estructura General del Algoritmo

A partir de los lineamientos citados en el Capítulo 2, en la sección 2.4, y los requerimientos expuestos anteriormente, se construyen los diagramas de bloques mostrados en las Figuras 3.2 y 3.3, donde se representa gráficamente el funcionamiento general del algoritmo esteganográfico, esto es, la incrustación y extracción de la información secreta.

En la Figura 3.2, la mayoría de los bloques corresponden a procedimientos obligatorios que se deben llevar a cabo en el transmisor, independientemente del *Proceso de Incrustación* implementado. Además, se resalta el ingreso de los audios portada y secreto con frecuencias de muestreo propias de su creación (Fs y Fss, respectivamente) y la respuesta del algoritmo en este punto que corresponde al audio *stego*, caracterizado por una frecuencia de muestreo idéntica a la del audio portada.

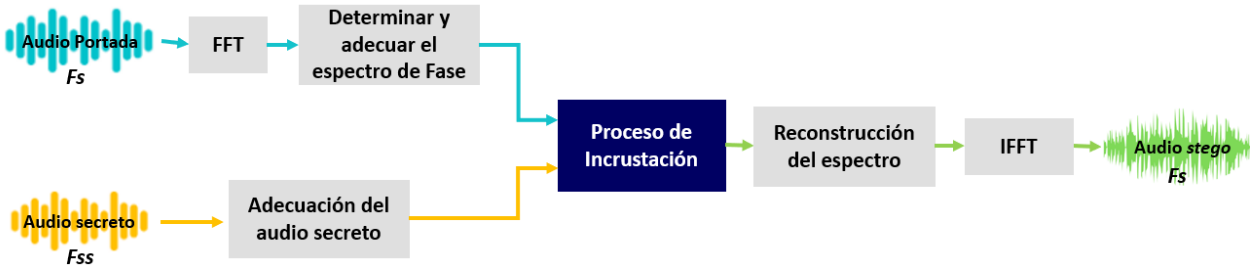


Figura 3.2: *Diagrama general del funcionamiento del algoritmo en el Transmisor.*

A continuación se explica el funcionamiento de los bloques invariantes que conforman el algoritmo implementado por el transmisor del mensaje para generar el audio *stego*.

- **FFT**: El audio portada ingresa a este bloque donde se aplica la FFT, obteniendo su equivalente complejo en el dominio de la frecuencia.
- **Determinar y adecuar el espectro de fase**: Partiendo de la señal compleja obtenida del bloque FFT, se determina su espectro de fase, el cual se somete a una adecuación dependiente del proceso de incrustación que se va a ejecutar.
- **Adecuación del audio secreto**: El audio secreto se debe someter a una adecuación que depende del proceso de incrustación del cual será partícipe; es por ello que su adaptación puede ser en términos de una limitación temporal o compresión (cuantificación y codificación de fuente).
- **Reconstrucción del espectro**: Después de la incrustación de la información en el espectro de fase, se reconstruye una aproximación de la señal compleja en la frecuencia mediante el espectro de magnitud original y el espectro de fase modificado, dando paso al espectro total del audio *stego*.
- **IFFT**: Se realiza la transformación del espectro del audio *stego* del dominio de la frecuencia al dominio temporal mediante la IFFT, dando fin a los procedimientos llevados a cabo en el Transmisor.

Por su lado, en el receptor, el algoritmo debe realizar la extracción del audio secreto, mediante los procedimientos expuestos en la Figura 3.3.

Figura 3.3: *Diagrama general del funcionamiento del algoritmo en el Receptor.*

Los primeros dos bloques, es decir, el cálculo de la FFT y determinación del espectro de fase, se ejecutan de forma similar a lo descrito anteriormente en el transmisor. Por su parte, el bloque posterior al *Proceso de Extracción* desarrolla la siguiente función:

- **Reconstrucción del audio secreto**: Dependiendo del proceso de incrustación implementado en el transmisor, el audio secreto pudo ser sometido a una adecuación. En este bloque se ejecuta el procedimiento contrario, con el fin de retornar sus características a un estado muy próximo al original[3], adaptando el audio secreto para su correcta reproducción.

En las figuras anteriores, Figuras 3.2 y 3.3, se destacan los bloques de *Proceso de Incrustación* y *Proceso de Extracción*, dado que su implementación puede variar según el enfoque que se desee emplear para la inserción del audio secreto. Partiendo de la información citada en las secciones 2.3 y 2.4 referente a las técnicas esteganográficas usadas en el dominio temporal y los procesos de incrustación centrados en la técnica de codificación de fase, se proponen dos enfoques, para los cuales se adaptan los diagramas de bloques generales.

3.3.1. Incrustación en el espectro de fase mediante la técnica *low bit encoding* condicionada al dominio transformado

Aunque la técnica *low bit encoding* es comúnmente implementada en el dominio temporal, es posible adaptarla al dominio transformado mediante la cuantificación y codificación del espectro, en este caso, de fase. Por su lado, también se debe cuantificar, codificar y generar la secuencia de bits que representan al audio secreto. A partir de las respectivas representaciones binarias, se realiza la modificación de los bits menos significativos de las muestras del espectro de fase codificadas en función de los bits del audio secreto.

Como se puede percibir, en este enfoque se establecen ciertos acondicionamientos para ejecutar los procesos de incrustación y extracción, los cuales se exponen en los diagramas de las Figuras 3.4 y 3.5.

Figura 3.4: *Diagrama general de algoritmo. Incrustación mediante low bit encoding.*

[3]Un archivo de audio digital puede contar con uno o dos canales de información (mono o estéreo); el algoritmo garantiza que el audio sea de tipo mono. Por otro lado, dependiendo del proceso de incrustación, el audio secreto es sometido a cuantificación y codificación de fuente, por lo que si la resolución aplicada no coincide con la usada en la creación del audio (por un sistema ajeno al objeto de este trabajo de grado), se puede generar distorsión. Adicionalmente, dependiendo del número de bits a incrustar, y el proceso de incrustación, puede ser necesario complementar su longitud con la concatenación de bits aleatorios.

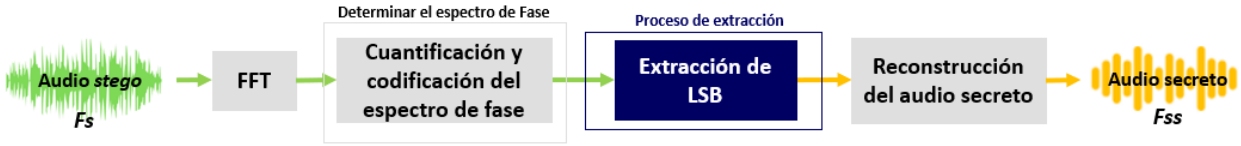


Figura 3.5: *Diagrama general de algoritmo. Extracción mediante low bit encoding.*

Para implementar *low bit encoding* se debe especificar el número de bits menos significativos (LSB) a modificar por cada muestra, para ello, se estipula:

$$T_{Bp} = (L_p)(B_p), \tag{3.1}$$

$$T_{Bs} = (L_s)(B_s), \tag{3.2}$$

donde:

T_{Bp}, T_{Bs}: número total de bits que componen el espectro de fase del audio portada (p) o el audio secreto (s) en el dominio del tiempo.

L_p, L_s: número de muestras que representan el espectro de fase del audio portada y del audio secreto en el dominio del tiempo.

B_p, B_s: número de bits que representan una de las muestras asociadas al audio portada y al audio secreto [4].

De las Ecuaciones 3.1 y 3.2, se deduce que:

$$n_{LSB} = \left\lceil \frac{T_{Bs}}{L_p} \right\rceil = \left\lceil \frac{L_s B_s}{L_p} \right\rceil, \tag{3.3}$$

donde, $n_{LSB} < B_p$ y $L_s B_s < L_p B_p$, siendo n_{LSB} el número de bits menos significativos por muestra disponibles para la incrustación.

- **Proceso de incrustación mediante *low bit encoding***: Aunque el proceso de incrustación se base en *low bit encoding*, éste puede adaptarse según la banda de frecuencias a modificar y la cantidad de bits por muestra que se van a alterar para insertar el audio secreto; es por esta razón que se proponen cuatro métodos enfocados en: incrustar la misma cantidad de bits en todas las frecuencias, incrustar una cantidad constante de bits pero sólo en una de cada dos componentes de frecuencia, incrustar diferentes cantidades de bits en bandas específicas del espectro (se dejan bandas de frecuencia sin modificar) e insertar una cantidad específica de bits por muestra dependiendo de la banda de frecuencia, haciendo un barrido continuo del espectro (no se dejan bandas de frecuencia sin modificar). Cada una de estas variantes se exponen con mayor rigor en el Capítulo 4.

- **Proceso de extracción mediante *low bit encoding***: Reconociendo a priori la variante implementada en el transmisor y sus parámetros asociados, se recupera la información llevando a cabo el procedimiento contrario, del cual se extrae la información que representa al mensaje secreto.

[4] El proceso de codificación de fuente asigna palabras código de longitud fija a cada una de las muestras.

3.3.2. Incrustación directa en el espectro de fase, mediante la diferencia entre muestras adyacentes

- **Proceso de incrustación directa**: Este enfoque consiste en incrustar las muestras de audio secreto directamente en el espectro de fase del audio portada. Para garantizar su extracción en el receptor se considera igualar las componentes de fase adyacentes y realizar la diferencia entre una de estas muestras y la muestra del audio secreto correspondiente.

Figura 3.6: *Diagrama del algoritmo. Incrustación directa.*

- **Proceso de extracción directa**: Dado que en el transmisor se garantiza la presencia de muestras de referencia en el espectro de fase, se realiza la diferencia entre muestras adyacentes, donde el resultado corresponde a las muestras temporales del audio secreto.

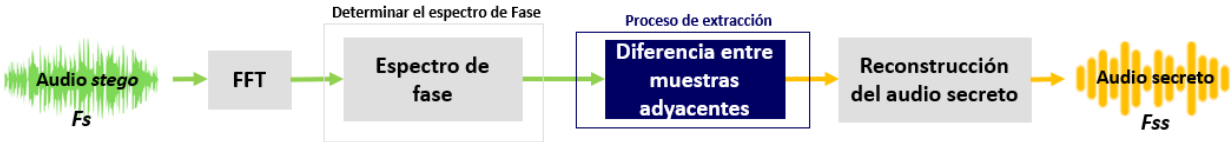


Figura 3.7: *Diagrama del algoritmo. Extracción directa.*

CAPÍTULO 4

IMPLEMENTACIÓN DEL ALGORITMO

4.1. Entorno de Desarrollo

Parte de la correcta implementación y funcionamiento del algoritmo de esteganografía de audio basado en codificación de fase diseñado en este trabajo de grado, depende del entorno de simulación, en este caso MATLAB. Dicho entorno debe permitir evaluar el comportamiento del algoritmo en diferentes condiciones y ajustar sus parámetros para optimizar su rendimiento; es por ello que se deben tener en cuenta diversos factores, en especial aquellos que están relacionados con la importación y exportación de datos, ya que en este escenario el algoritmo debe ser capaz de importar las señales de audio portada y audio secreto y, a su vez, exportar el audio *stego* resultante, junto con el audio secreto recuperado en el receptor.

Partiendo de lo anterior, y previamente a la ejecución de pruebas, es necesario garantizar que el algoritmo desarrollado exporte correctamente el archivo de audio *stego* generado, ya que es importante asegurar que la técnica de esteganografía de audio utilizada es efectiva y que los datos ocultos pueden ser recuperados correctamente, por ello, bajo este contexto se debe tener en cuenta las funciones *audioread* y *audiowrite* de MATLAB.

La función *audioread* permite la lectura de audio en diferentes formatos como: *FLAC, M4A, MP4, OGG, MP3, WAV*, entre otros. Según las especificaciones incluidas en los atributos de la función, la información extraída puede tener una representación propia de cada formato; éste es el caso de *dataType*, el cual puede especificarse como *'native'* o *'double'*, donde para la primera opción, según el formato del audio, cada muestra puede representarse en un rango de 0 a 255 (para una resolución de 8 bits), de -32768 a 32768 (para una resolución de 16 bits), entre otros [36]. Si no es especificado, por defecto se realiza la lectura del audio bajo el tipo de dato *'double'* el cual acota el valor de las muestras entre -1 y 1.

Para la generación del audio *stego* en el transmisor y el audio secreto en el receptor se recurre a la función *audiowrite*; esta función permite generar audios en diferentes formatos, como: *FLAC, M4A, MP4, OGG, WAV*; con el fin de usar un formato que soporte MATLAB y que, a su vez, no genere distorsiones adicionales en el audio *stego*, se selecciona *WAV*. *Audiowrite* recibe algunos parámetros a especificar como *BitsPerSample*, que se refiere a la cantidad de bits que se utilizan para representar cada muestra de audio; su valor por defecto es 16 pero al utilizar los formatos *FLAC* o *WAV*, éste puede tomar

valores de 8, 16, 24, 32 y 64; para el caso específico del formato WAV este valor debe establecerse en 32 o mayor [37], dado que esta resolución es la usada en la lectura de los archivos de audio al usar un *dataType* igual a *'double'*; es necesario que la resolución entre la lectura y escritura concuerde para evitar la inserción de distorsiones adicionales al audio *stego* y asegurar la fidelidad del audio secreto en el receptor.

4.2. Implementación de Variantes del Algoritmo

Dado que los métodos asociados al enfoque de *low bit encoding* requieren de la adecuación de los audios portada y secreto, en su respectivo dominio, su implementación se describe a continuación.

4.2.1. Adecuación de audios portada y secreto

El enfoque de incrustación en el espectro de fase mediante la técnica *low bit encoding* condicionada al dominio transformado exige la determinación y adecuación del espectro de fase del audio portada y la adecuación del audio secreto.

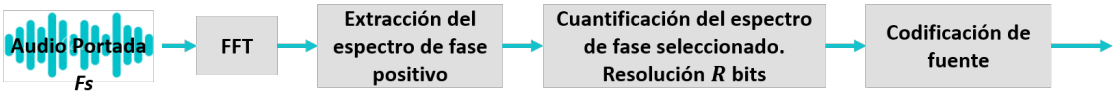


Figura 4.1: *Diagrama de flujo para el proceso de manipulación del audio portada.*

Cada uno de los bloques de la Figura 4.1 hacen parte de la adecuación del audio portada. Después de determinar la frecuencia de muestreo del archivo de audio y realizar la transformación de este audio al dominio de la frecuencia, se extrae el espectro de fase positivo, esto, a partir del vector de frecuencias generado por la frecuencia de muestreo, el cual permite seleccionar las componentes positivas del espectro de fase.

No es posible manipular el espectro de fase completo dado que se generarían modificaciones erróneas al momento de insertar la información, es decir, al incrustar la información se alteraría la simetría del espectro de fase, por lo que al aplicar la IFFT el audio *stego* tendría una naturaleza compleja en el dominio del tiempo, afectando la fidelidad y claridad de la señal de audio generada, y causando distorsiones que podrían ser perceptibles por el HAS.

Después de la transformación, el espectro de fase positivo se somete a la cuantificación y codificación de fuente con una resolución de R bits, brindando muestras adecuadas para realizar la incrustación del audio secreto.

Por su lado, el audio secreto debe someterse a cierta manipulación con el propósito de adaptarlo para su posterior incrustación, este proceso hace referencia al bloque

de “cuantificación, codificación y secuencia binaria” (ver Figura 3.4) que se ilustra en la Figura 4.2.

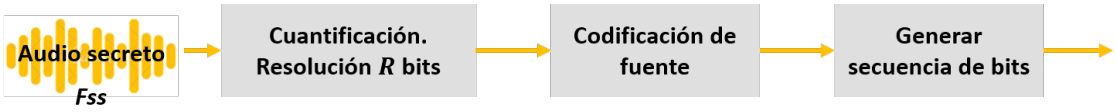


Figura 4.2: *Diagrama de flujo para el proceso de adecuación del audio secreto.*

El audio secreto en su representación temporal se somete a una cuantificación y codificación de fuente, obteniendo muestras de R bits. Estas muestras se concatenan ordenadamente en un único vector, preparando la secuencia para el proceso de incrustación.

Basándose en los diagramas generales expuestos en el Capítulo 3, se presentan las variantes del algoritmo, según el enfoque del *Proceso de Incrustación* implementado.

4.2.2. Método uno

En esta primera versión del algoritmo se hace uso de la técnica *low bit encoding* donde, después de adecuar los audios portada y secreto, se recurre a modificar los LSB de cada una de las muestras que representan al espectro de fase positivo del audio portada; dichas muestras son una cantidad finita representadas por L_P. Un ejemplo de ello se ilustra en la Figura 4.3, donde se presentan las muestras discretas del espectro de fase del audio portada con una $R = 8$ bits, cada una ordenada desde el Bit Más Significativo (MSB, *Most Significant Bit*) hacia el LSB; en este caso se han incrustado los bits del audio secreto en los 3 bits LSB ($n_{LSB} = 3$) de 40 muestras ($L_P = 40$).

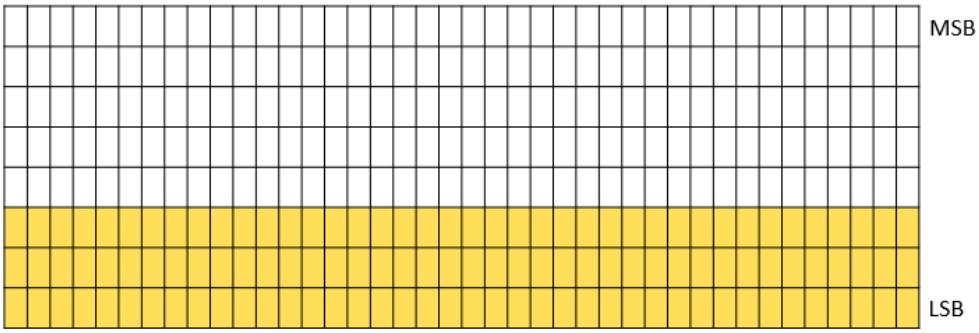


Figura 4.3: *Método uno. Incrustación de $n_{LSB} = 3$ bits en todas las muestras.*

El número de bits que se pueden incrustar y que representan al audio secreto, esta regido por la cantidad de muestras que conforman el espectro de fase positivo del audio portada, al igual que el número total de bits que componen el audio secreto, es con ello que se plantea la Ecuación 4.1, en donde se establece el tamaño del audio secreto medido en segundos, a partir del tamaño del audio portada, con el propósito de reconocer

los límites del algoritmo para ejecutar la incrustación. Asimismo, en la Ecuación 4.2 se establece la duración mínima que debe tener el audio portada medido en segundos, en función del audio secreto que se desee incrustar.

$$A_s\,[segundos] \leq \frac{F_s(B_p-1)}{2F_{ss}B_s}A_p, \tag{4.1}$$

$$A_p\,[segundos] \geq \frac{2F_{ss}B_s}{F_s(B_p-1)}A_s, \tag{4.2}$$

donde:
A_s: duración temporal en segundos del audio secreto.
A_p: duración temporal en segundos del audio portada.
B_s: número de bits que representan una muestra del audio secreto o resolución (R) del audio secreto.
B_p: número de bits que representan una muestra del espectro de fase del audio portada o resolución (R) del audio portada.
F_{ss}: frecuencia de muestreo del audio secreto.
F_s: frecuencia de muestreo del audio portada.

Después de llevar a cabo la incrustación en el dominio de la frecuencia, se genera el audio *stego* en el dominio del tiempo, mediante la decodificación de fuente del espectro de fase positivo, a partir del cual se crea un espectro de fase con simetría impar; éste es usado en la reconstrucción de la señal en el dominio de la frecuencia. Como último paso, se aplica la FT inversa.

En el receptor, el algoritmo recupera la información secreta incrustada en transmisión aplicando la FFT al audio *stego* recibido y determinando el espectro de fase positivo. Seguidamente, realiza el proceso de cuantificación y codificación de fuente con una resolución de R bits, tomando en consideración el número de bits modificados por muestra (n_{LSB}), los cuales el receptor debe conocer a priori, junto con la frecuencia de muestreo (F_{ss}) para lograr extraer los LSB de cada una de las muestras y dar paso a la reconstrucción del audio secreto.

4.2.3. Método dos

En esta versión del algoritmo, a diferencia del caso anterior, se modifica una de cada dos componentes de frecuencia del espectro de fase; para ejemplificar este tipo de incrustación se presenta la Figura 4.4, en donde se asume la modificación de 3 bits de información por una de cada dos muestras.

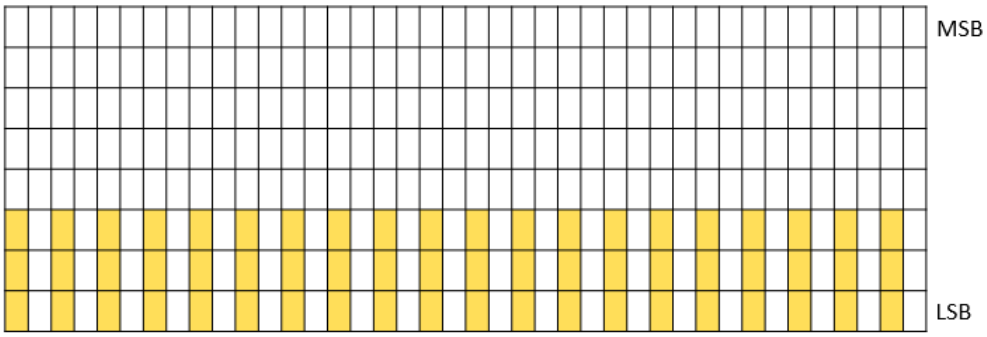


Figura 4.4: *Método dos. Incrustación de $n_{LSB} = 3$ bits en cada muestra impar.*

La cantidad de bits a incrustar en este método no está limitada únicamente por el espectro de fase positivo de la señal de audio portada, sino también porque la alteración de las muestras no es continua, es por ello que la capacidad de incrustación con respecto al primer caso se reduce a la mitad; partiendo de esto, es posible calcular la duración temporal del audio secreto que se puede ocultar, considerando la duración del audio portada, o la duración necesaria del audio portada para ocultar un audio secreto deseado, tal y como se muestra en las Ecuaciones 4.3 y 4.4 respectivamente.

$$A_s\,[segundos] \leq \frac{F_s(B_p - 1)}{4F_{ss}B_s}A_p, \tag{4.3}$$

$$A_p\,[segundos] \geq \frac{4F_{ss}B_s}{F_s(B_p - 1)}A_s, \tag{4.4}$$

donde:
A_s: duración temporal en segundos del audio secreto.
A_p: duración temporal en segundos del audio portada.
B_s: número de bits que representan una muestra del audio secreto o resolución (R) del audio secreto.
B_p: número de bits que representan una muestra del espectro de fase del audio portada o resolución (R) del audio portada.
F_{ss}: frecuencia de muestreo del audio secreto.
F_s: frecuencia de muestreo del audio portada.

Por su parte, el receptor debe conocer previamente el número de bits incrustados en cada muestra, con el objetivo de extraerlos después de realizar la cuantificación y codificación del espectro de fase positivo del audio *stego* recibido; posteriormente se reconstruye la secuencia de bits del audio secreto y con ello sus muestras, así, reconociendo a priori su frecuencia de muestreo (F_{ss}), se genera el archivo de audio para su reproducción.

4.2.4. Método tres

Bajo la premisa de que el HAS es más sensible a alteraciones en ciertas bandas de frecuencias, se desarrolla esta versión del algoritmo, la cual se basa en los diagramas expuestos en las Figuras 3.4 y 3.5. Al igual que las anteriores versiones, en el transmisor, los audios portada y secreto sufren su respectiva adecuación.

Para iniciar con el proceso de incrustación, es necesario definir las bandas de frecuencia sobre las cuales se realizará la inserción, además del porcentaje de bits que serán insertados en cada una; esto se ve representado en la Figura 4.5, donde se han establecido tres bandas de inserción, cada una con un número determinado de bits disponibles para la incrustación.

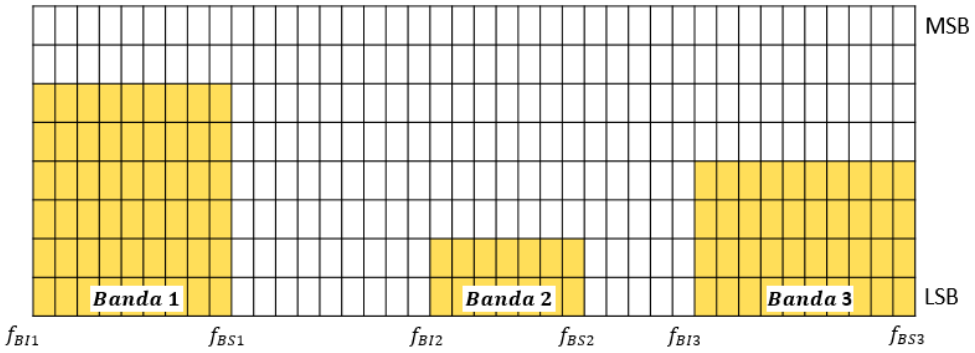


Figura 4.5: *Método tres. Alteración por bandas de las componentes fase.*

Tanto la longitud de las bandas como el porcentaje de bits disponibles para la inserción en cada una restringe la cantidad de información a insertar; basándose en estos atributos, se puede estimar la longitud temporal de audio secreto que es posible incrustar en términos del audio portada o la longitud del audio portada necesaria para incrustar el audio secreto deseado, mediante las Ecuaciones 4.5 y 4.6 respectivamente.

$$A_s\,[segundos] \leq \frac{(B_p - 1)\,\alpha}{B_s F_{ss}} A_p, \tag{4.5}$$

$$A_p\,[segundos] \geq \frac{B_s F_{ss}}{(B_p - 1)\,\alpha} A_s, \tag{4.6}$$

donde:

$$\alpha = \sum_{n=1}^{W} (f_{BSn} - f_{BIn})\, B_{Kn},$$

$$0 \leq B_{Kn} \leq 1 \qquad \wedge \qquad f_{BSn} > f_{BIn},$$

A_s: duración temporal en segundos del audio secreto.
A_p: duración temporal en segundos del audio portada.
B_s: número de bits que representan una muestra del audio secreto o resolución (R) del audio secreto.
B_p: número de bits que representan una muestra del espectro de fase del audio portada o resolución (R) del audio portada.
F_{ss}: frecuencia de muestreo del audio secreto.
W: número de bandas consideradas para la incrustación.
B_{Kn}: cantidad de inserción de información en una banda definida en el rango de $0\ (0\,\%)$ a $1\ (100\,\%)$.
f_{BIn}, f_{BSn}: frecuencias inferior y superior que limitan una banda.

El receptor debe conocer previamente las frecuencias límite de cada banda, así como el número de bits incrustados en cada una mediante los porcentajes. Con estos parámetros, el receptor realiza la cuantificación y codificación del espectro de fase positivo del audio *stego* recibido. Seguidamente, identifica las bandas asociadas a la incrustación del audio secreto y extrae los bits deseados, para reconstruir la secuencia de bits del audio secreto, sus muestras y, finalmente, el archivo de audio adecuado para su reproducción.

4.2.5. Método cuatro

Al igual que en el caso anterior y partiendo de la premisa en donde el HAS es más sensible a ciertas bandas de frecuencia, especialmente en el rango de $1.5KHz$ a $4KHz$ [9], se establece este método, en el cual el número de bits a incrustar por banda se define de acuerdo al rango de frecuencias que la compone, es decir, se incrusta un número específico de bits en cada una de las muestras acorde al rango de frecuencias al cual pertenezca la muestra; cabe resaltar que a diferencia del caso anterior, en esta variante sí se modifican todas las componentes del espectro de fase positivo. A modo ilustrativo se presenta un ejemplo en la Figura 4.6.

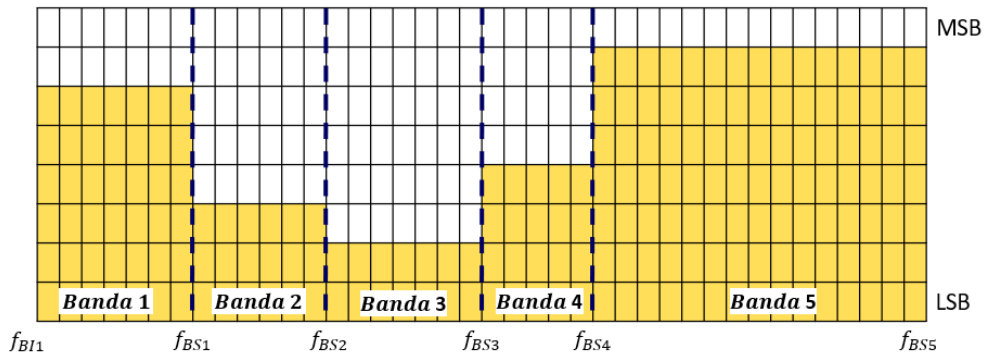


Figura 4.6: *Método cuatro. Alteración de todas las componentes del espectro de fase en diferentes proporciones.*

En esta versión del algoritmo es necesario establecer las frecuencias límite entre las bandas y el porcentaje de inserción en cada una de ellas. Establecidos los parámetros adecuados, el algoritmo ejecuta la FFT sobre el audio portada, obtiene su espectro de fase positivo que, posteriormente, se cuantifica y codifica; mientras tanto, el audio secreto se comprime, obteniendo la secuencia de bits que lo representa. Identificadas las bandas de inserción y la cantidad de bits que es posible incrustar en cada muestra, se procede a reemplazar los LSB por la secuencia que representa al audio secreto, de forma ordenada hasta incrustar toda la información disponible.

La cantidad de información a incrustar se encuentra restringida por los parámetros usados en esta versión. Para garantizar que la información secreta deseada sea totalmente incrustada, se recurre a las Ecuaciones 4.7 y 4.8.

$$A_s\,[segundos] \leq \frac{(B_p - 1)\,\beta}{B_s F_{ss}} A_p, \tag{4.7}$$

$$A_p\,[segundos] \geq \frac{B_s F_{ss}}{(B_p - 1)\,\beta} A_s, \tag{4.8}$$

donde:

$$\beta = f_{BSW} B_{KW} - f_{BI1} B_{K1} + \sum_{n=1}^{W-1} f_{BSn} \left(B_{Kn} - B_{K(n+1)}\right),$$

$$0 \leq B_{Kn} \leq 1 \qquad \wedge \qquad f_{BSn} > f_{BIn}$$

A_s: duración temporal en segundos del audio secreto.
A_p: duración temporal en segundos del audio portada.
B_s: número de bits que representan una muestra del audio secreto o resolución (R) del audio secreto.
B_p: número de bits que representan una muestra del espectro de fase del audio portada o resolución (R) del audio portada.
W: número de bandas consideradas para subdividir el espectro de fase.
F_{ss}: frecuencia de muestreo del audio secreto.
B_{Kn}: cantidad de inserción de información en una banda definida en el rango de $0\ (0\,\%)$ a $1\ (100\,\%)$.
f_{BIn}, f_{BSn}: frecuencias inferior y superior que limitan una banda.

Finalizada la incrustación, se construye el espectro de fase y se combina con el espectro de magnitud original del audio portada para generar el audio *stego* mediante la IFFT.

En cuanto al receptor, éste debe conocer con anterioridad los parámetros usados en el transmisor para recuperar adecuadamente el audio secreto; la extracción se ejecuta como lo demuestra la Figura 3.5.

4.2.6. Método cinco

Esta versión abandona por completo la técnica *low bit encoding* y recurre a la modificación directa del espectro de fase ilustrado en la Figura 3.6.

El audio portada se debe someter a la FFT para extraer su espectro de fase positivo. Dicho espectro es discreto en frecuencia, aunque continuo en magnitud; basándose en esa discretización frecuencial, se igualan las magnitudes de las muestras adyacentes, esto con el objetivo de obtener una referencia para la posterior recuperación del audio secreto incrustado [24]; adicionalmente, éstas se someten a una atenuación de $-3dB$. Atenuar el espectro es necesario debido a que éste se encuentra acotado entre los valores de $-\pi$ y π, los cuales son límites que no deben superarse durante la inserción del audio secreto, esto, para evitar ambigüedades al momento de implementar la IFFT.

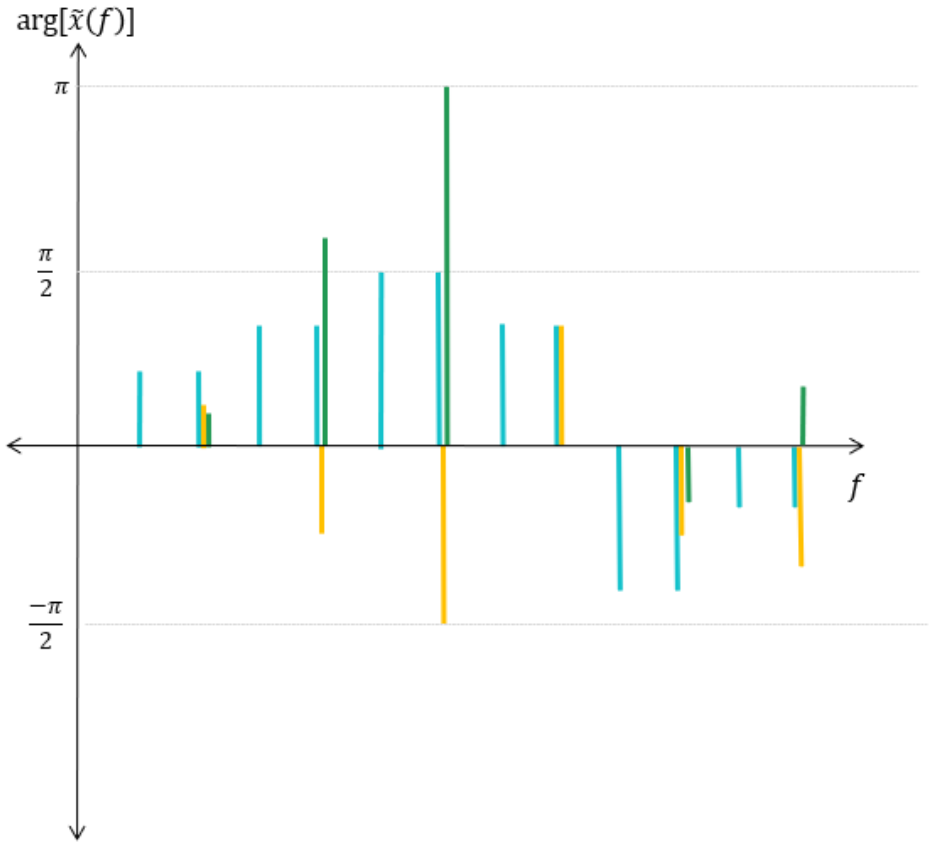


Figura 4.7: *Método cinco. Proceso de incrustación por diferencia de componentes de fase.*

Posteriormente, se calcula la diferencia entre cada una de las muestras pares del espectro de fase positivo y las muestras temporales del audio secreto, de forma ordenada. Es así como en la Figura 4.7, las muestras de color verde representan el resultado de dicha operación. Concluida la incrustación, se construye el espectro de la señal y se aplica la IFFT, dando como resultado el audio *stego*.

En el receptor, se extraen las muestras temporales del audio *stego* a partir de la implementación de la FFT, la extracción de su espectro de fase positivo y la ejecución de la operación usada en el transmisor: cada muestra temporal del audio secreto corresponde a la diferencia entre dos muestras adyacentes (impar y par) del audio *stego*. Una vez recuperadas todas las muestras, se reconstruye el audio secreto.

Dadas las cualidades de este método es posible calcular la duración temporal del audio secreto o el audio portada a usar en el proceso de incrustación mediante las Ecuaciones 4.9 y 4.10.

$$A_s\,[segundos] \leq \frac{F_s}{4F_{ss}} A_p, \tag{4.9}$$

$$A_p\,[segundos] \geq \frac{4F_{ss}}{F_s} A_s, \tag{4.10}$$

donde:
A_s: duración temporal en segundos del audio secreto.
A_p: duración temporal en segundos del audio portada.
F_{ss}: frecuencia de muestreo del audio secreto.
F_s: frecuencia de muestreo del audio portada.

4.3. Algoritmo de Esteganografía de Audio Basado en la Técnica Temporal *Low Bit Encoding*

Con el objetivo de medir y comparar el desempeño de las diferentes variantes del algoritmo de esteganografía de audio basado en codificación de fase con respecto a un enfoque más tradicional, se implementa un algoritmo basado en la técnica temporal *low bit encoding* en el dominio temporal. Los bloques de la Figura 4.8 representan los procesos que ejecuta este algoritmo en el transmisor. Dichos bloques se describen a continuación.

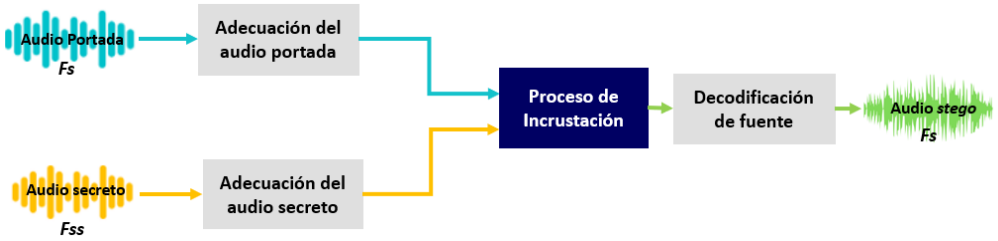


Figura 4.8: *Diagrama general del funcionamiento del algoritmo temporal. Transmisor.*

- **Adecuación del audio portada**: El audio portada se somete a una adecuación previo al proceso de incrustación; en este caso su adaptación se da en términos de la cuantificación y codificación de fuente.

- **Adecuación del audio secreto**: El audio secreto se somete a una adecuación que depende del proceso de incrustación del cual será partícipe, en este caso *low bit encoding*; es por ello que su adaptación también se da en términos de cuantificación y codificación de fuente.

- **Proceso de incrustación**: En este bloque se lleva a cabo la modificación de la señal de audio portada a partir del reemplazo de los LSB, de cada una de sus muestras resultantes del proceso de codificación, por los bits que representan cada muestra del audio secreto a incrustar. Este número de bits por muestra a modificar se calcula a partir de la Ecuación 4.11, que resulta de una adecuación de la Ecuación 3.3.

$$n_{LSB} = \left\lceil \frac{T_{Bs}}{L_{pt}} \right\rceil = \left\lceil \frac{L_s B_s}{L_{pt}} \right\rceil , \tag{4.11}$$

  donde:
  n_{LSB}: número de bits menos significativos por muestra disponibles para la incrustación.
  T_{Bs}: número total de bits que componen el audio secreto (s) en el dominio del tiempo.
  L_{pt},L_s: número de muestras que representan el audio portada o el audio secreto en el dominio del tiempo.
  B_{pt}, B_s: número de bits que representan una muestra o resolución (R) de los audios portada y secreto.

  Además, se debe garantizar que:

$$n_{LSB} < B_{pt} \qquad \wedge \qquad L_s B_s < L_{pt} B_{pt}$$

  Basándose en estas restricciones, es posible calcular la duración del audio secreto a partir de la duración del audio portada o viceversa, mediante las Ecuaciones 4.12 y 4.13.

$$A_s\,[segundos] \leq \frac{F_s(B_{pt}-1)}{F_{ss}B_s}A_p, \tag{4.12}$$

$$A_p\,[segundos] \geq \frac{F_{ss}B_s}{F_s(B_{pt}-1)}A_s, \tag{4.13}$$

- **Decodificación de fuente**: En este bloque la señal modificada por el proceso de incrustación se decodifica con el fin de obtener una versión reconstruida de la señal de audio portada, dando como resultado el audio *stego*.

Por otra parte, en el receptor el algoritmo lleva a cabo la extracción del audio secreto, mediante el procedimiento expuesto en la Figura 4.9.

Figura 4.9: *Diagrama general del funcionamiento del algoritmo temporal. Receptor.*

- **Cuantificación y codificación del audio *stego***: En este bloque se cuantifica y codifica la señal de audio *stego*, generando así una secuencia discreta de valores de amplitud con una resolución igual a la implementada en transmisión, que posteriormente permite realizar el proceso de extracción.
- **Proceso de extracción**: Reconociendo a priori los parámetros usados en el transmisor, se lleva a cabo la extracción de los bits que componen información que representa al mensaje secreto.
- **Reconstrucción del audio secreto**: Obtenidos los bits del proceso de extracción, éstos se ordenan correctamente con el fin de obtener las muestras codificadas del audio secreto. Una vez determinadas las muestras, se someten a la decodificación de fuente para generar la señal de audio secreto adecuada para su reproducción.

4.4. Pruebas de Validación

Los métodos de codificación de fase presentados difieren en el proceso de incrustación y extracción de la información secreta, lo cual puede beneficiar o perjudicar las características esteganográficas ideales (Figura 2.1). Dado que se usan señales de audio como archivos base para el proceso esteganográfico, se recurre a analizar la imperceptibilidad de cada método, puesto que es una característica de gran relevancia según el enfoque de este trabajo de grado, por lo que se debe considerar la sensibilidad del HAS al momento de excluir los métodos menos adecuados mediante pruebas de validación.

Para estas pruebas de validación se cuenta con un banco de audios compuesto por 33 canciones de diversos géneros musicales (clásica, pop, ranchera, rock, tropical, urbana y vallenato), de las cuales 30 de ellas son utilizadas como audios portada y 5 como audios secretos. Partiendo de esto, en la simulación, cada uno de los audios portada se combina con cada uno de los audios secretos, dando como resultado 150 audios *stego*, los cuales se evalúan a partir de su valor de SNR, la cual se calcula comparando la señal de audio portada utilizada y la señal de audio *stego* resultante de cada método, siendo una medida objetiva que permite estimar el grado de distorsión o ruido presente en el audio *stego* en comparación con el audio portada.

En cada uno de los métodos se incrusta la misma cantidad de información secreta, dando paso a la inserción de $2'822.400$ bits en archivos de audio de 50 segundos de duración con frecuencia de muestreo de $44,1KHz$. Así mismo, se consideran los parámetros expuestos en la Tabla 4.1 para los *métodos tres y cuatro*.

Tabla 4.1: Pruebas de validación. Parámetros implementados en *métodos tres y cuatro*.

Método	**Banda [Hz]**	**Número de LSB**
Tres	0 - 200	6
	500 - 800	2
	12.000 - 22.050	6
Cuatro	0 - 1.000	3
	1.000 - 2.000	2
	2.000 - 5.000	3
	5.000 - 7.000	5
	7.000 - 22.050	6

En la Figura 4.10 es notoria la afectación que el método usado para la inserción de la información puede ocasionar a la imperceptibilidad en el audio *stego*: la modificación directa de las muestras del espectro de fase generan cambios muy ostensibles en la señal de audio, causando mayor distorsión; sin embargo, estos cambios pueden reducirse mediante el uso de otra técnica como *low bit encoding*, aunque ésta conlleve procesos adicionales como la adecuación del espectro de fase y el audio secreto.

A partir de los resultados obtenidos en la Figura 4.10 y tomando en consideración que una SNR elevada es deseable, puesto que se puede asociar con un cambio poco perceptible en el audio *stego*, al ser comparado con el audio portada; es evidente que los *métodos uno, tres y cuatro* destacan con valores promedio de SNR por encima de $26.53dB$, en comparación a los valores de SNR negativos obtenidos en el *método cinco* dando paso a la exclusión de este método.

De forma similar el *método dos* se considera inadecuado para realizar pruebas poste-

riores, puesto que en este trabajo de grado se prioriza la imperceptibilidad y la capacidad de incrustación de información, las cuales no tienen valores adecuados en este método. En la Figura 4.10 se muestra que la media de los resultados del *método dos* se encuentra por debajo del valor de $26.53dB$, el cual se toma como referencia, dado que corresponde a la cota inferior de los métodos más destacados en cuanto a imperceptibilidad. Por otro lado, dejar una muestra sin modificar disminuye la capacidad de incrustación de la información, lo que para estas pruebas implica que $n_{LSB2} < n_{LSB1}$.

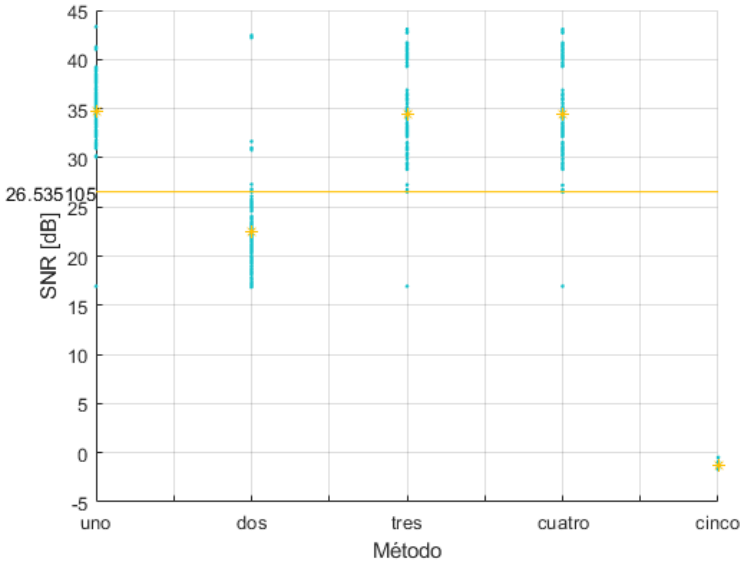


Figura 4.10: *Pruebas de validación. Resultados de simulación.*

Con el fin de verificar la veracidad de los resultados de simulación, se realiza una prueba preliminar de percepción, para la cual se cuenta con la participación de 13 personas[5] entre los 19 y 24 años de edad. Como primera medida se verifica el límite superior auditivo de cada participante[6] reproduciendo una serie de tonos de diferentes frecuencias, iniciando por un tono de $20KHz$ hasta un tono de $12KHz$ (ver Figura 4.11).

[5]En el Apéndice B se muestra el consentimiento informado que cada uno de los participantes aceptó antes de realizar las pruebas.

[6]Esta prueba busca confirmar la importancia de contar con personas de diferentes edades en las pruebas subjetivas, puesto que este factor puede influir en la valoración que realicen; sin embargo, con el fin de tener un caso más crítico se consideran personas cercanas a los 20 años, buscando que su rango de audición sea aún bastante amplio y se puedan detectar distorsiones en altas frecuencias.

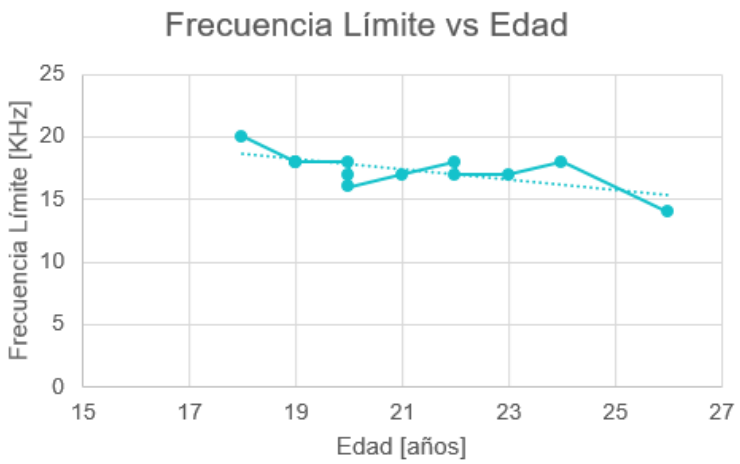


Figura 4.11: *Resultados de prueba preliminar de percepción de tonos.*

Posteriormente, se llevó a cabo la reproducción de 10 audios *stego* seleccionados del banco resultante de la prueba de simulación, garantizando que la cantidad de información secreta incrustada y la duración de los audios sean las mismas; adicionalmente, se utilizaron los mismos audífonos y el mismo nivel de volumen para homogeneizar las condiciones de las pruebas. La selección de audios para el desarrollo de esta prueba se realizó bajo los siguientes criterios:

- Cada uno de los 10 audios portada utilizados es diferente.
- El audio secreto contenido en los 10 audios *stego* seleccionados es el mismo para todos los casos.
- Los audios *stego* seleccionados representan diferentes valores de SNR.

En estas pruebas cada participante, al escuchar un audio *stego* asigna una puntuación de 1 a 5 dependiendo de la calidad del audio, i.e., se realiza una prueba de MOS. En la Tabla 4.2 se expone la MOS como resultado de las calificaciones promedio brindadas por todos los participantes para cada uno de los audios presentados junto con información relevante para su respectivo análisis.

Tabla 4.2: Resultados de pruebas preliminares. BER, SNR y MOS de audios *stego* en prueba preliminar de percepción.

Audio *stego*	Género	Método	BER	SNR [dB]	MOS
Ae27	Vallenato	*tres*	0	$32,63$	$4,64$
Ae11	Ranchera	*dos*	$6,37 \times 10^{-6}$	$26,53$	$3,92$
Ae22	Tropical	*dos*	0	$25,45$	$2,85$
Ae4	Pop	*dos*	0	$23,92$	$2,85$
Ae7	Pop	*dos*	$2,12 \times 10^{-6}$	$23,32$	$2,92$
Ae8	Pop	*dos*	$3,54 \times 10^{-7}$	$22,24$	$2,5$
Ae6	Pop	*dos*	0	$21,94$	$2,57$
Ae12	Ranchera	*dos*	$1,34 \times 10^{-5}$	$20,53$	$3,07$
Ae18	Rock	*dos*	0	$18,55$	$3,71$
Ae21	Tropical	*dos*	$1,52 \times 10^{-5}$	$17,20$	$2,21$

En la Figura 4.12 se presenta la relación entre el valor de SNR de cada uno de los audios *stego* y su MOS como resultado de la prueba; cabe resaltar que como comportamiento ideal se esperaría que a medida que aumenta el valor de SNR, la calidad del audio mejore, considerando que se ha garantizado la misma cantidad de información oculta en todos los casos; sin embargo, en los resultados obtenidos no se cumple con esta premisa en totalidad, ya que, como se evidencia en la diferencia entre la gráfica de los resultados y la línea recta esperada, algunos de los audios *stego* obtuvieron una MOS más alta que la pronosticada. Tal es el caso del audio *Ae18* que, en comparación con otros audios, posee un valor de SNR mucho menor y cuya calificación se encuentra por encima de los audios *Ae22*, *Ae4*, *Ae7*, *Ae8*, *Ae6*, *Ae12* y *Ae21*.

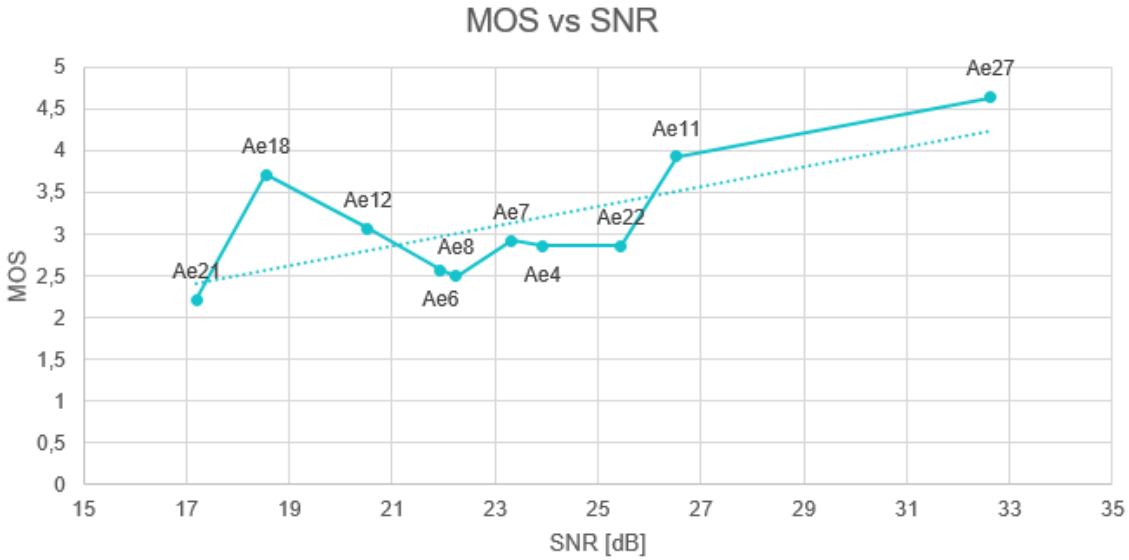


Figura 4.12: *Resultados de la prueba preliminar de percepción SNR vs MOS.*

Con el objetivo de verificar la MOS calculada para *Ae18* se realiza una prueba adicional que consiste en repetir la prueba anterior, donde 13 personas evalúan 10 audios *stego*; no obstante, nueve de los audios hacen parte de la primera prueba, mientras el décimo corresponde al audio *Ae24* que presenta una SNR de $18,32dB$, cercana a la otorgada por *Ae18* (ver Tabla 4.2), con lo cual se obtienen los resultados mostrados en la Figura 4.13.

Tabla 4.3: Resultados de las segundas pruebas preliminares de percepción. Promedio de calificaciones de audios *stego*.

Audio *stego*	**Género**	**Método**	**BER**	**SNR [dB]**	**MOS**
Ae27	Vallenato	*tres*	0	$32,63$	$4,38$
Ae11	Ranchera	*dos*	$6,37 \times 10^{-6}$	$26,53$	$3,76$
Ae22	Tropical	*dos*	0	$25,45$	$2,84$
Ae4	Pop	*dos*	0	$23,92$	3
Ae7	Pop	*dos*	$2,12 \times 10^{-6}$	$23,32$	$2,84$
Ae8	Pop	*dos*	$3,54 \times 10^{-7}$	$22,24$	$2,92$
Ae6	Pop	*dos*	0	$21,94$	3
Ae12	Ranchera	*dos*	$1,34 \times 10^{-5}$	$20,53$	$3,07$
Ae24	Urbana	*dos*	0	$18,32$	$2,07$
Ae21	Tropical	*dos*	$1,52 \times 10^{-5}$	$17,20$	$2,61$

En la Figura 4.13 se observa que, en la segunda prueba la relación entre el valor de la SNR y la MOS es mucho más congruente con la premisa establecida (a mayor SNR mayor MOS), en comparación con la primera prueba (Figura 4.12).

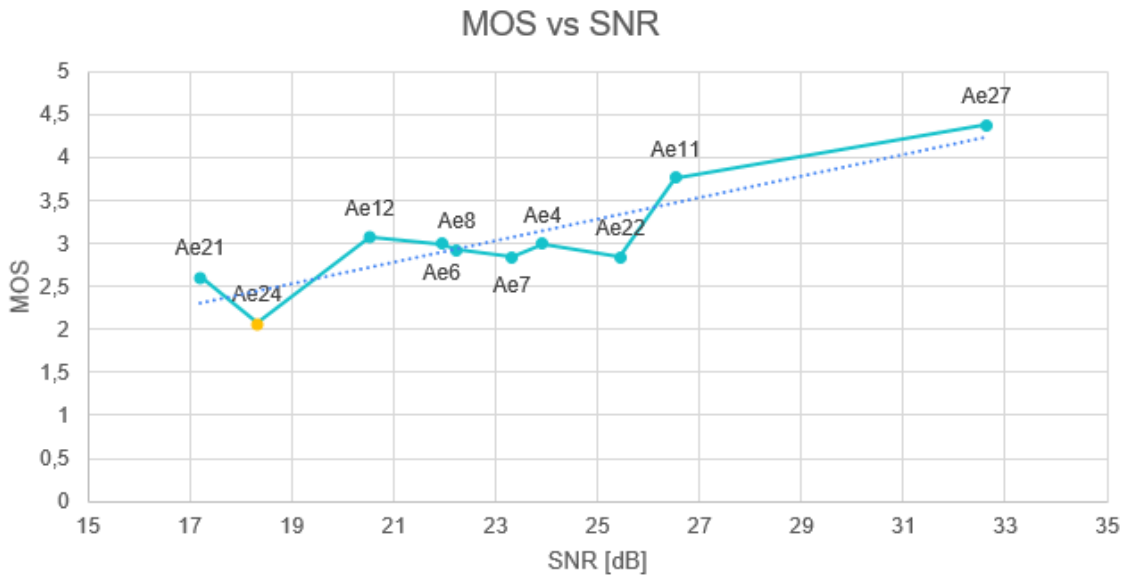


Figura 4.13: *Resultados de la segunda prueba preliminar de percepción SNR vs MOS.*

Dado el cambio de audios *stego* con SNR muy próximas entre sí (*Ae18*, *Ae24*) es posible asumir que el audio portada influye considerablemente en la calidad del audio *stego* resultante, por lo que se hace necesaria una caracterización más pormenorizada de los audios portada según su género; este factor se aborda en la siguiente sección.

Además de la imperceptibilidad del audio *stego*, es necesario corroborar la correcta extracción del audio secreto en el receptor; para ello, los audios *stego* usados en las pruebas preliminares de percepción son procesados por su respectivo receptor. Recuperado el audio secreto se procede a calcular su BER, dando como resultado lo expuesto en las Tablas 4.2 y 4.3 para el procesamiento de los 11 audios *stego*.

Es notoria la proximidad de los valores de BER a cero, indicando que la información secreta recuperada es muy próxima a la versión incrustada, esto es, no se presentan pérdidas significativas de la información confidencial, validando la posibilidad de recuperar la información secreta adecuadamente al concluir el proceso esteganográfico por parte del receptor[7], es por ello que esta métrica de evaluación objetiva no se tendrá en cuenta mas adelante.

4.5. Caracterización de Audios Portada

A partir de los resultados obtenidos en la sección anterior, es posible deducir que el desempeño del algoritmo está relacionado en gran medida con la selección del audio portada, ya que para ejecutar dichas pruebas de validación se incrusta un mismo audio secreto en diferentes audios portada y la imperceptibilidad de los audios *stego* varía considerablemente de acuerdo a la señal de audio portada utilizada. Es por ello que se deben buscar las características deseables en un audio portada que lo hacen más apropiado para el ocultamiento de información.

Aunque una caracterización rigurosa de las señales de audio desborda el alcance de este trabajo de grado, se opta por abordar dos medios: una caracterización desde el dominio de la frecuencia y otra desde el dominio del tiempo. En el dominio de la frecuencia se utilizan los espectrogramas y los valores de la desviación estándar de su espectro de magnitud. Por otro lado, para el dominio del tiempo se utiliza la energía total de cada señal de audio y, nuevamente, la desviación estándar de los valores de amplitud de las muestras.

Cabe mencionar que la caracterización de audios se podría extender hacia los audios confidenciales, sin embargo, mediante una prueba de simulación expuesta en el apéndice D se evidencia que las características de la información secreta son irrelevantes para la

[7]Como se ha mencionado en las anteriores secciones, el receptor debe conocer de antemano los parámetros implementados en el transmisor; para ello, se recurre a la generación de una Clave que el transmisor envía al receptor con anticipación (ver Apéndice E).

imperceptibilidad del audio *stego*; sí es primordial que dicha información pueda representarse como una secuencia binaria.

4.5.1. Caracterización en el dominio de la frecuencia

4.5.1.1. Espectrogramas

En la Figura 4.10 es posible identificar que ciertos audios portada producen valores de SNR elevados, generando una brecha considerable entre su valor específico y el valor promedio de SNR obtenido en el método al cual fue sometido para dar paso al respectivo audio *stego*; tal es el caso del audio portada que generó valores de SNR entre 42 y 43 *dB* a lo largo de los *métodos uno, dos, tres y cuatro* que, a partir de ahora, se denominará *Ap1*. Con esa cualidad a favor, se procede a obtener su espectrograma.

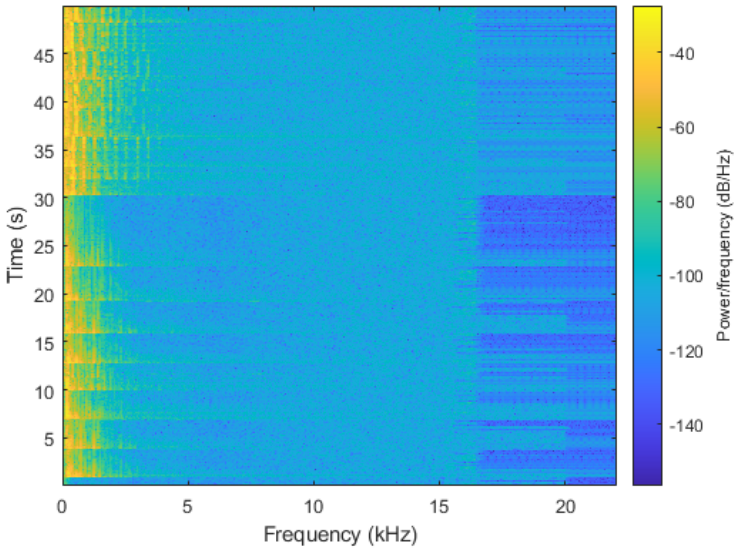


Figura 4.14: *Espectrograma del audio portada Ap1.*

En la Figura 4.14 se observa la Densidad Espectral de Potencia (PSD, *Power Spectral Density*) del *Ap1* a lo largo de ventanas temporales de 20 *ms* de duración. Es notorio que los valores más elevados de densidad se encuentran concentrados en la banda de 0 - 3 *KHz*, aproximadamente, durante los 50 segundos de duración del audio.

El espectrograma del *Ap1* se considera, por el desempeño del mismo, como una característica deseable en un audio portada, i.e., valores elevados concentrados en la banda de 0 - 3 *KHz*, por lo que se procede a determinar los espectrogramas de los otros audios portada usados en las pruebas preliminares (ver Apéndice C); con el fin de sintetizar la información obtenida, se evalúa el porcentaje de PSD, en rangos de 3 *dB* asociados a niveles de intensidad, que se concentran en las componentes de frecuencia hasta los 3 *KHz*, dando paso a la Figura 4.15.

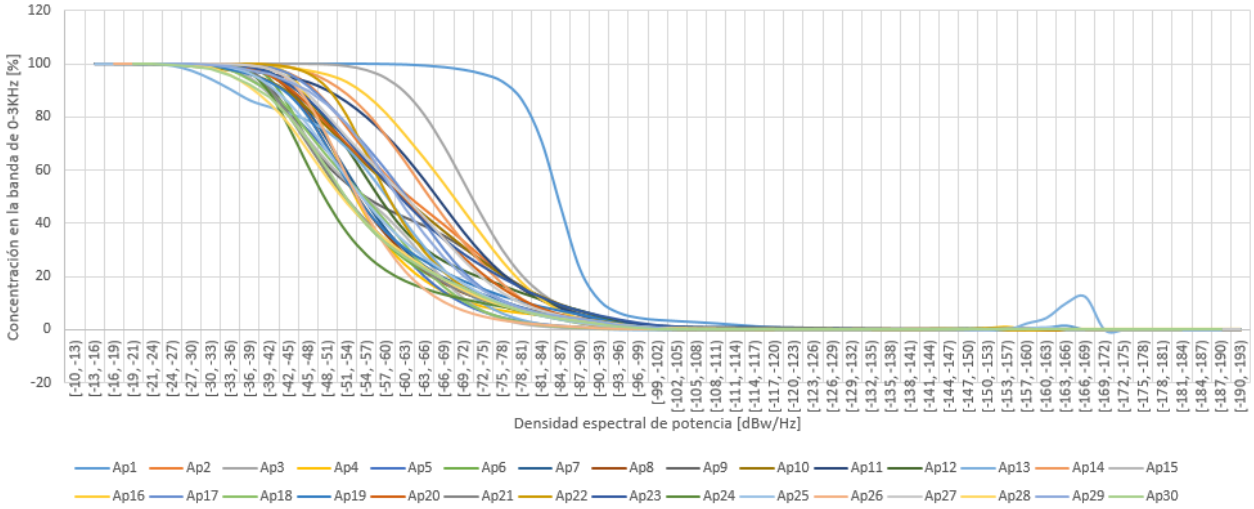


Figura 4.15: *Concentración de la PSD de los audios portada en la banda de* $0 - 3KHZ$.

Aunque la Figura 4.15 demuestra la particularidad del audio *Ap1*, no brinda la suficiente claridad en cuanto al comportamiento de los demás audios portada, por esta razón, se decide asumir la clasificación por géneros musicales.

Para garantizar la equidad en número de audios por género, es necesario modificar el banco de audios, sin alterar aquellos usados en las pruebas preliminares de percepción; de los resultados es posible observar que la acumulación de densidad de potencia en banda base $(0 - 3KHz)$ no influye directamente en la SNR calculada (ver Apéndice C), pero sí brinda una aproximación a la imperceptibilidad del audio *stego* ante el HAS. Un ejemplo de ello ocurre en los audios *stego Ae4*, *Ae6*, *Ae7*, *Ae8* y *Ae11*, *Ae12* que hacen parte de los géneros Pop y Ranchera respectivamente (ver Tabla 4.3); los porcentajes de acumulación de la PSD de sus audios portada se exponen en la Figura 4.16, donde se percibe que el *Ap12* presenta un bajo decaimiento al compararlo con los audios asociados al Pop, aunque su SNR sea menor. Así mismo, se observa que, en la Tabla 4.3, el audio *stego* generado a partir de *Ap12* (*Ae12*) tiene una mejor calificación que el audio *stego* producto de *Ap11* (*Ae11*). Por su lado, al estar *Ap11* más hacia la derecha que *Ap12* en la Figura 4.16, *Ap11* es la mejor opción para ser empleado como audio portada.

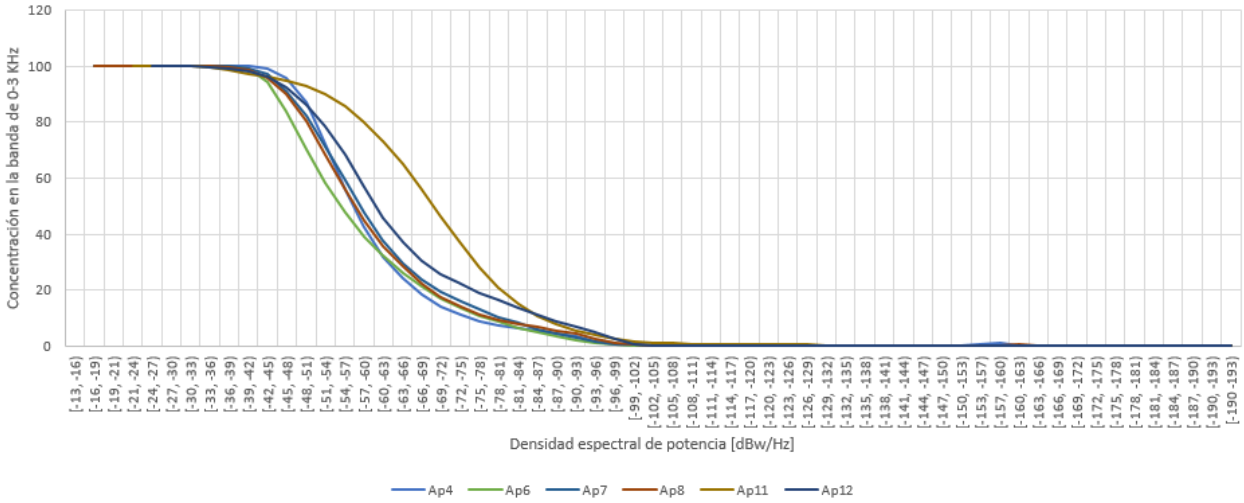


Figura 4.16: *Comparación entre pruebas preliminares de percepción para el género Pop y resultados de espectrogramas.*

Lo anterior puede ser usado como una forma de seleccionar un buen audio portada entre múltiples opciones, pero no es un método totalmente infalible; en la Figura 4.16, puede notarse que las gráficas de asociadas a los audios *Ap4*, *Ap6*, *Ap7* y *Ap8* no presentan una diferenciación significativa entre ellas, esto es, las concentraciones de la PSD en la banda de 0-3 *KHz* son muy próximas, además de que se alternan para ciertos niveles, generando entrelazamientos. Las calificaciones obtenidas para los cuatro audios *stego* (*Ae4*, *Ae6*, *Ae7*, *Ae8*) no difieren considerablemente, por lo que en este caso la caracterización indica que los audios tienen un nivel aproximado de cualidades que los llevan a brindar audios *stego* con MOS muy cercanas entre si; dado el caso en que la MOS de los audios no concuerde con esa proximidad entre las concentraciones de PSD, sería necesario apoyarse de otros medios para obtener una caracterización más certera y fiable.

4.5.1.2. Desviación estándar

A partir de la clasificación por géneros musicales de los audios portada, se recurre a un análisis estadístico mediante el cálculo de la desviación estándar del espectro de magnitud de cada una de las señales de audio portada, obteniendo así la dispersión de las amplitudes con respecto a su media; esto permite identificar algunas características propias de los géneros musicales, de esta manera, una desviación estándar baja indica que las amplitudes están relativamente cerca de la media y una desviación estándar alta apunta a que las amplitudes están más dispersas, de tal forma que algunas de las amplitudes están más alejadas de la media que otras. En la Figura 4.17 se presenta el promedio de las desviaciones estándar por cada género músical evaluado, en donde destaca la música Clásica, ya que presenta la desviación estándar más baja comparada con los otros seis géneros musicales evaluados, dando así un indicio de su favorabilidad como audios portada, ya que de acuerdo a los resultados obtenidos a priori existe una relación

predominante entre este género y su comportamiento ideal para el ocultamiento de información, lo cual puede atribuirse a la poca variabilidad de las amplitudes que componen el espectro de magnitud de este tipo de señales.

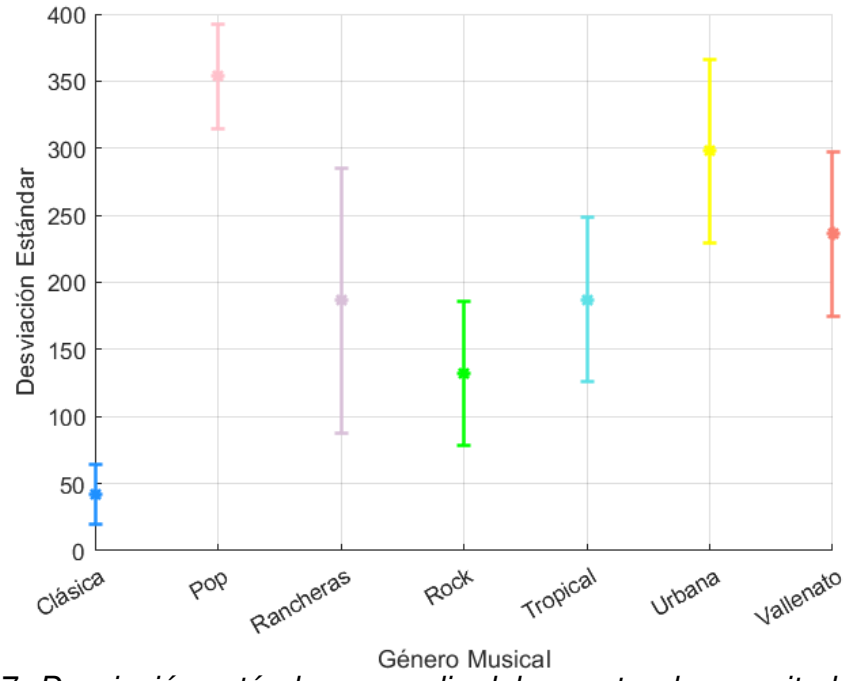


Figura 4.17: *Desviación estándar promedio del espectro de magnitud por género musical.*

4.5.2. Caracterización en el dominio temporal

4.5.2.1. Energía

De manera similar al enfoque aplicado en el dominio transformado, en el caso del dominio temporal también es necesario llevar a cabo una breve caracterización de los audios portada utilizados, ya que éstos influyen en gran medida en la imperceptibilidad del audio secreto incrustado, esto, de acuerdo a los valores de SNR obtenidos para el *método uno temporal* (ver Apéndice C, Tabla C.1), en donde se observa que dichos valores son totalmente opuestos con los valores obtenidos en el *método uno* del dominio transformado; es apreciable que para los géneros musicales en donde los valores de SNR son máximos en el dominio temporal, resultan mínimos en el dominio transformado, por lo cual se recurre a hacer un análisis de la energía[8] presente en las señales de audio portada, obteniendo así los resultados mostrados en la Figura 4.18.

[8]La implementación que se realiza de *audioread* en MATLAB, garantiza la lectura de los audios acotándolos entre -1 y 1, dando paso a una comparación justa de energía entre las señales.

A partir de la Figura 4.18 es posible evidenciar que el género de música Clásica presenta la menor acumulación de energía en comparación a géneros como Pop y Urbano, los cuales a su vez destacan por su imperceptibilidad frente a la incrustación de información. Esto debido a que estas señales tienen una mayor capacidad para encubrir cambios sin ser percibidos, puesto que las alteraciones en fragmentos de alta energía[9] pueden pasar desapercibidas, lo que lleva a un mejor enmascaramiento de la información secreta, lo cual resulta conveniente al implementar este método.

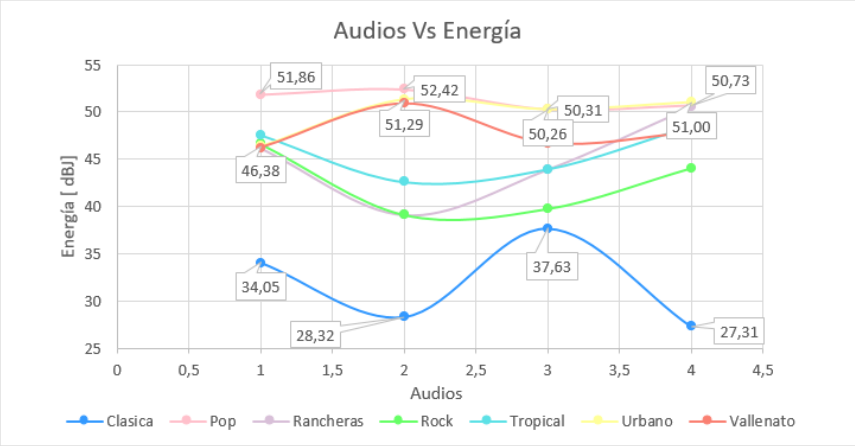


Figura 4.18: *Energía [dBJ] por Género musical.*

Asimismo, en la Figura 4.19 se presenta el promedio de energía por cada uno de los géneros musicales empleados, respaldando lo expuesto anteriormente, en donde la música Clásica presenta la menor acumulación de energía con $31,82$ *dBJ*, mientras que el Pop destaca con $51,33$ *dBJ* en promedio, seguido de la música Urbana con $49,73$ *dBJ*.

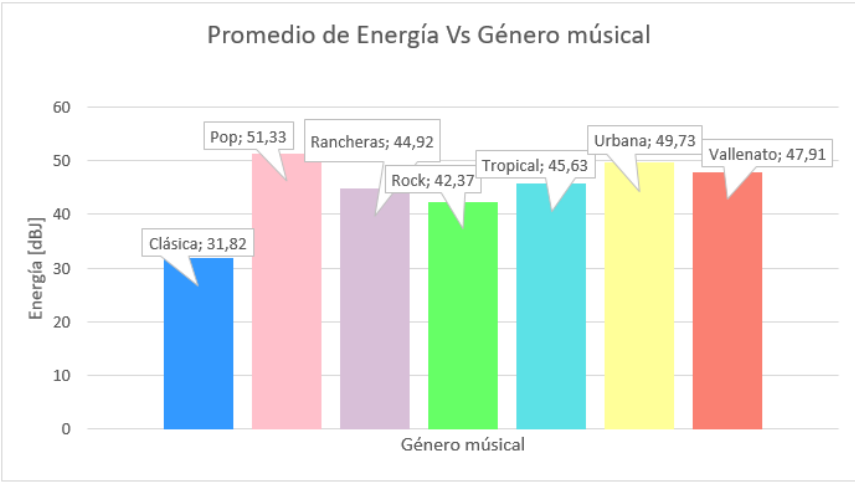


Figura 4.19: *Promedio de Energía [dBJ] por Género musical.*

[9]Entiéndase como distorsión la percepción de un piso de ruido el cual se disimula en señales de alta energía, i.e., valores altos de amplitud, que pueden encubrir con mayor efectividad este piso de ruido.

4.5.2.2. Desviación estándar

Al igual que en el dominio transformado, en el dominio temporal se calcula la desviación estándar de cada una de las señales de audio portada, obteniendo así la dispersión de las amplitudes con respecto a su media, permitiendo identificar algunas características como la presencia de sonidos con alta variabilidad, lo cual se refleja en una desviación estándar elevada o, en caso contrario, señales de audio más constantes, en donde se obtienen desviaciones estándar mucho más bajas. En la Figura 4.20 se presenta el promedio de la desviación estándar por cada género evaluado, en donde de forma similar al dominio transformado destacan el género de música Clásica por su desviación estándar más baja y el género Pop por presentar la desviación estándar más elevada. Sin embargo, como se mencionó anteriormente, son las canciones de este último las que muestran un mejor desempeño como audios portada en el dominio temporal, por lo cual se infiere que la presencia de sonidos altamente variables resultan convenientes en este método apoyando así el análisis de la energía previamente expuesto.

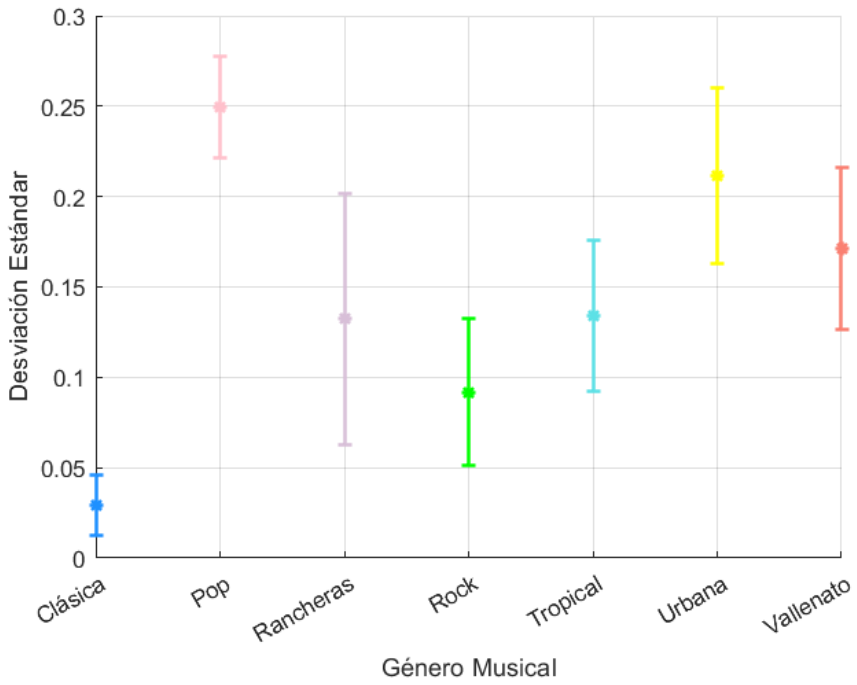


Figura 4.20: *Desviación estándar promedio de la amplitud de muestras por género musical.*

CAPÍTULO 5

ANÁLISIS DE RESULTADOS

En este punto, el algoritmo basado en codificación de fase propuesto cuenta con tres variantes que se han denominado *método uno*, *método tres* y *método cuatro*. Dado que la diferencia entre las tres variantes radica en la manera en que se incrusta el mensaje secreto, cada método cuenta con una capacidad máxima de incrustación predefinida (ver Capítulo 4) y brinda una imperceptibilidad evaluada a través de las medidas objetivas SNR, SSIM y PEAQ-ODG (ver Apéndice A)[10] y subjetivas como MOS y CMOS. Adicionalmente, se cuenta con un versión temporal del primer método, denominado *método uno temporal*, con el que se busca comparar las características esteganográficas de interés.

Para evaluar la capacidad de incrustación e imperceptibilidad, en este capítulo se presentan los resultados de simulación (medidas objetivas) de cada método, para lo cual inicialmente se muestran los resultados obtenidos por cada audio *stego* para una cantidad de información incrustada fija y posteriormente se muestran los resultados de cada método al variar la cantidad de información incrustada, lo cual equivale a aumentar la duración del audio secreto. Seguidamente, se seleccionan los audios *stego* más adecuados para la evaluación de la imperceptibilidad de forma subjetiva, con el fin de complementar el análisis del desempeño de los algoritmos propuestos en términos de esa característica.

Las simulaciones se realizan a partir de 28 audios portada y 5 audios secretos, usando los formatos *MP3* y *WAV*. Cada audio portada cuenta con una duración de 90 segundos, una frecuencia de muestreo de $44,1\ KHz$ y una resolución de 8 bits. Por otro lado, los audios secretos cuentan con una frecuencia de muestreo de $48\ KHz$ y una resolución de 8 bits.

Dado que los *métodos tres y cuatro* requieren de la definición de las bandas de incrustación y número de bits LSB a modificar, en la Tabla 5.1 se presentan éstos parámetros.

[10] Gracias a las pruebas preliminares realizadas en la implementación se evidenció que la SNR como medida objetiva puede tener diferencias con las evaluaciones subjetivas, así que para las pruebas de este capítulo se consideró necesario complementar los resultados objetivos con otras medidas que permitan estimar la similitud entre los audios portada y *stego*, buscando encontrar una medida más cercana a las evaluaciones subjetivas.

Tabla 5.1: Análisis de resultados. Parámetros implementados en *métodos tres y cuatro*.

Método	Banda [Hz]	Número de LSB
Tres	0 - 200	6
	500 - 800	4
	10.000 - 22.050	6
Cuatro	0 - 1.000	3
	1.000 - 2.000	2
	2.000 - 5.000	2
	5.000 - 7.000	4
	7.000 - 22.050	6

5.1. Capacidad de Incrustación

Cada método cuenta con una capacidad de incrustación máxima propia de su diseño, la cual se puede calcular a partir de las Ecuaciones planteadas en el Capítulo 4. Tomando como referencia la duración del audio portada, frecuencias de muestreo de los audios portada y secreto, y las bandas de incrustación indicadas en la Tabla 5.1, en la Tabla 5.2 se muestran la capacidad máxima de cada método.

Tabla 5.2: Capacidad de incrustación máxima por método.

Dominio	Método	Capacidad de Incrustación máxima [seg]	Capacidad de Incrustación máxima [bits]
Frecuencia	*uno*	$36,1758$	$13'891.500$
	tres	$18,334$	$7'040.250$
	cuatro	$24,346$	$9'349.200$
Tiempo	*uno*	$72,321$	$27'783.000$

De forma general, el *método uno temporal* cuenta con la mayor capacidad de incrustación, siendo el doble del permitido en el *método uno*, tal y como se deduce a partir de las Ecuaciones 4.1 y 4.12; dado que es necesario garantizar la simetría impar del espectro de fase cuando se construye el espectro del audio *stego* resultante, la capacidad en el dominio de la frecuencia como mínimo se reduce en un $50\,\%$ en comparación con el *método uno temporal*.

Además, de las variantes del algoritmo de codificación de fase, el *método tres* es la variante con menor capacidad de incrustación, y esto se debe a la ausencia de incrustación de información secreta en dos bandas ($200Hz$ - $500Hz$ y $800Hz$ - $10.000Hz$), lo cual disminuye la cantidad de información secreta que es posible adicionar en el audio portada, dando una reducción del $74,661\,\%$ en comparación al *método uno temporal*.

5.2. Resultados Objetivos

Las métricas de evaluación objetivas elegidas pretenden determinar de forma cuantitativa la imperceptibilidad de la información secreta incrustada en un audio *stego*, con lo cual se busca estimar el desempeño de los diferentes métodos al mantener o variar la cantidad de información secreta.

5.2.1. Desempeño del Algoritmo: Cantidad Fija de Información Secreta

Idealmente las medidas objetivas deberían proporcionar una correcta valoración sobre la imperceptibilidad del audio *stego*; no obstante, se debe tener presente que las alteraciones realizadas (ya sea en el tiempo o la frecuencia) pueden o no ser perceptibles para el HAS, lo que conlleva a validar los resultados mediante las medidas subjetivas MOS y CMOS. Para los resultados de simulación expuestos en esta sección, se realiza la incrustación de 7 segundos de audio secreto en cada audio portada, que equivale a la adición de $2'688.000$ bits de información secreta.

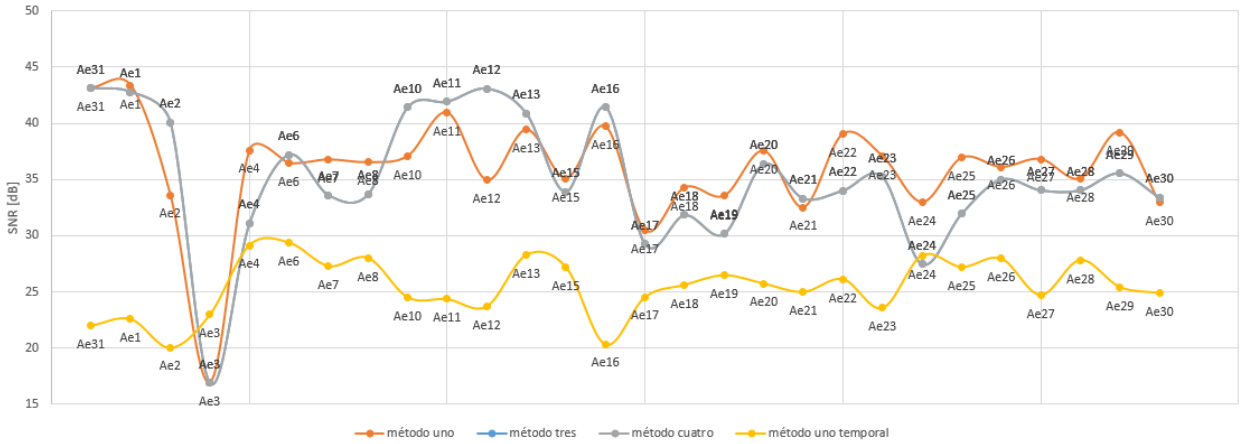


Figura 5.1: *Métrica de evaluación SNR para los audios stego de las versiones del algoritmo propuesto.*

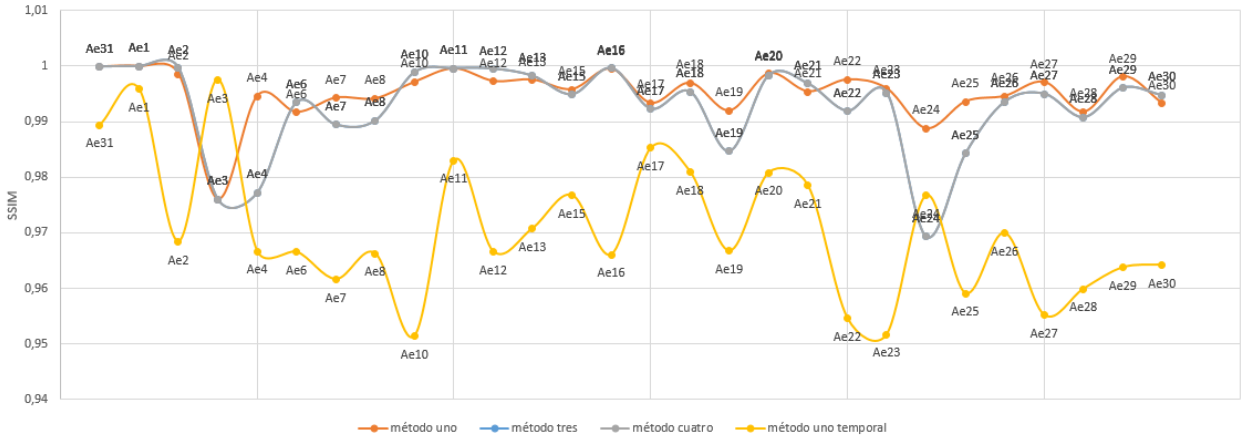


Figura 5.2: *Métrica de evaluación SSIM para los audios stego de las versiones del algoritmo propuesto.*

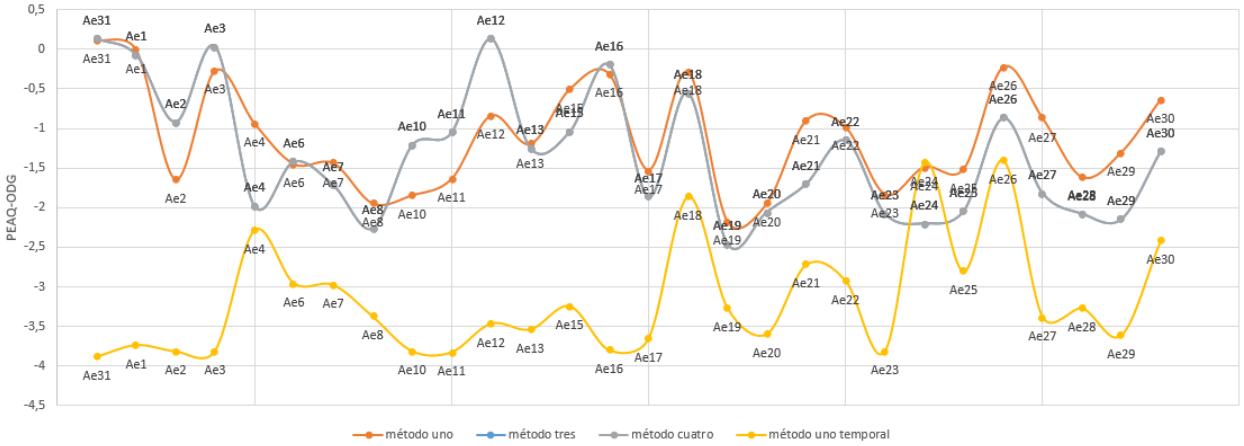


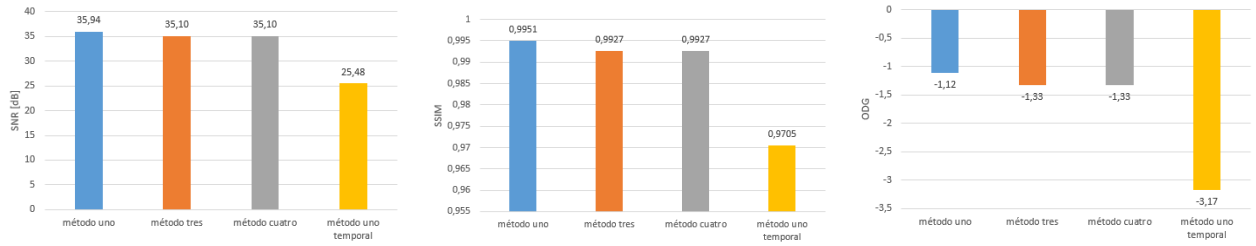
Figura 5.3: *Métrica de evaluación PEAQ-ODG para los audios stego de las versiones del algoritmo propuesto.*

El desempeño de las versiones del algoritmo es equivalente en las Figuras 5.1, 5.2 y 5.3, donde se exponen los resultados obtenidos por los cuatro métodos comparados para cada uno de los 28 audios portada considerados. Es importante resaltar que los *métodos tres y cuatro* mantienen los mismos valores de SNR, SSIM y PEAQ-ODG en todos los casos. Adicionalmente, ambos métodos alternan sus resultados junto al *método uno*, por lo cual se recurre a una evaluación porcentual para determinar el número de veces que el *método uno* presenta un mejor desempeño que los *métodos tres y cuatro*, de esta forma, para la SNR y el SSIM se obtiene un $64,28\,\%$, y para el PEAQ-ODG un $71,42\,\%$.

Sin embargo, aunque las versiones propuestas de incrustación en el dominio de la frecuencia no brindan un campeón incontrovertible, sí se puede concluir que los métodos esteganográficos de codificación de fase superan la imperceptibilidad del *método uno temporal*, puesto que aproximadamente el $92,85\,\%$ de los audios *stego* generados

mediante codificación de fase superan los niveles de imperceptibilidad de los audios generados mediante el método temporal, lo cual lleva a determinar que el método *low bit encoding* en el dominio del tiempo genera alteraciones en el audio portada que son más propensas a ser descubiertas por el HAS.

En la Figura 5.4 es notorio como el promedio de SNR de los *métodos uno, tres y cuatro* superan en $9,62$ *dB* al *método uno temporal*; así mismo, el SSIM del *método uno temporal*, aunque es mayor al $90\,\%$, se encuentra por debajo de los demás métodos como mínimo por un $2,21\,\%$, diferencia que, si se emplea el PEAQ-ODG, es más sustancial, dado que el *método uno temporal* con un valor promedio de $-3,17$, es el único con una calidad considerada como *irritante* (-3).



(a) SNR (b) SSIM (c) PEAQ-ODG

Figura 5.4: *Promedio de métricas objetivas obtenidas en cada método.*

Dado el diseño de cada versión del algoritmo, el *método uno* calcula el número de LSB que deben ser modificados para incrustar el mensaje secreto, a diferencia de los *métodos tres y cuatro* que cuentan con un número de LSB por banda predefinido. Esto hace que los $2'688.000$ bits de información secreta se almacenen en el rango de $12,09KHz$ a $20,05KHz$ con un $n_{LSB} = 3$ para el *método uno* y con un $n_{LSB} = 6$ para los *métodos tres y cuatro*, en la tercera y quinta banda respectivamente, en el rango de $17,07KHz$ a $20,05KHz$ como se muestra en la Figura 5.5; en estas condiciones, las tres versiones brindan resultados muy beneficiosos al compararlos con los dados por el método temporal, dada la modificación de componentes de elevada frecuencia.

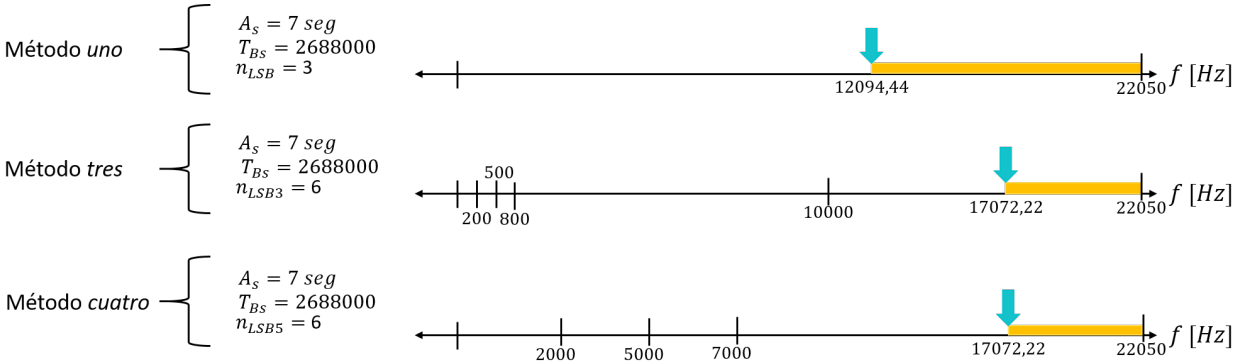


Figura 5.5: *Cantidad fija de información secreta. Componentes de frecuencia alteradas.*

5.2.2. Desempeño del algoritmo: Variación de la Cantidad de Información Secreta

Como se mencionó en la sección 5.1, cada una de las versiones del algoritmo propuesto cuenta con una capacidad de incrustación máxima; sin embargo, es necesario evaluar si el variar la cantidad de información incrustada, por debajo de dichos máximos, altera la imperceptibilidad de los audios *stego*. Para ello, se establecen ocho escenarios distintos de incrustación de audio secreto (ver Tabla 5.3).

Tabla 5.3: Variación de la cantidad de información secreta.

Audio secreto	
Duración [seg]	**Número de bits**
1	384.000
3	$1'152.000$
5	$1'920.000$
7	$2'688.000$
9	$3'456.000$
11	$4'224.000$
13	$4'992.000$
15	$5'760.000$

Variar la cantidad de información secreta repercute en el desempeño de los audios *stego* en cuanto a su imperceptibilidad. Tomando como referencia las Figuras 5.6 y 5.7, los niveles de SNR y SSIM en los *métodos tres y cuatro* se mantienen constantes hasta la incrustación de 9 segundos, aproximadamente la modificación de $3'456.000$ bits de los audios portada, lo cual se convierte en un punto de inflexión puesto que posteriormente los valores de SNR y SSIM decrecen. Por su parte, el *método uno* tiende a mantener los valores de SNR y SSIM constantes hasta los 13 segundos ($4'992.000$ bits); sin embargo, se presentan mínimos en estas medidas en la incrustación de audios secretos de 5 y 7 segundos de duración. En paralelo, el algoritmo temporal presenta un comportamiento monótonamente decreciente, es decir, entre mayor información se incrusta en el audio, la SNR y el SSIM decrecen, indicando un deterioro en la imperceptibilidad del audio *stego*.

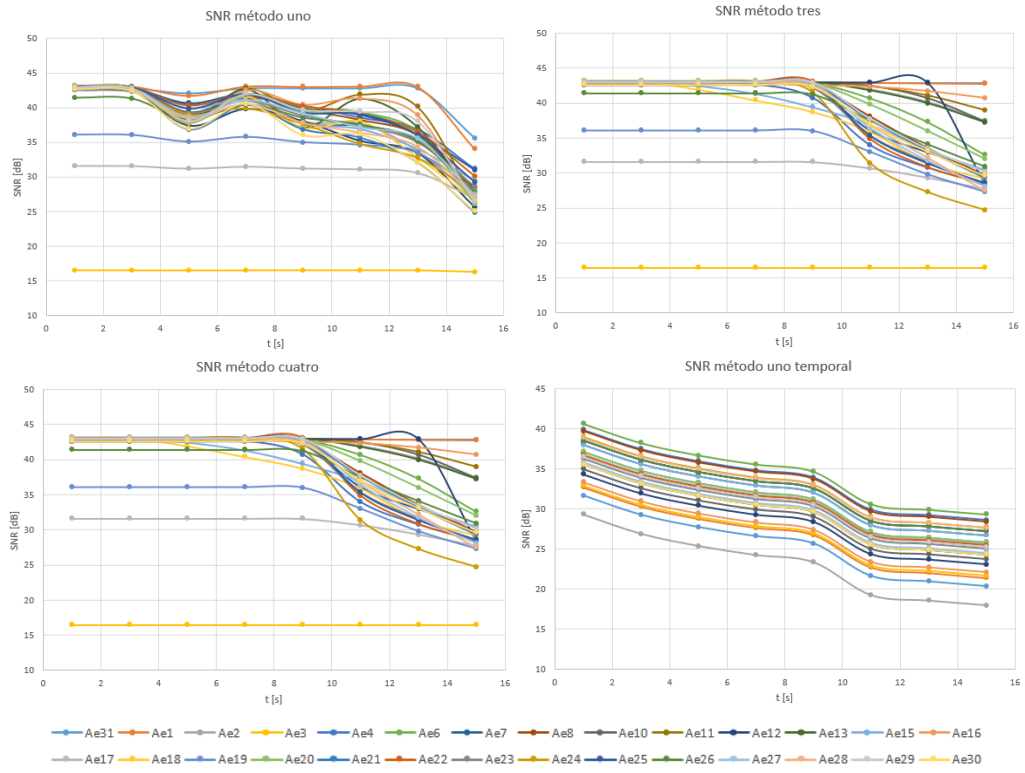


Figura 5.6: *SNR en función de la variación de la cantidad información secreta.*

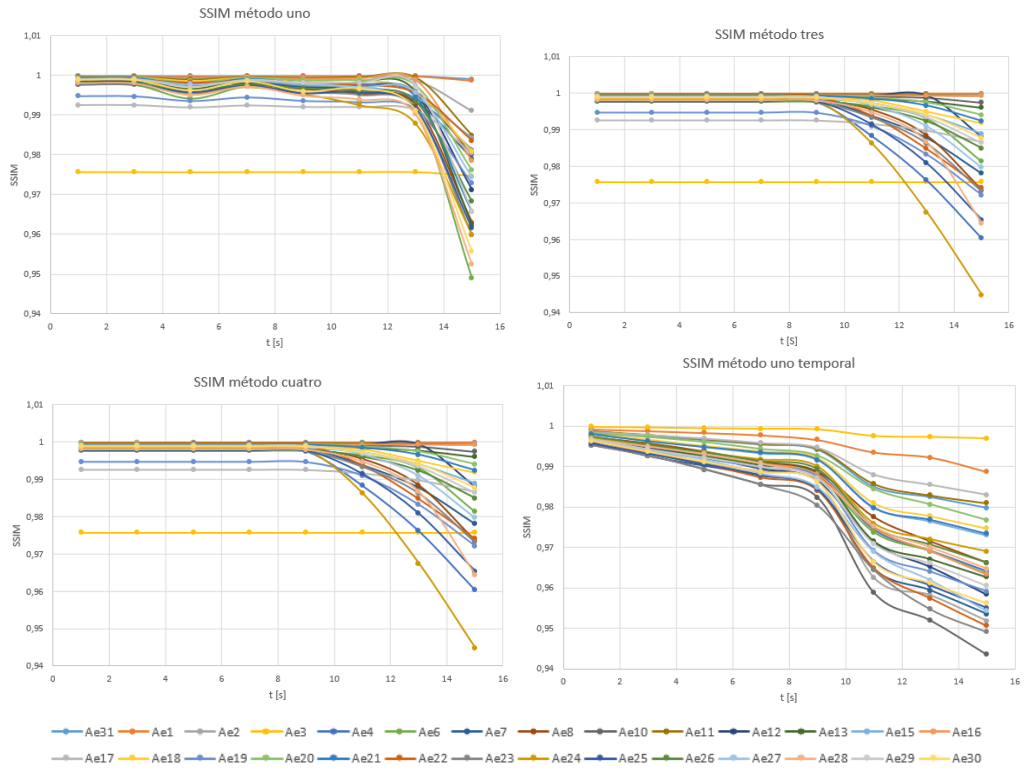


Figura 5.7: *SSIM en función de la variación de la cantidad de información secreta.*

Para el PEAQ-ODG, los *métodos tres y cuatro* mantienen el comportamiento constante hasta los 9 segundos; después la calidad del audio se deteriora considerablemente, aunque la caída en los valores obtenidos depende del audio portada usado en la construcción del audio *stego*. Por otro lado, el *método uno* discrepa con el comportamiento anterior, puesto que presenta un deterioro inmediato de la calidad del audio, el cual incrementa conforme se adiciona mayor información secreta; del mismo modo, el *método uno temporal* presenta un decaimiento de la calidad de los audios *stego*, aunque, a diferencia del *método uno*, la calidad de los audios puede establecerse como un *poco molesto* (-2) o *irritante* (-3), incluso desde la mínima cantidad de información secreta incrustada (1 segundo).

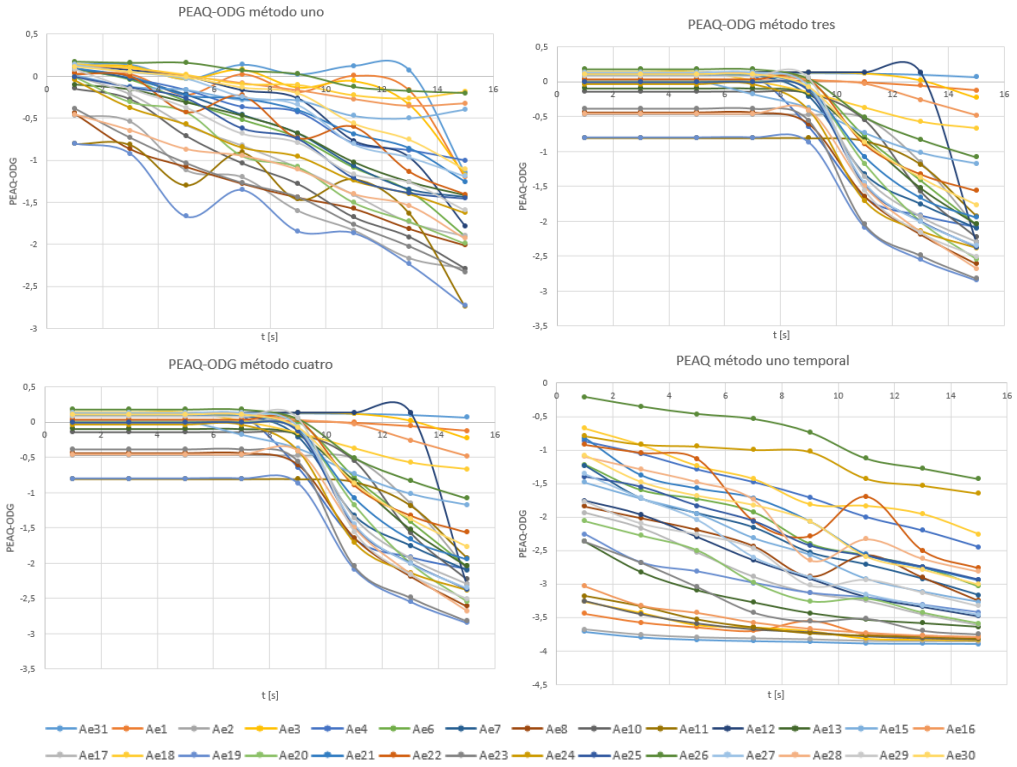


Figura 5.8: *PEAQ-ODG en función de la variación de la cantidad de información secreta.*

El comportamiento constante en las Figuras 5.6 y 5.7 se atribuye a la modificación de la fase de las componentes de alta frecuencia, que en el caso del *método uno* corresponden a frecuencias que se encuentran entre $3,56KHz$ y $22,05KHz$, mientras que para los *métodos tres y cuatro* se perturban en el rango de $15,65KHz$ a $22,05KHz$. Particularmente *el método uno* presenta dos mínimos en las medidas a los 5 y 9 segundos, debido a que para estos casos los valores de n_{LSB} son iguales a 1 y 2, dando paso a la modificación de las componentes en el rango $716,66Hz - 22,05KHz$ y $2,85KHz - 22,05KHz$, respectivamente, lo que se ve como una alteración directa a la mayoría de componentes cercanas al origen, especialmente en el primer caso. Después de cada mínimo, las medidas de SNR y SSIM se elevan, a los 7 y 11 segundos, debido a que en este punto los n_{LSB}

del método se incrementa a 2 y 3 bits, según corresponda, dando paso a la modificación de las componentes de frecuencia entre $7,11KHz$ a $22,05KHz$ y $6,04KHz$ a $22,05KHz$. En la Figura 5.9 es posible ver de forma gráfica dichas particularidades.

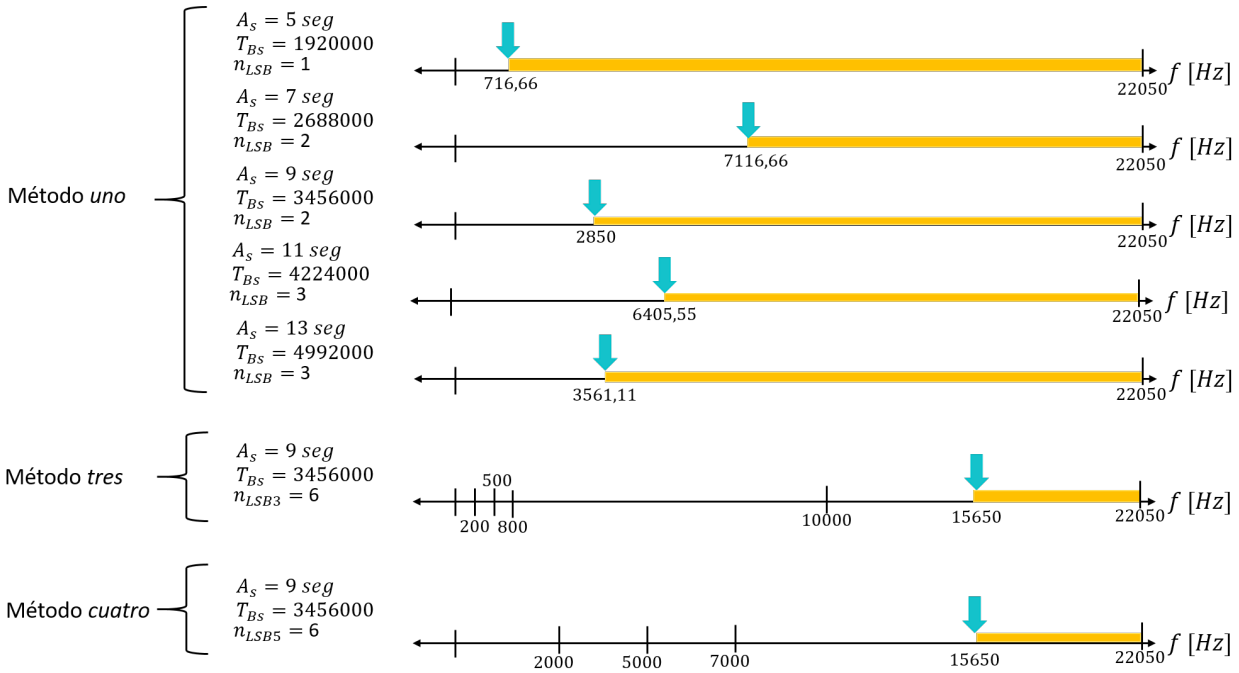


Figura 5.9: *Variación de información secreta. Alteración de componente de baja frecuencia en método uno.*

Dado que los *métodos tres y cuatro* cuentan con mayor rigurosidad en cuanto al n_{LSB} de cada banda (n_{LSB} fijos) y éstos son elevados en las componentes de alta frecuencia, para los casos planteados en la Tabla 5.3, la información secreta no es la suficiente para modificar las componentes de baja frecuencia (el mensaje secreto es menor a la capacidad de incrustación máxima brindada por los métodos), generando un comportamiento constante de la calidad de audio hasta una duración de audio secreto de 9 y un decremento de su calidad para las incrustaciones de los audios secretos con 11, 13 y 15 segundos de duración. Cabe resaltar que a los 9 segundos, en ambos métodos se alteran las componentes de frecuencia entre $15,65KHz$ y $22,05KHz$, con un $n_{LSB} = 6$ bits; mientras que para los 11 segundos, el rango modificado es de $14,23KHz$ a $22,05KHz$, lo cual respalda que las alteraciones en componentes de alta frecuencia tienden a ser menos perceptibles para el HAS.

5.3. Resultados Subjetivos

Ya que es necesario validar los resultados objetivos del algoritmo de esteganografía de audio diseñado mediante resultados subjetivos (MOS, CMOS), se proponen dos evaluaciones con el propósito de medir el desempeño del algoritmo, a partir de la capacidad

de incrustación e imperceptibilidad de la información secreta en audios *stego*. Para ello, se cuenta con una muestra poblacional de 32 participantes con edades entre 19 y 29 años, para abarcar un mayor rango de frecuencias auditivas. Adicionalmente, cabe mencionar que el desarrollo de las pruebas se ejecutan bajo un entorno en las mismas condiciones ambientales, asegurando la incrustación del mismo audio secreto, el uso de los mismos audífonos y nivel de volumen del audio para todos los participantes, garantizando así que los resultados obtenidos sean lo más comparables posibles y que la variabilidad en éstos se deba principalmente a las diferencias de los parámetros evaluados, en lugar de condiciones externas.

5.3.1. Desempeño del algoritmo: Comparación entre variantes

Las pruebas se realizan siguiendo los casos establecidos en las pruebas objetivas; sin embargo, es necesario ejecutar una primera evaluación para estimar el rendimiento de las distintas variantes del algoritmo y determinar, así, cuál presenta un mejor desempeño en cuanto a imperceptibilidad. Esta depuración inicial es necesaria para asegurar que la siguiente evaluación no sea tan extensa, ya que no sería necesario calificar tantos audios *stego* por método. Así, en la primera evaluación, se estima la calidad de los audios mediante la MOS, variando la cantidad de información secreta (7, 9 y 13 segundos) y el género musical de los audios *stego* para cada uno de los tres métodos diseñados (*método uno*, *tres*, *cuatro*) y el *método uno temporal*. Esto permite medir la imperceptibilidad de cada uno de los métodos por separado frente a distintos valores de incrustación. Los escenarios propuestos se exponen en la Tabla 5.4.

Tabla 5.4: Parámetros para las pruebas subjetivas al variar la cantidad de información secreta.

Cantidad de información secreta incrustada por género musical [seg]							
Género musical	Pop *(Ae4)*	Tropical *(Ae19)*	Urbana *(Ae26)*	Ranchera *(Ae13)*	Vallenato *(Ae30)*	Clásica *(Ae2)*	Rock *(Ae17)*
Método *Uno*	7	7	7	9	9	13	13
Método *Tres*	7	7	7	9	9	13	13
Método *Cuatro*	7	7	7	9	9	13	13
Método *Uno Temporal*	7	7	7	9	9	13	13

Cada participante, posterior a la escucha de los audios, asigna una evaluación de acuerdo a la calidad percibida, siguiendo una escala de 1 *(muy mala)* a 5 *(muy buena)* como se muestra en la Figura 5.10.

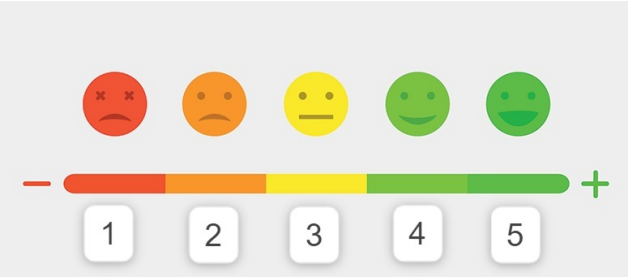


Figura 5.10: *Escala de MOS usada en la evaluación de versiones del algoritmo.*

Los resultados promedio obtenidos para cada uno de los métodos se muestran en la Figura 5.11, en donde es posible observar que el *método cuatro* obtuvo la mejor calificación, con una MOS de $4,1308$ seguido del *método tres* con una MOS de $4,0709$, el *método uno* con $3,9285$ y finalmente el *método uno temporal* en último lugar con un MOS de $3,1547$.

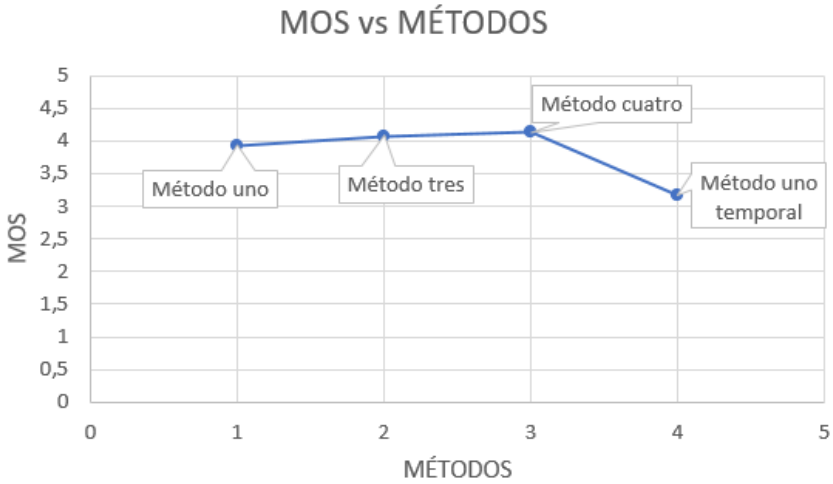


Figura 5.11: *Resultados de la prueba de desempeño de los métodos propuestos vs MOS.*

De forma similar se procede a realizar la comparación entre cuatro audios portada originales y cuatro audios *stego* respectivamente (ver Tabla 5.5), seleccionados bajo el criterio del mejor resultado de SNR obtenido en cada método. Los audios *stego* contienen la misma cantidad de información secreta (7 segundos), y representan cada una de las tres variantes del algoritmo (*método uno*, *tres* y *cuatro*) y el *método uno temporal*. Los participantes asignan un valor numérico de 1 a 5, donde 3 indica que los audios se escuchan *aproximadamente igual*, 1 o 2 si el audio *stego* se escucha *mucho peor* o *peor* que el audio original y 4 o 5 si consideran que el audio *stego* se escucha *mejor* o *mucho mejor* que el original, siguiendo la métrica CMOS que se ilustra gráficamente en la Figura 5.12.

Tabla 5.5: Parámetros para pruebas subjetivas de comparación.

Método	**Audio Portada**	**Audio *stego***
Método *Uno*	*Ap1*	*Ae1*
Método *Tres*	*Ap31*	*Ae31*
Método *Cuatro*	*Ap31*	*Ae31*
Método *Uno Temporal*	*Ap6*	*Ae6*

Figura 5.12: *Escala de CMOS usada en la evaluación de versiones del algoritmo.*

Los resultados obtenidos se muestran en la Figura 5.13, en donde el *método uno* y el *método uno temporal*, obtuvieron una calificación de CMOS inferior a $2,8$ situándolos en la escala de *peor* (2) frente a la calidad del audio *stego* comparado con la del audio portada, mientras que los *métodos tres* y *cuatro* sobresalen con valores de CMOS mayores a 3, ubicándolos en la escala de *aproximadamente igual* frente a la misma comparación, lo que indica que la calidad del audio *stego* es similar a la del audio portada.

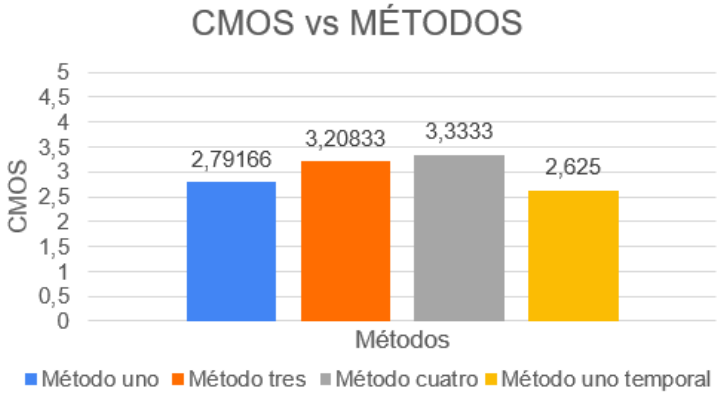


Figura 5.13: *Resultados de la prueba de desempeño de los métodos propuestos vs CMOS.*

Estos resultados, junto con los resultados objetivos expuestos en la sección 5.2, dan paso a que los *métodos tres y cuatro* sean considerados como aquellos que, para una

cantidad de información secreta a incrustar dada, ofrecen un mejor desempeño, ya que cuentan con mayor imperceptibilidad en comparación a los demás métodos implementados, lo cual puede deberse principalmente a que en su diseño la incrustación de la información se da en su mayoría en las bandas de frecuencia más elevadas, por lo que estas alteraciones pueden resultar más inapreciables por el HAS. Asimismo, es posible evidenciar que el *método uno temporal* exhibe un menor nivel de imperceptibilidad en comparación con los métodos implementados con base en la codificación de fase, lo cual se ve reflejado en las calificaciones más bajas tanto en MOS como en CMOS.

Con base a estos resultados, el desarrollo de la segunda evaluación de imperceptibilidad, en la cual se mide el desempeño del algoritmo en cuanto a la cantidad fija de información secreta y la variación de la misma, se tiene únicamente en cuenta la implementación del *método cuatro*, ya que aunque su rendimiento es similar al *método tres*, este cuenta con mayor capacidad de incrustación, resultando más idóneo para el desarrollo de las pruebas.

5.3.2. Desempeño del Algoritmo: Cantidad Fija de Información Secreta

Con el fin de verificar el papel que tiene el género musical al cual pertenece el audio portada usado en el proceso esteganográfico, se selecciona un audio *stego* aleatoriamente de cada uno de los siete géneros incluidos en la evaluación (ver Tabla 5.6), garantizando que se han construido a partir de una incrustación de 13 segundos de información secreta mediante el *método cuatro*, esto es, audios *stego* que almacenan $4'992.000$ bits.

Tabla 5.6: Evaluación subjetiva de audios *stego*. Clasificación por géneros.

Género	Clásica	Ranchera	Rock	Tropical	Vallenato	Urbana	Pop
Audio *stego*	*Ae2*	*Ae13*	*Ae17*	*Ae19*	*Ae30*	*Ae26*	*Ae4*

Los participantes de esta segunda evaluación, brindan su calificación respecto a la calidad de cada audio de 1 *(muy malo)* a 5 *(muy bueno)*, dando paso a una MOS por audio *stego* que se expone en la Figura 5.14, la cual evidencia que no sólo la cantidad de información secreta hace o no menos imperceptible un audio *stego*, sino que también la naturaleza del audio portada del cual partió. Además se exponen discrepancias entre las métricas objetivas, la clasificación por género obtenida en la subsección 4.5.1.2 y los resultados subjetivos obtenidos.

En términos generales, la música Clásica sobresale con valores de SNR y MOS elevados, mientras en PEAQ-ODG cuenta con una calidad aproximada a 1 (*perceptible pero no molesto*), destacandose como el género con mejor respuesta ante imperceptibilidad; según los resultados de desviación estándar indicados en la subsección 4.5.1.2, el siguiente género con mejor respuesta corresponde a Rock, pero éste se encuentra en la

tercera posición en términos de MOS. Por otra parte, el género de música Urbana obtuvo una calidad aproximada a 3 (*regular*) la cual ha sido predicha al ocupar el sexto lugar de la clasificación mediante desviación estándar, aunque la PEAQ-ODG indica una calidad de audio cercana a *perceptible pero no molesto* (-1).

Dadas las inconsistencias, se procede a evaluar cual de las métricas o características brindan información más acertada sobre el comportamiento de audios *stego* en términos del género al cual pertenece, deduciendose que la clasificación por desviación estándar es la más cercana a la MOS, dado que los resultados subjetivos indican que los géneros se pueden clasificar en el orden: Clásica, Ranchera, Rock, Vallenato, Tropical, Urbana y Pop; comparándolo con la clasificación brindada por las demás métricas, de tal forma que si el género se encuentra una posición por encima o por debajo de su clasificación real, se asume como relativamente válido. La desviación estándar (ver Figura 4.17) discrepa con los resultados de MOS al intercambiar las posiciones de los géneros Ranchera y Rock, Vallenato y Tropical, lo cual no es un resultado fatídico como la estimación de calidad (PEAQ-ODG) de *Ae26* como *perceptible pero no molesto* (-1). De los resultados es indispensable destacar el buen desempeño de los género de música Clásica, Ranchera y Rock, brindando valores de MOS cercanos a una calidad de 4 $(bueno)$.

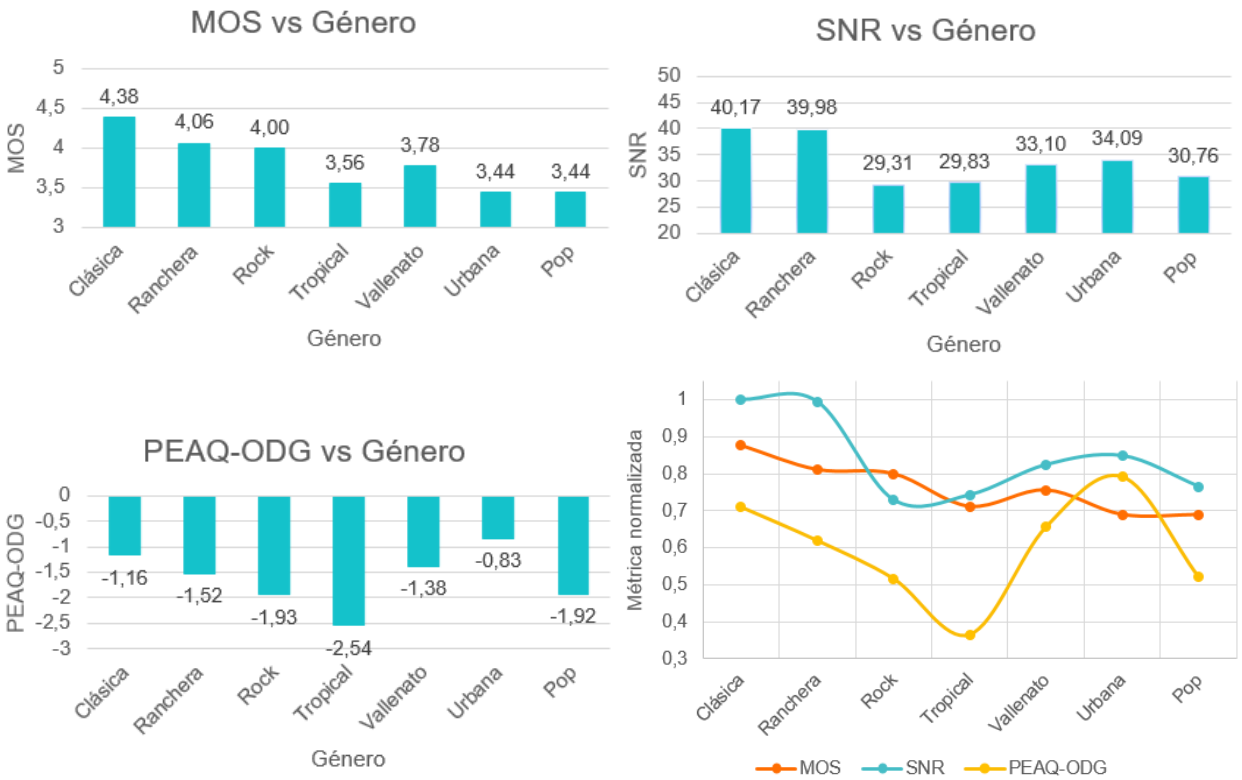


Figura 5.14: *Resultados de la evaluación de desempeño de los géneros utilizados.*

5.3.3. Desempeño del Algoritmo: Variación de la Cantidad de Información Secreta

A partir de la clasificación previamente realizada de los audios portada (sección 4.5), se evalúa el rendimiento del algoritmo con base en la variación de la cantidad de información secreta incrustada, para ello se eligen los géneros musicales con mejor desempeño, siendo éstos, la música Clásica y la Ranchera. Así, para el desarrollo de esta prueba se consideran tres casos de incrustación: 11, 13 y 15 segundos representados por $4'224.000$, $4'992.000$ y $5'760.000$ bits respectivamente. Los escenarios implementados se ilustran en la Tabla 5.7.

Tabla 5.7: Evaluación subjetiva de audios *stego*. Variación en la cantidad de información secreta

Escenario	**Audio Portada**	**Audio *stego***	**Cantidad de información secreta incrustada [seg]**
Uno	*Ap31*	*Ae31*$_1$	11
Dos		*Ae31*$_2$	13
Tres		*Ae31*$_3$	15
Cuatro	*Ap13*	*Ae13*$_1$	11
Cinco		*Ae13*$_2$	13
Seis		*Ae13*$_3$	15

Para evaluar subjetivamente la imperceptibilidad de la información secreta, a medida que se incrementa su valor en los audios *stego*, se utiliza la métrica de evaluación CMOS (ver Figura 5.12). Los participantes deben comparar la similitud entre la calidad de los audios portada y los audio *stego* a escuchar basándose únicamente en la calidad percibida de los audios, de esta manera una calificación de 3 indica una similitud *aproximadamente igual*, mientras que una calificación de 1 o 2 sugiere una similitud *mucho peor* o *peor* y 4 o 5 refleja una calidad superior del audio *stego* en comparación con el audio portada *(mejor o mucho mejor)*. Los resultados obtenidos se muestran en la Figura 5.15.

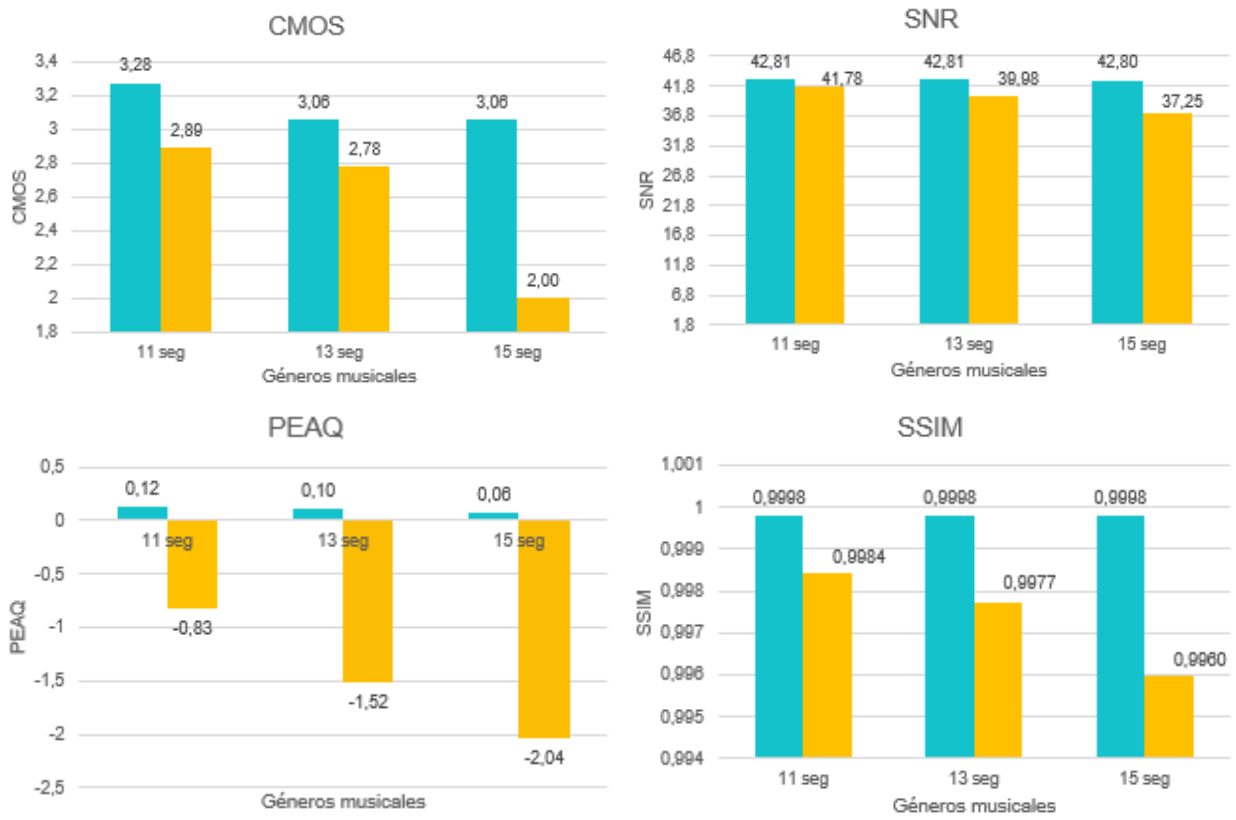


Figura 5.15: *Resultados de la evaluación de desempeño ante la variación de la cantidad de información secreta.*

En promedio las puntuaciones obtenidas por los participantes es consistente con el desempeño de los audios portada, como se había inferido en la clasificación por géneros musicales. En general, la música Clásica (de color azul, ver Figura 5.15) bajo los tres escenarios propuestos obtiene resultados iguales o ligeramente superiores a 3, situándose sobre la escala de *aproximadamente igual* frente a la comparación del audio portada original versus los audios *stego*, indicando así que pese a que se incrementa la cantidad de información secreta en este tipo de audios portada la imperceptibilidad se mantiene estable ante el HAS. Asimismo, estos resultados son consistentes con los obtenidos a partir de las métricas de evaluación objetivas, ya que tanto para la SNR como para el SSIM mantienen valores superiores a los 40 *dB* y muy cercanos a 1 respectivamente, además, la métrica PEAQ-ODG permanece cercana a cero en los tres casos,mostrando así el buen desempeño del algoritmo al implementar este tipo de señales como medios portada.

Por su parte, el género musical Ranchera (de color amarillo, ver Figura 5.15) obtiene un desempeño menor en comparación con la música Clásica en términos de imperceptibilidad; en este caso las puntuaciones de CMOS se encuentran en la escala de *peor*, con valores que van desde $2,89$ hasta 2 respectivamente, es decir, la imperceptibilidad en

este tipo de señales es más baja entre mayor cantidad de información secreta se incruste en el audio portada. Estos resultados, al igual que en el caso anterior, se mantienen consistentes con los resultados objetivos, ya que, aunque los valores de SNR sean altos, manteniéndose por encima de los 35 *dB*, los valores de PEAQ-ODG alcanzan valores bajos que llegan a los $-2,04$, situando al audio *stego* con mayor información secreta en las escala de *ligeramente molesto*; de igual forma los valores de SSIM presentan diferencias más notorias en comparación a la música Clásica, alejándose del valor deseado (1) a medida que incrementa la longitud del audio secreto. Sin embargo, cabe resaltar que aunque el rendimiento de este último género musical sea inferior al de la música Clásica, no puede decirse que es inadecuado para una implementación esteganográfica, ya que en ninguno de los casos alcanza la escala de *irritante o muy molesto*.

CAPÍTULO 6

CONCLUSIONES Y TRABAJOS FUTUROS

6.1. Conclusiones

- Para desarrollar un algoritmo esteganográfico efectivo, es esencial inicialmente identificar el dominio y las características esteganográficas prioritarias, lo que proporciona la base necesaria para definir los procesos y parámetros adecuados. Un diseño nítido y preciso del algoritmo no sólo simplifica su implementación, sino que también posibilita la identificación de los parámetros cruciales que, de acuerdo con los criterios establecidos, dificultan la detección y extracción del mensaje secreto por parte de posibles intrusos.

- Es innegable que la capacidad de incrustación en un audio portador tiende a aumentar a medida que su duración se extiende. Sin embargo, también es importante destacar que existen otras características del audio portador que, al ser contrastadas con las del audio secreto, pueden contribuir a una mayor capacidad de incrustación. Ejemplos de estas características incluyen la frecuencia de muestreo y la resolución en bits por muestra, las cuales deben tener un valor más elevado en el audio portador para lograr este objetivo de manera efectiva.

- Un método de incrustación temporal, como el *método uno temporal*, cuenta con una capacidad de incrustación mayor que la ofrecida por un método de incrustación en el dominio de la Trasformada de Fourier; sin embargo, usar una transformación garantiza cambios menos abruptos en los archivos de audio *stego*, reduciendo la posibilidad de que la información secreta sea detectada por el HAS. Además, brinda cierto grado de seguridad ante la extracción de la información secreta, puesto que ésta sólo se puede recuperar siempre y cuando se reconozcan los parámetros usados en el transmisor, especialmente en los *métodos tres y cuatro*.

- La modificación de la fase de las componentes de frecuencia cercanas al origen ejerce un impacto significativo en el rendimiento del algoritmo en términos de imperceptibilidad. Por lo tanto, es beneficioso establecer previamente los límites de alteración en las bandas de frecuencia más sensibles para el (HAS). Es por esta razón que el *método cuatro* destaca como la versión más adecuada del algoritmo propuesto. En términos generales, la modificación de las componentes de fase de alta frecuencia, superiores a $10KHz$, asegura una mayor imperceptibilidad según el HAS; sin embargo, es importante tener en cuenta que la capacidad auditiva individual puede variar en términos de frecuencia.

- El audio portada seleccionado para el proceso esteganográfico basado en codificación de fase es relevante para la imperceptibilidad del audio *stego* ante posibles observadores dado que sus características intrínsecas pueden o no contribuir a que el audio *stego* pase desapercibido ante el HAS, por lo que es recomendable reconocer la importancia del género musical al cual pertenece el audio: los géneros con mejor desempeño en esta evaluación poco rigurosa son Clásica, Ranchera y Rock.

- La energía acumulada en las componentes de baja frecuencia de un audio portada, vista a través de espectrogramas, confiere ventajas significativas a la esteganografía basada en codificación de fase. Además, la poca variabilidad en los espectros de magnitud de las señales también resulta beneficiosa, ya que la uniformidad y consistencia dificultan la detección del mensaje secreto por parte del HAS. Asimismo, es importante destacar que la naturaleza de la información secreta, ya sea voz, audio u otros tipos representados en forma de bits, no tiene un impacto significativo en la imperceptibilidad del audio *stego*.

6.2. Trabajos Futuros

- Definir rigurosamente todas las características y parámetros que hacen de un audio el mejor candidato para un proceso esteganográfico basado en codificación de fase. Evaluar su veracidad y establecer un método o metodología ideal para su caracterización.

- Establecer las componentes de fase límite para la cual los valores de LSB del *método uno* no generen degradaciones abruptas a lo largo de la variación de información secreta incrustada.

- Variar los límites de las bandas de los *métodos tres y cuatro*, además del número de bits admitidos por banda, con el fin de establecer los valores óptimos para los parámetros que permitan obtener los mejores resultados en cuanto a capacidad de incrustación e imperceptibilidad.

- Establecer el límite máximo de la resolución para el espectro de fase del audio portada en comparación con la resolución utilizada para la cuantificación del audio secreto, que permita incrementar la capacidad de incrustación del algoritmo sin perjudicar la imperceptibilidad o introducir distorsión.

- Utilizar señales de voz como medio portada en un proceso esteganográfico de audio, para comparar así su desempeño frente a las señales de música.

- Diseñar e implementar un algoritmo de esteganografía de audio empleando otro dominio transformado.

- Diseñar e implementar un algoritmo de estegoanálisis que evalúe el rendimiento del algoritmo de esteganografía de audio basado en codificación de audio propuesto, en términos de su robustez ante intentos de detección y capacidad para preservar la integridad del mensaje secreto.

REFERENCIAS BIBLIOGRÁFICAS

[1] UIT-T, “Subjective performance assessment of telephone-band and wideband digital codecs,” Recomendación P.830, Unión Internacional de Telecomunicaciones, Ginebra, Feb. 1996.

[2] J. M. Ramírez, H. A. Romo, and M. M. Silva, *Telecomunicaciones Digitales*. Universidad del Cauca, first ed., 2020.

[3] C. J. Carrillo, *Fundamentos del Análisis de Fourier*. Universidad de Vigo, 2003.

[4] Z. Fan, *Seminar Notes: The Mathematics of Music*. 2010.

[5] B. Gold, N. Morgan, and D. Ellis, *Speech and Audio Signal Processing*. Wiley, second ed., 2011.

[6] A. Khetrapal, “How does the ear work?,” *News-Medical.net*, Apr. 2017. Section: Health.

[7] RAE, “audio.” `https://dle.rae.es/audio`, 2022.

[8] A. García, *Técnicas de control de sonido en directo*. Ediciones Paraninfo, S.A., Mar. 2017.

[9] P. S. Sundar, C. Chowdhury, and S. Kamarthi, “Evaluation of human ear anatomy and functionality by axiomatic design,” *Biomimetics*, vol. 6, no. 2, 2021.

[10] I. Cobeta, F. Núñez, and S. Fernández, *Patología de la Voz*. Marge Médica Books, first ed., 2013.

[11] T. Ogunfunmi, R. Togneri, and M. Narasimha, *Speech and Audio Processing for Coding, Enhancement and Recognition*. Springer, 2015.

[12] UIT-T, “Pulse Code Modulation (PCM) of voice frequencies,” Recommendation G.711, International Telecommunication Union, Ginebra, June 1990.

[13] UIT-T, “Códec de voz de doble velocidad para la transmisión en comunicaciones multimedia a 5,3 y 6,3 kbit/s,” Recomendación G.723.1, Unión Internacional de Telecomunicaciones, Ginebra, June 2006.

[14] UIT-T, “Modulación por impulsos codificados diferencial adaptativa (MICDA) a 40, 32, 24, 16 Kbit/s. Anexo B: Formato de paquete, identificador de capacidad y parámetros de capacidad para la señalización H.245,” Recomendación G.726, Unión Internacional de Telecomunicaciones, Ginebra, July 2003.

[15] V. Melchoir, “High resolution audio: A history and perspective,” May 2019.

[16] Google Cloud, “Introduction to audio encoding.” https://cloud.google.com/speech-to-text/docs/encoding#uncompressed_audio.

[17] F. Posada Prieto, *Multimedia y Web 2.0*. Ministerio Eduacación de España, first ed., 2012.

[18] T. Connaghan, “Formatos de archivos de audio.” https://emastered.com/es/blog/audio-file-formats, oct 2022.

[19] F. Djebbar, B. Ayad, K. A. Meraim, and H. Hamam, “Comparative study of digital audio steganography techniques,” *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2012, Dec. 2012.

[20] G. A. Bernal Patiño, “Recuantificación de la fase de una señal de audio para el ocultamiento de información,” Instituto Politécnico Nacional. Ciudad de México. Aug. 2019.

[21] A. A. Alsabhany, F. Ridzuan, and A. H. Azni, “The Adaptive Multi-Level Phase Coding Method in Audio Steganography,” *IEEE Access*, vol. 7, pp. 129291–129306, 2019.

[22] M. Nutzinger and J. Wurzer, “A Novel Phase Coding Technique for Steganography in Auditive Media,” in *Proc. 2011 Sixth International Conference on Availability, Reliability and Security*, pp. 91–98, IEEE, Aug. 2011.

[23] P. Jayaram, H. R. Ranganatha, and H. S. Anupama, “Information Hiding Using Audio Steganography - a Survey,” *The International Journal of Multimedia & Its Applications (IJMA)*, vol. 3, no. 3, 2011.

[24] N. Parab, M. Nathan, and K. Talele, “Audio steganography using differential phase encoding,” *Communications in Computer and Information Science*, vol. 145 CCIS, pp. 146–151, 2011. ISBN: 9783642202087.

[25] J. Riera del Moral, *Estudio de distintos métodos de audio watermarking*. bachelorThesis, June 2020. Accepted: 2021-03-10T09:36:24Z.

[26] S. P. Rajput, K. P. Adhiya, and G. K. Patnaik, “An Efficient Audio Steganography Technique to Hide Text in Audio,” in *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, pp. 1–6, Aug. 2017.

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

[28] J. Delgado, Pablo. Herre, “Can we still use peaq? a performance analysis of the itu standard for the objective assessment of perceived audio quality,” *IEEE Access*, 2020.

[29] UIT-T, “Método para mediciones objetivas de la calidad de audio percibida,” Recomendación BS.1387, Unión Internacional de Telecomunicaciones, 1998.

[30] UIT-T, “Método para mediciones objetivas de la calidad de audio percibida,” Recomendación BS.1387-1, Unión Internacional de Telecomunicaciones, 1998-2001.

[31] S. Stephencwelch, “Perceptual-coding-in-python.” `https://github.com/stephencwelch/Perceptual-Coding-In-Python/tree/master/PEAQPython/PQevalAudioMATLAB`. 2014.

[32] UIT-T, “Method for objective measurements of perceived audio quality,” Recomendación BS.1387-2, International Telecommunication Union, 2023.

[33] M. M. Silva Zambrano, “Cuantificación de señales de audio utilizando wavelets,” p. 59, 2022. Universidad del Cauca. Popayán.

[34] A. Hevner, “A Three Cycle View of Design Science Research,” *Scandinavian Journal of Information Systems*, vol. 19, Jan. 2007.

[35] M. Arias Chaves, “La ingeniería de requerimientos y su importancia en el desarrollo de proyectos de software,” *InterSedes: Revista de las Sedes Regionales*, 2005.

[36] MathWorks, “audioread - read audio file.” `https://la.mathworks.com/help/matlab/ref/audioread.html#btiabil-1-dataType`. 2023a.

[37] MathWorks, “audiowrite - write audio file.” `https://la.mathworks.com/help/matlab/ref/audiowrite.html`. 2023a.

[38] 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)*, pp. 301–304, 2013.

[39] C. S. Hernandez Vargas, “Comunicación encubierta de mensajes de voz dentro de audio,” Universidad Militar Nueva Granada. Bogotá.

[40] A. Alsabhany, F. Ridzuan, and A. Halim, “A hybrid method for data communication using encrypted audio steganography,” *ResearchGate*, 2017.

[41] CiberProtector, “Encriptar y desencriptar ficheros online.” `https://ciberprotector.com/encriptar-desencriptar-ficheros-online/`. Sitio web de CiberProtector.

APÉNDICE A

MÉTRICAS DE EVALUACIÓN

La evaluación de un audio puede hacerse a través de medidas subjetivas u objetivas. Para realizar una evaluación subjetiva es necesaria la participación de diferentes personas que califiquen el audio según una escala predefinida; no obstante, se debe tener en cuenta que en la evaluación subjetiva puede estar condicionada por el estado de ánimo de las personas o factores externos, como los ambientales; es por esto que lo deseable es tener un número significativo de participantes, lo cual hace que estas evaluaciones sean complejas y/o costosas si se desean resultados muy precisos. Por lo anterior, se puede recurrir a una evaluación objetiva, la cual puede apoyar calificaciones subjetivas de una cantidad baja de participantes, obteniendo resultados válidos.

Las métricas subetivas MOS, CMOS y las métricas objetivas BER, SNR, SSIM, PEAQ se han descrito en el Capítulo 2, mientras la métrica M-NRMSE se detalla a continuación.

- **M-NRMSE**: El error cuadrático medio normalizado promedio (M-NRMSE, *Mean Normalized Root Mean Square Error*), calculado mediante la ecuación A.1, es una métrica de error utilizada para medir la calidad de señales de audio, permitiendo cuantificar la diferencia entre dos señales de audio, en este caso una señal original (audio portada) y su versión modificada (audio *stego*); su valor se encuentra acotado en el rango de [0,1], en donde un resultado igual a 1 indica que la señal modificada es muy similar a la original, mientras que un valor cercano a 0 significa que las señales son muy diferentes [38].

$$M - NRMSE = 1 - \frac{1}{2} \cdot \sqrt{\frac{\sum [x[n] - w[n]]^2}{\sum x^2[n]}}, \tag{A.1}$$

  donde $x[n]$ representa la señal de audio portada y $w[n]$ la señal de audio *stego*.

Con el fin de reconocer las métricas objetivas más adecuadas para la evaluación de los audios *stego* generados mediante el algoritmo propuesto, se recurre a comparar el resultado subjetivo MOS obtenido en las pruebas de validación expuestas en el Capítulo 4 y las métricas objetivas SNR, SSIM, PEAQ-ODG y M-NRMSE, tal y como se presenta

en la Tabla A.1.

Tabla A.1: Medida subjetiva y objetivas de audios *stego* resultantes de las pruebas de validación.

Audio	**MOS**	**SNR**	**SSIM**	**M-NRMSE**	**PEAQ-ODG**
Ae27	$4,64$	$32,63$	$0,9931$	$0,9623$	$-3,1760$
Ae11	$3,92$	$26,53$	$0,9893$	$0,9769$	-3.5240
Ae22	$3,85$	$25,45$	$0,9465$	$0,9733$	$-2,8130$
Ae4	$2,85$	$23,92$	$0,8941$	$0,9680$	$-2,5966$
Ae7	$2,92$	$23,32$	$0,9001$	$0,9658$	$-3,0216$
Ae8	$2,5$	$22,24$	$0,8843$	$0,9615$	$-3,5378$
Ae6	$2,57$	$21,94$	$0,8473$	$0,9531$	$-3,2410$
Ae12	$3,07$	$20,53$	$0,9344$	$0,9530$	$-3,5072$
Ae18	$3,71$	$18,55$	$0,9057$	$0,9401$	$-0,9644$
Ae21	$2,21$	$17,20$	$0,8766$	$0,9308$	$-2,8130$
Ae24	$2,21$	$18,32$	$0,7953$	$0,9325$	$-3,1830$

Dada la pequeña variación del M-NRMSE, se descarta de las métricas usadas para la evaluación objetiva de los resultados; con ello, las métricas subjetivas y objetivas implementadas son: MOS, CMOS, SNR, SSIM y PEAQ-ODG.

APÉNDICE B

PRUEBAS PRELIMINARES DE PERCEPCIÓN

Las pruebas preliminares de percepción de los audios *stego*, construidos a partir de los métodos iniciales planteados para el algoritmo de codificación de fase, se realizan en dos secciones: una prueba de audición y una prueba de percepción de calidad de los audios *stego*.

Para la prueba de audición se usan tonos desde los $20KHz$ hasta los $12KHz$ con una intensidad de $-3dB$. Seguidamente, para la prueba de percepción de calidad, se reproducen los 10 audios seleccionados (ver Tabla 4.2 o 4.3), de tal forma que cada participante brinde su respectiva calificación en una escala de 5 a 1, con el objetivo de calcular así el MOS: muy buena (5), buena (4), regular (3), mala (2) o muy mala (1).

Antes de que un participante inicie con la prueba preliminar de percepción, se le indica y acepta los términos expuestos a continuación, de lo contrario no se inicia la prueba.

> **Consentimiento Informado**
> Gracias por participar en la *prueba preliminar de percepción* asociada al trabajo de grado *Algoritmo de Esteganografía de Audio Basado en Codificación de Fase*.
>
> Al registrarse en este documento, usted admite que participa voluntariamente de ésta *prueba preliminar de percepción* guiada por las estudiantes que desarrollan el trabajo de grado ya mencionado.
>
> También, declara que ha sido informado sobre su libertad de detener y/o abandonar la prueba en cualquier momento y por cualquier motivo, si así lo desea. En este caso, los datos recopilados no serán parte de los resultados de la *prueba preliminar de percepción*.

APÉNDICE C

BANCO DE AUDIOS

En este Apéndice se exponen los audios portada utilizados y su respectiva caracterización para el desarrollo del presente trabajo de grado, esto con el objetivo de proporcionar mayor información que contribuya a la comprensión del análisis realizado. En la Tabla C.1 se presenta la nomenclatura de los audios portada, el género músical al que pertenecen y el valor de SNR en dB de los audios *stego* resultantes de cada uno de los métodos, considerando la misma información secreta en todos los casos. Adicionalmente se adjuntan los resultados complementarios del proceso de caracterización de los audios, como los espectrogramas de las señales de audio utilizadas en las pruebas preliminares de percepción, la concentración de PSD en banda base (hasta $3KHz$), al igual que las tablas de resultado que contienen información relevante sobre la desviación estándar en el dominio temporal y transformado.

Tabla C.1: Lista de audios portada y SNR de audios *stego* resultantes de pruebas de validación.

Audio	**Género audio Portada**	**SNR Método uno [dB]**	**SNR Método dos [dB]**	**SNR Método tres [dB]**	**SNR Método cuatro [dB]**	**SNR Método cinco [dB]**
$Ae_{31,1}$	Clásica	$43,145$	$43,102$	$40,575$	$43,137$	$-0,771$
$Ae_{31,2}$	Clásica	$43,145$	$43,101$	$40,543$	$43,137$	$-0,804$
$Ae_{31,3}$	Clásica	$43,145$	$43,108$	$40,859$	$43,138$	$-0,776$
$Ae_{31,4}$	Clásica	$43,145$	$43,104$	$40,621$	$43,137$	$-0,780$
$Ae_{31,5}$	Clásica	$43,145$	$43,104$	$40,640$	$43,137$	$-0,762$
$Ae_{1,1}$	Clásica	$43,330$	$42,279$	$42,735$	$42,735$	$-1,580$
$Ae_{1,2}$	Clásica	$43,330$	$42,227$	$42,709$	$42,7086$	$-1,661$
$Ae_{1,3}$	Clásica	$43,337$	$42,454$	$42,813$	$42,812$	$-1,609$
$Ae_{1,4}$	Clásica	$43,335$	$43,304$	$42,746$	$42,741$	$-1,621$
$Ae_{1,5}$	Clásica	$43,335$	$42,320$	$42,745$	$42,747$	$-1,612$
$Ae_{2,1}$	Clásica	$32,309$	$18,213$	$39,356$	$39,357$	$-1,665$

$Ae_{2,2}$	Clásica	$32,134$	$18,130$	$39,291$	$39,284$	$-1,678$
$Ae_{2,3}$	Clásica	$32,829$	$18,999$	$39,720$	$39,707$	$-1,682$
$Ae_{2,4}$	Clásica	$32,541$	$18,371$	$39,413$	$39,413$	$-1,682$
$Ae_{2,5}$	Clásica	$32,426$	$18,408$	$39,429$	$39,449$	$-1,674$
$Ae_{3,1}$	Clásica	$16,936$	$16,862$	$16,938$	$16,938$	$-1,293$
$Ae_{3,2}$	Clásica	$16,936$	$16,859$	$16,938$	$16,938$	$-1,273$
$Ae_{3,3}$	Clásica	$16,936$	$16,873$	$16,938$	$16,938$	$-1,285$
$Ae_{3,4}$	Clásica	$16,936$	$16,864$	$16,938$	$16,938$	$-1,289$
$Ae_{3,5}$	Clásica	$16,936$	$16,865$	$16,938$	$16,938$	$-1,2353$
$Ae_{4,1}$	Pop	$37,069$	$23,924$	$30,317$	$30,336$	$-1,200$
$Ae_{4,2}$	Pop	$36,933$	$23,810$	$30,243$	$30,226$	$-1,187$
$Ae_{4,3}$	Pop	$37,401$	$24,587$	$30,970$	$30,915$	$-1,186$
$Ae_{4,4}$	Pop	$37,170$	$24,045$	$30,400$	$30,404$	$-1,189$
$Ae_{4,5}$	Pop	$37,118$	$24,026$	$30,437$	$30,444$	$-1,163$
$Ae_{5,1}$	Pop	$34,713$	$20,896$	$36,381$	$36,406$	$-1,410$
$Ae_{5,1}$	Pop	$34,507$	$20,686$	$36,328$	$36,343$	$-1,403$
$Ae_{5,1}$	Pop	$35,329$	$21,955$	$36,872$	$36,877$	$-1,404$
$Ae_{5,1}$	Pop	$34,888$	$21,060$	$36,475$	$36,476$	$-1,404$
$Ae_{5,1}$	Pop	$34,813$	$21,070$	$36,525$	$36,511$	$-1,409$
$Ae_{6,1}$	Pop	$35,561$	$21,941$	$36,921$	$35,927$	$-1,500$
$Ae_{6,2}$	Pop	$35,420$	$21,781$	$36,841$	$35,870$	$-1,512$
$Ae_{6,3}$	Pop	$35,927$	$22,595$	$36,447$	$36,434$	$-1,512$
$Ae_{6,4}$	Pop	$35,712$	$22,022$	$36,037$	$36,034$	$-1,510$
$Ae_{6,5}$	Pop	$35,635$	$22,080$	$36,070$	$36,050$	$-1,518$
$Ae_{7,1}$	Pop	$35,775$	$23,324$	$32,660$	$32,409$	$-1,647$
$Ae_{7,2}$	Pop	$35,627$	$23,158$	$32,563$	$32,303$	$-1,638$
$Ae_{7,3}$	Pop	$36,124$	$23,946$	$32,097$	$32,994$	$-1,638$
$Ae_{7,4}$	Pop	$35,932$	$23,435$	$32,724$	$32,512$	$-1,631$
$Ae_{7,5}$	Pop	$35,834$	$23,439$	$32,722$	$32,462$	$-1,603$
$Ap_{8,1}$	Pop	$35,839$	$22,247$	$32,407$	$32,392$	$-1,466$
$Ap_{8,2}$	Pop	$35,697$	$22,168$	$32,284$	$32,288$	$-1,470$
$Ap_{8,3}$	Pop	$36,190$	$22,900$	$32,085$	$32,048$	$-1,469$
$Ap_{8,4}$	Pop	$35,958$	$22,358$	$32,517$	$32,518$	$-1,466$
$Ap_{8,5}$	Pop	$35,881$	$22,373$	$32,526$	$32,558$	$-1,464$
$Ae_{9,1}$	Pop	$36,238$	$22,694$	$34,628$	$34,633$	$-1,328$
$Ae_{9,2}$	Pop	$36,059$	$22,587$	$34,544$	$34,570$	$-1,376$
$Ae_{9,3}$	Pop	$36,509$	$22,245$	$35,173$	$34,194$	$-1,360$
$Ae_{9,4}$	Pop	$36,306$	$22,850$	$34,741$	$34,745$	$-1,361$
$Ae_{9,5}$	Pop	$36,254$	$22,804$	$34,745$	$34,742$	$-1,370$
$Ae_{10,1}$	Ranchera	$34,766$	$21,016$	$40,575$	$40,573$	$-1,345$
$Ae_{10,2}$	Ranchera	$34,606$	$20,903$	$40,543$	$40,552$	$-1,347$

$Ae_{10,3}$	Ranchera	$35,171$	$21,673$	$40,859$	$40,857$	$-1,355$
$Ae_{10,4}$	Ranchera	$35,030$	$21,172$	$40,621$	$40,632$	$-1,351$
$Ae_{10,5}$	Ranchera	$34,935$	$21,214$	$40,640$	$40,6550$	$-1,358$
$Ae_{11,1}$	Ranchera	$38,906$	$26,534$	$41,516$	$41,528$	$-0,992$
$Ae_{11,2}$	Ranchera	$38,787$	$26,525$	$41,507$	$41,508$	$-0,989$
$Ae_{11,3}$	Ranchera	$39,227$	$27,297$	$41,692$	$41,689$	$-0,985$
$Ae_{11,4}$	Ranchera	$39,107$	$27,828$	$41,559$	$41,559$	$-0,985$
$Ae_{11,5}$	Ranchera	$38,028$	$26,794$	$41,569$	$41,566$	$-0,991$
$Ae_{12,1}$	Ranchera	$34,408$	$20,536$	$43,076$	$43,076$	$-1,347$
$Ae_{12,2}$	Ranchera	$34,186$	$20,375$	$43,076$	$43,076$	$-1,339$
$Ae_{12,3}$	Ranchera	$34,795$	$20,206$	$43,076$	$43,076$	$-1,343$
$Ae_{12,4}$	Ranchera	$34,524$	$20,644$	$43,076$	$43,076$	$-1,342$
$Ae_{12,5}$	Ranchera	$34,465$	$20,663$	$43,076$	$43,076$	$-1,338$
$Ae_{13,1}$	Ranchera	$35,815$	$22,203$	$40,278$	$40,271$	$-1,302$
$Ae_{13,2}$	Ranchera	$35,654$	$22,066$	$40,217$	$40,224$	$-1,309$
$Ae_{13,3}$	Ranchera	$36,230$	$22,897$	$40,558$	$40,558$	$-1,306$
$Ae_{13,4}$	Ranchera	$35,007$	$22,396$	$40,335$	$40,339$	$-1,310$
$Ae_{13,5}$	Ranchera	$35,988$	$22,389$	$40,340$	$40,349$	$-1,308$
$Ae_{14,1}$	Rock	$41,070$	$30,837$	$40,129$	$40,126$	$-1,430$
$Ae_{14,2}$	Rock	$41,035$	$30,807$	$40,055$	$40,058$	$-1,465$
$Ae_{14,3}$	Rock	$41,263$	$31,659$	$40,392$	$40,403$	$-1,458$
$Ae_{14,4}$	Rock	$41,121$	$30,949$	$40,169$	$40,170$	$-1,4615$
$Ae_{14,5}$	Rock	$41,101$	$30,986$	$40,171$	$40,172$	$-1,430$
$Ae_{15,1}$	Rock	$33,638$	$19,680$	$32,929$	$32,907$	$-1,378$
$Ae_{15,2}$	Rock	$33,497$	$19,583$	$32,796$	$32,804$	$-1,373$
$Ae_{15,3}$	Rock	$34,053$	$19,337$	$32,432$	$33,441$	$-1,371$
$Ae_{15,4}$	Rock	$33,823$	$19,864$	$32,985$	$33,014$	$-1,369$
$Ae_{15,5}$	Rock	$33,768$	$19,844$	$32,026$	$32,016$	$-1,386$
$Ae_{16,1}$	Rock	$37,791$	$24,950$	$41,065$	$41,049$	$-1,599$
$Ae_{16,2}$	Rock	$37,649$	$24,773$	$40,973$	$40,967$	$-1,607$
$Ae_{16,3}$	Rock	$38,164$	$25,748$	$41,330$	$41,333$	$-1,608$
$Ae_{16,4}$	Rock	$37,953$	$25,139$	$41,108$	$41,103$	$-1,608$
$Ae_{16,5}$	Rock	$37,893$	$25,076$	$41,091$	$41,084$	$-1,613$
$Ae_{17,1}$	Rock	$30,133$	$22,583$	$28,853$	$28,868$	$-1,304$
$Ae_{17,2}$	Rock	$30,090$	$22,458$	$28,826$	$28,827$	$-1,310$
$Ae_{17,3}$	Rock	$30,203$	$22,098$	$29,069$	$29,095$	$-1,314$
$Ae_{17,4}$	Rock	$30,154$	$22,736$	$28,911$	$28,904$	$-1,311$
$Ae_{17,5}$	Rock	$30,152$	$22,710$	$28,912$	$28,911$	$-1,307$
$Ae_{18,1}$	Rock	$32,613$	$18,559$	$31,044$	$31,011$	$-1,259$
$Ae_{18,2}$	Rock	$32,380$	$18,356$	$30,930$	$31,011$	$-1,271$
$Ae_{18,3}$	Rock	$32,101$	$19,369$	$31,544$	$31,011$	$-1,271$

$\mathsf{Ae}_{18,4}$	Rock	$32,803$	$18,656$	$31,125$	$31,011$	$-1,267$
$\mathsf{Ae}_{18,5}$	Rock	$32,718$	$18,698$	$31,120$	$31,011$	$-1,270$
$\mathsf{Ae}_{19,1}$	Tropical	$32,341$	$20,613$	$29,393$	$29,402$	$-1,343$
$\mathsf{Ae}_{19,2}$	Tropical	$32,220$	$20,521$	$29,337$	$29,343$	$-1,329$
$\mathsf{Ae}_{19,3}$	Tropical	$32,583$	$20,286$	$29,879$	$29,867$	$-1,334$
$\mathsf{Ae}_{19,4}$	Tropical	$32,443$	$20,754$	$29,494$	$29,467$	$-1,330$
$\mathsf{Ae}_{19,5}$	Tropical	$32,433$	$20,785$	$29,484$	$29,513$	$-1,330$
$\mathsf{Ae}_{20,1}$	Tropical	$36,500$	$23,097$	$34,924$	$34,914$	$-1,014$
$\mathsf{Ae}_{20,2}$	Tropical	$36,294$	$23,919$	$34,888$	$34,893$	$-1,019$
$\mathsf{Ae}_{20,3}$	Tropical	$36,849$	$23,737$	$34,514$	$35,535$	$-1,010$
$\mathsf{Ae}_{20,4}$	Tropical	$36,610$	$23,186$	$34,061$	$34,070$	$-1,013$
$\mathsf{Ae}_{20,5}$	Tropical	$36,565$	$23,211$	$34,077$	$34,105$	$-1,013$
$\mathsf{Ae}_{21,1}$	Tropical	$31,374$	$17,204$	$32,501$	$32,476$	$-1,183$
$\mathsf{Ae}_{21,2}$	Tropical	$31,173$	$17,043$	$32,438$	$32,398$	$-1,168$
$\mathsf{Ae}_{21,3}$	Tropical	$31,764$	$17,832$	$32,212$	$32,219$	$-1,168$
$\mathsf{Ae}_{21,4}$	Tropical	$31,515$	$17,343$	$32,583$	$32,592$	$-1,169$
$\mathsf{Ae}_{21,5}$	Tropical	$31,424$	$17,323$	$32,613$	$32,626$	$-1,131$
$\mathsf{Ae}_{22,1}$	Tropical	$38,201$	$25,455$	$33,052$	$33,055$	$-1,298$
$\mathsf{Ae}_{22,2}$	Tropical	$38,052$	$25,273$	$32,996$	$33,017$	$-1,284$
$\mathsf{Ae}_{22,3}$	Tropical	$38,511$	$26,097$	$33,591$	$33,585$	$-1,282$
$\mathsf{Ae}_{22,4}$	Tropical	$38,309$	$25,550$	$33,147$	$33,173$	$-1,280$
$\mathsf{Ae}_{22,5}$	Tropical	$38,253$	$25,557$	$33,185$	$33,155$	$-1,294$
$\mathsf{Ae}_{23,1}$	Urbana	$35,679$	$22,048$	$34,126$	$34,159$	$-0,485$
$\mathsf{Ae}_{23,2}$	Urbana	$35,507$	$21,901$	$34,093$	$34,072$	$-0,470$
$\mathsf{Ae}_{23,3}$	Urbana	$35,985$	$22,603$	$34,784$	$34,789$	$-0,472$
$\mathsf{Ae}_{23,4}$	Urbana	$35,796$	$22,091$	$34,263$	$34,283$	$-0,472$
$\mathsf{Ae}_{23,5}$	Urbana	$35,745$	$22,128$	$34,290$	$34,296$	$-0,483$
$\mathsf{Ae}_{24,1}$	Urbana	$32,386$	$18,324$	$26,579$	$26,591$	$-0,938$
$\mathsf{Ae}_{24,2}$	Urbana	$32,210$	$18,171$	$26,535$	$26,510$	$-0,918$
$\mathsf{Ae}_{24,3}$	Urbana	$32,753$	$18,910$	$27,247$	$27,209$	$-0,920$
$\mathsf{Ae}_{24,4}$	Urbana	$32,501$	$18,390$	$26,759$	$26,698$	$-0,916$
$\mathsf{Ae}_{24,5}$	Urbana	$32,422$	$18,403$	$26,761$	$26,745$	$-0,939$
$\mathsf{Ae}_{25,1}$	Urbana	$35,215$	$21,514$	$30,853$	$30,760$	$-0,954$
$\mathsf{Ae}_{25,2}$	Urbana	$34,980$	$21,290$	$30,797$	$30,650$	$-0,954$
$\mathsf{Ae}_{25,3}$	Urbana	$35,646$	$22,197$	$31,470$	$31,399$	$-0,951$
$\mathsf{Ae}_{25,4}$	Urbana	$35,376$	$21,619$	$31,017$	$30,885$	$-0,954$
$\mathsf{Ae}_{25,5}$	Urbana	$35,276$	$21,690$	$31,046$	$30,911$	$-0,970$
$\mathsf{Ae}_{26,1}$	Urbana	$35,106$	$21,600$	$33,992$	$33,979$	$-1,462$
$\mathsf{Ae}_{26,2}$	Urbana	$34,980$	$21,448$	$33,949$	$33,939$	$-1,462$
$\mathsf{Ae}_{26,3}$	Urbana	$35,449$	$22,226$	$34,482$	$34,521$	$-1,472$
$\mathsf{Ae}_{26,4}$	Urbana	$35,254$	$21,690$	$34,156$	$34,091$	$-1,471$

$Ae_{26,5}$	Urbana	$35,163$	$21,757$	$34,100$	$34,123$	$-1,494$
$Ae_{27,1}$	Vallenato	$36,066$	$22,529$	$32,630$	$32,633$	$-1,377$
$Ae_{27,2}$	Vallenato	$35,897$	$22,407$	$32,562$	$32,583$	$-1,382$
$Ae_{27,3}$	Vallenato	$36,372$	$23,108$	$33,200$	$33,239$	$-1,384$
$Ae_{27,4}$	Vallenato	$36,176$	$22,644$	$32,745$	$32,735$	$-1,385$
$Ae_{27,5}$	Vallenato	$36,114$	$22,645$	$32,738$	$32,775$	$0-1,371$
$Ae_{28,1}$	Vallenato	$33,346$	$19,362$	$32,454$	$32,439$	$-1,404$
$Ae_{28,2}$	Vallenato	$33,171$	$19,246$	$32,367$	$32,354$	$-1,396$
$Ae_{28,3}$	Vallenato	$33,772$	$20,025$	$33,116$	$33,062$	$-1,396$
$Ae_{28,4}$	Vallenato	$33,526$	$19,523$	$32,548$	$32,586$	$-1,395$
$Ae_{28,5}$	Vallenato	$33,449$	$19,500$	$32,583$	$32,614$	$-1,401$
$Ae_{29,1}$	Vallenato	$36,142$	$22,665$	$34,174$	$34,152$	$-1,480$
$Ae_{29,2}$	Vallenato	$35,998$	$22,527$	$34,123$	$34,066$	$-1,469$
$Ae_{29,3}$	Vallenato	$36,521$	$23,308$	$34,691$	$34,751$	$-1,468$
$Ae_{29,4}$	Vallenato	$36,352$	$22,833$	$34,345$	$34,308$	$-1,469$
$Ae_{29,5}$	Vallenato	$36,351$	$22,886$	$34,288$	$34,305$	$-1,465$
$Ae_{30,1}$	Vallenato	$31,121$	$16,966$	$32,244$	$32,221$	$-0,949$
$Ae_{30,2}$	Vallenato	$30,971$	$16,895$	$32,149$	$32,141$	$-0,948$
$Ae_{30,3}$	Vallenato	$31,547$	$17,615$	$32,761$	$32,761$	$-0,938$
$Ae_{30,4}$	Vallenato	$31,311$	$17,106$	$32,325$	$32,329$	$-0,939$
$Ae_{30,5}$	Vallenato	$31,240$	$17,139$	$32,302$	$32,375$	$-0,939$

En la Tabla C.2 se presentan los géneros de los audios secretos utilizados para el desarrollo de las pruebas de validación.

Tabla C.2: Géneros de los audios secretos.

Audio	**Género músical**
As1	Rock
As2	Pop
As3	Voz
As4	Pop
As5	Rock

A continuación se presentan los espectrogramas de los audios portada pertenecientes al género de música Clásica (ver Tabla C.3), además de los audios portada que constituyen los audios *stego* usados en la prueba preliminar de percepción (ver Tabla C.4).

Tabla C.3: Espectrogramas de audios portada asociados a música clásica.

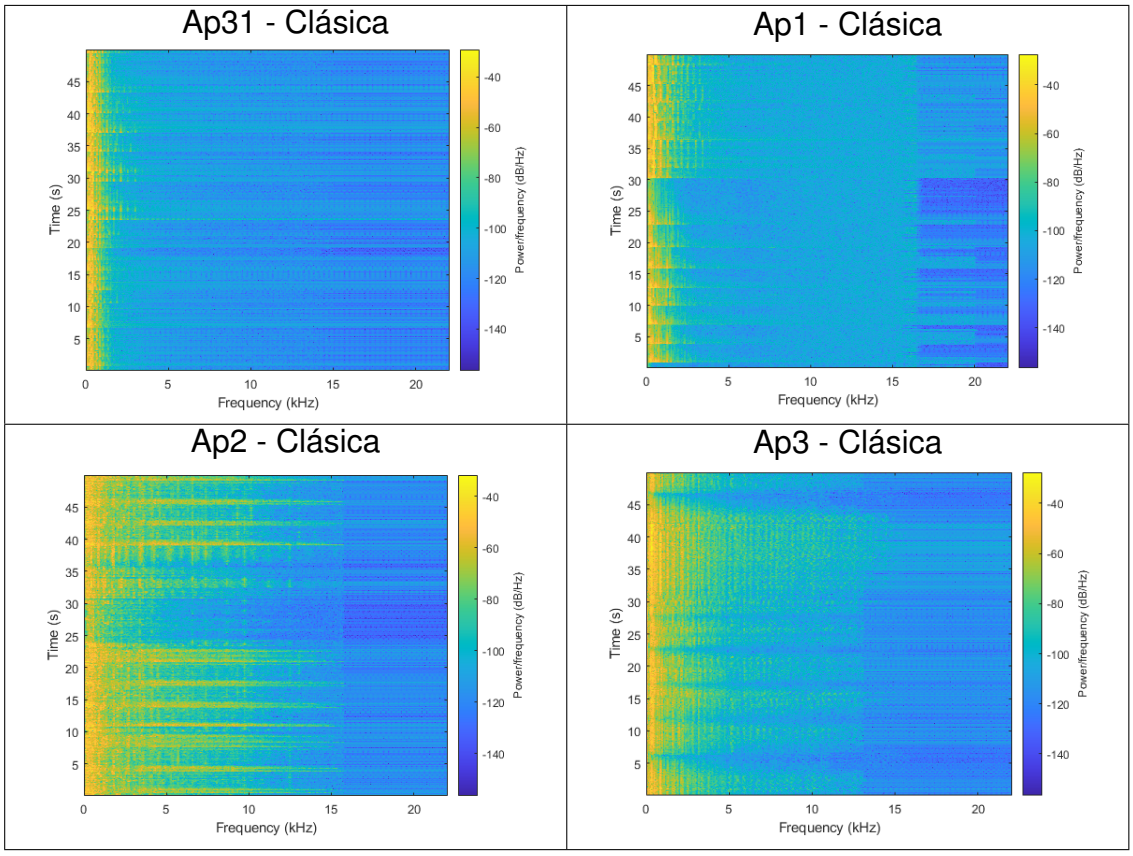
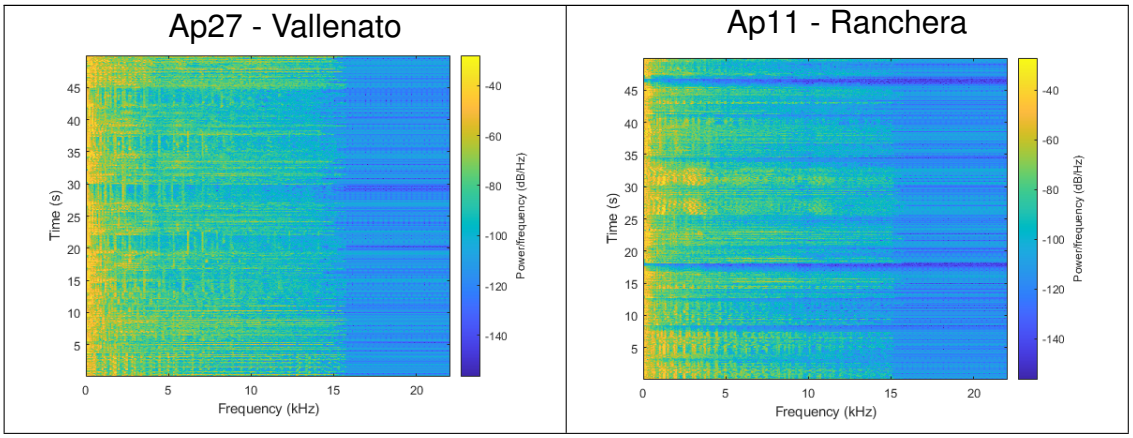
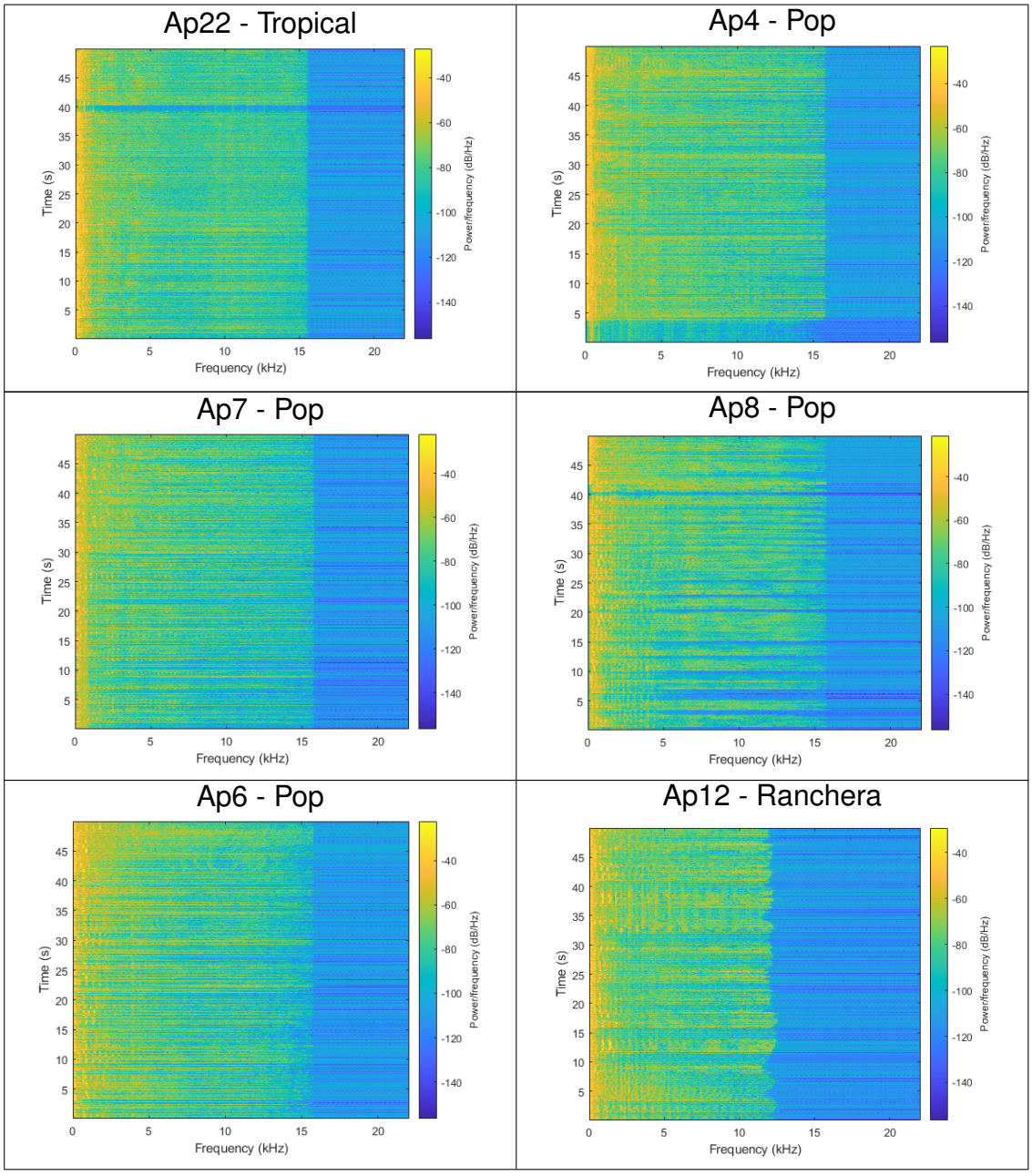
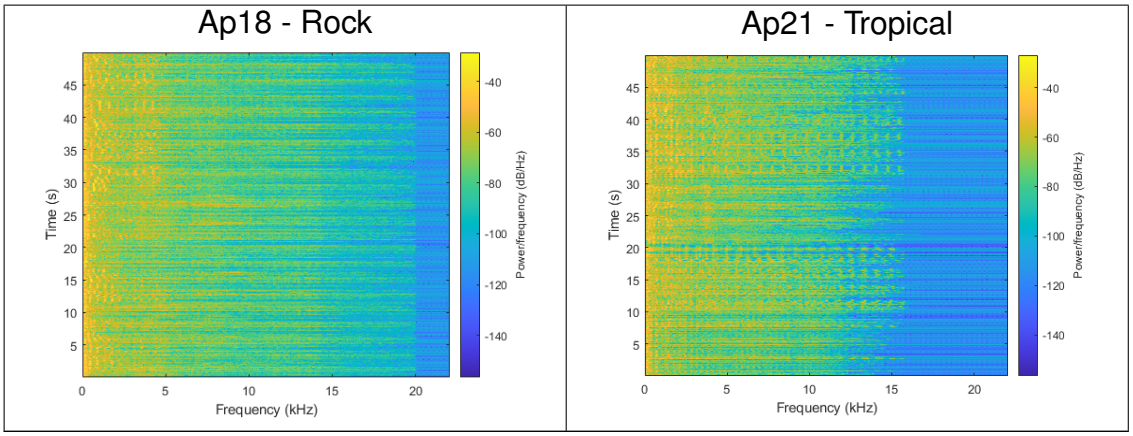


Tabla C.4: Espectrogramas de audios portada usados a la prueba preliminar de percepción.




Ap22 - Tropical
Ap4 - Pop
Ap7 - Pop
Ap8 - Pop
Ap6 - Pop
Ap12 - Ranchera
Time (s)
Frequency (kHz)
Power/frequency (dB/Hz)



Con el fin de sintetizar la información obtenida con la PSD de los audios listados y resaltados de color negro[11] en la Tabla C.1, y dado que se ha identificado la importancia de los valores que toma la PSD en la banda de 0 a $3KHz$, se generan las siguientes gráficas en las cuales el eje *x* representa los posibles valores en dB que puede tomar la PSD y el eje *y* indica el porcentaje de la PSD, de cada grupo de intensidades, que se encuentra acumulada en la banda de 0 a $3KHz$. Cada gráfica muestra la comparación de los 4 audios que se consideran por género musical.

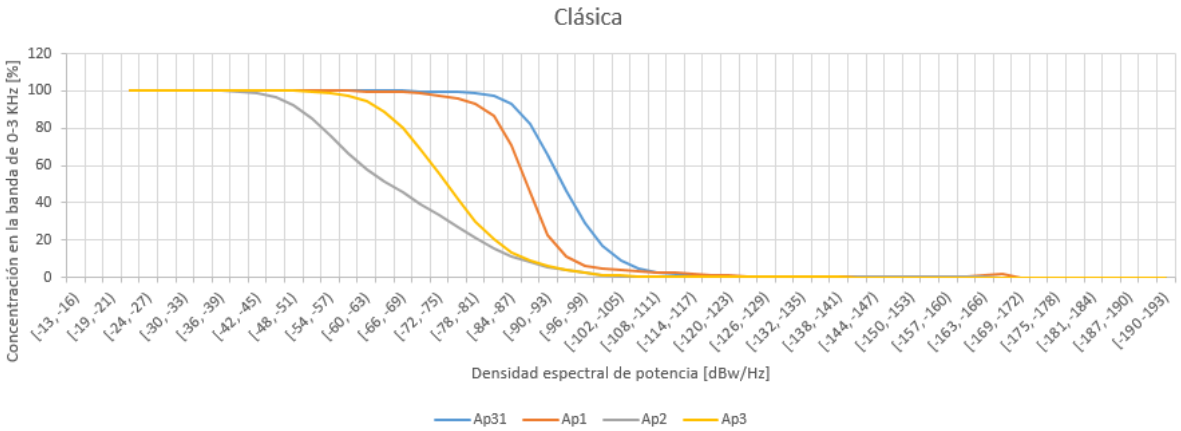


Figura C.1: *Concentración de PSD en banda de* $0 - 3KHz$ *para audios clasificados como música clásica.*

[11] En la Tabla C.1, los audios resaltados en color azul hacen parte del banco de audios del cual partieron las pruebas de validación. Cuando las pruebas de validación pasan a su fase de percepción, se ve la necesidad de caracterizar los audios portada, lo que conlleva a modificar el banco de audios con el fin de que cada género músical contenga una misma cantidad de audios; de ésta transición, los audios resaltados se ven inevitablemente excluidos.

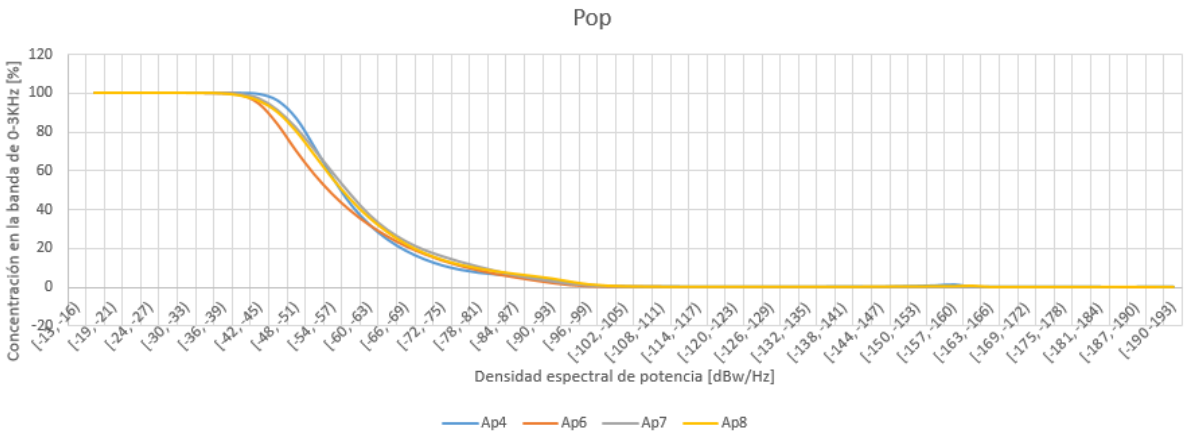


Figura C.2: *Concentración de PSD en banda de* $0 - 3KHz$ *para audios clasificados como música pop.*

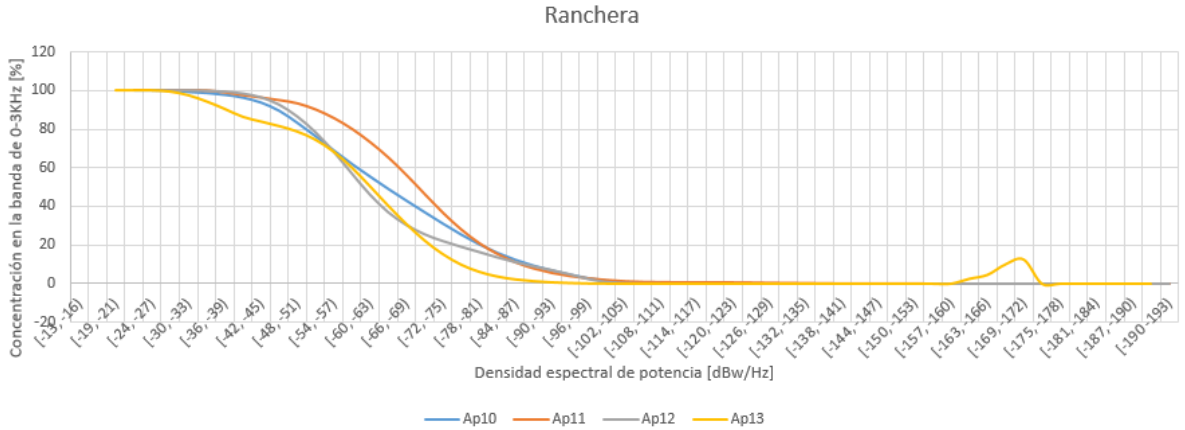


Figura C.3: *Concentración de PSD en banda de* $0 - 3KHz$ *para audios clasificados como música ranchera.*

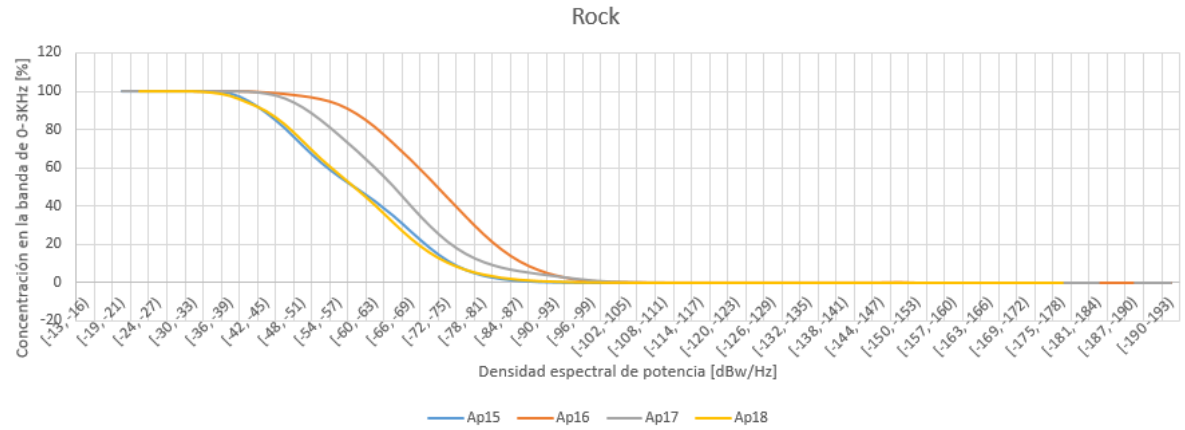


Figura C.4: *Concentración de PSD en banda de* $0 - 3KHz$ *para audios clasificados como música rock.*

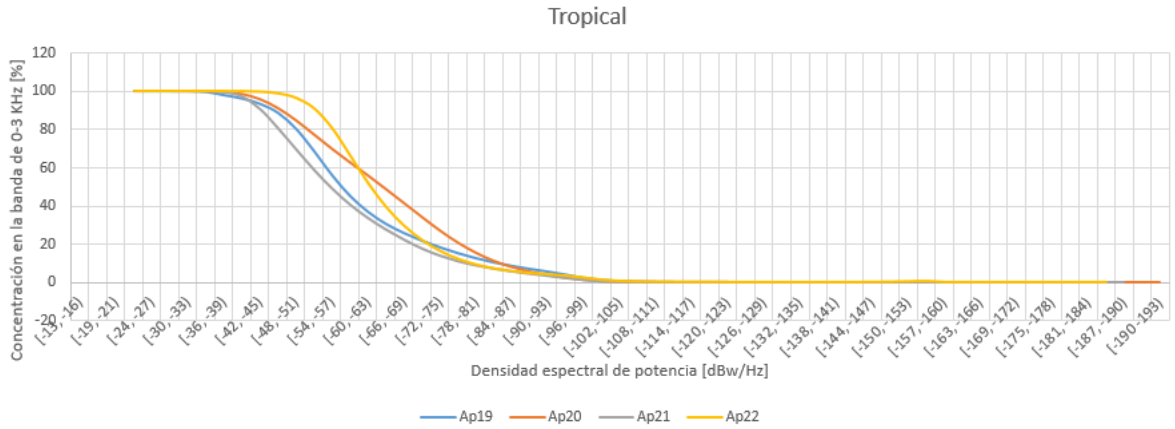


Figura C.5: *Concentración de PSD en banda de* $0 - 3KHz$ *para audios clasificados como música tropical.*

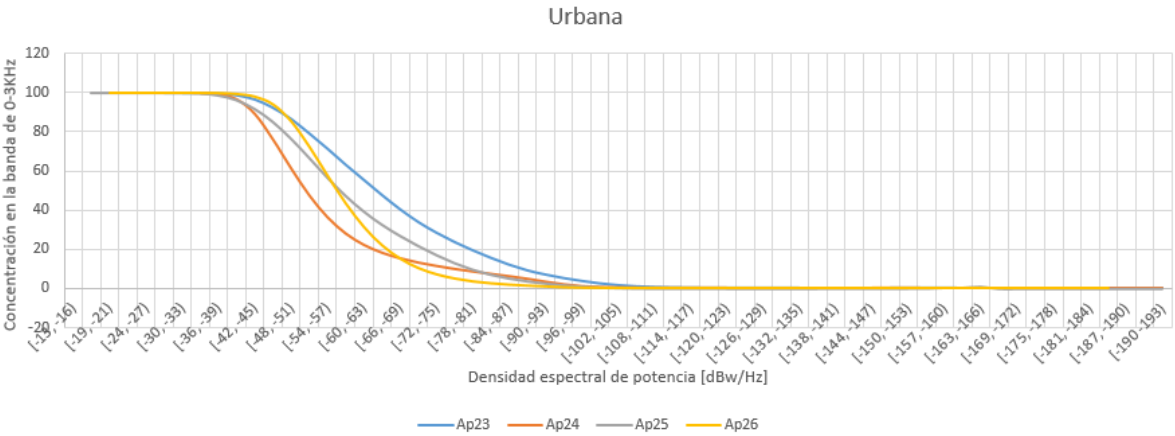


Figura C.6: *Concentración de PSD en banda de* $0 - 3KHz$ *para audios clasificados como música urbana.*

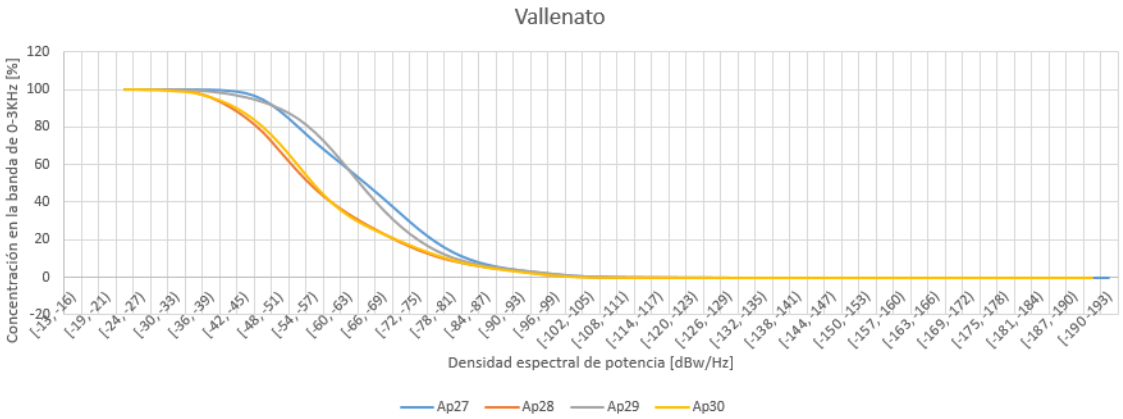


Figura C.7: *Concentración de PSD en banda de* $0 - 3KHz$ *para audios clasificados como música vallenato.*

Las Figuras C.1 a C.7 demuestran cómo la PSD brinda una guía hacia lo beneficioso que puede ser un audio como archivo portada -una mayor acumulación de PSD en banda base es lo deseado- aunque no se cumple en todos los casos ya sea por el género al cuál pertenece el audio, la poca definición de la concentración de PSD en banda base u otras características que pueden ser estudiadas con mayor rigor, pero que sobrepasan los requisitos de este trabajo.

Es posible resaltar algunos casos que permiten verificar su utilidad:

- Para los audios categorizados como música clásica, según la Figura C.1, *Ap31*, *Ap1*, *Ap3* y *Ap2* es el orden adecuado para los audios en términos de lo aptos que pueden ser como audios portada; no obstante, *Ap3* cuenta con SNR menor que la obtenida por *Ap2* (ver Tabla C.1).
- En la Figura C.4, el mejor audio a seleccionar como portada es *Ap16* y la segunda mejor opción corresponde a *Ap17*; mientras que, según la PSD los audios *Ap15* y *Ap18* no se pueden discernir, por lo que se debe recurrir a otros métodos de caracterización para clasificarlos adecuadamente, aunque en este punto es posible afirmar que los audios *stego* resultantes no superan en imperceptibilidad a los audios generados con *Ap16* y *Ap17*.
- La Figura C.7 no define si el audio *Ap27* podría ser un mejor audio portada que el audio *Ap29*, pero sí permite deducir que ambos son mejores que los audios *Ap28* y *Ap30*; si se recurre a los resultados expuestos en la Tabla C.1, los valores de SNR de los audios *stego* generados a partir de *Ap27* y *Ap29* son más altos que los obtenidos mediante *Ap28* y *Ap30*, específicamente en los *métodos uno y dos*, aunque los resultados de los *métodos tres, cuatro y cinco* son realmente próximos entre sí, validando, en cierta medida, lo anteriormente mencionado.

Finalmente se anexan los resultados del análisis estadístico correspondiente a la desviación estándar de cada una de las señales de audio presentadas en la Tabla, C.5, y que fueron utilizadas para la caracterización de los audios portada en el Capítulo 4.

Tabla C.5: Desviación estándar de audios *stego* resultantes de las pruebas de validación.

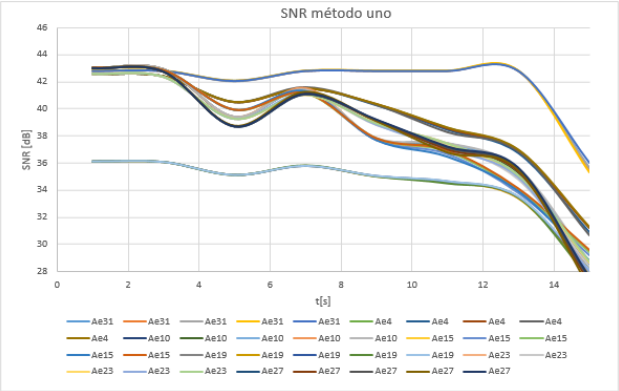
Audio	**Desviación estándar del espectro de magnitud**	**Desviación estándar de la señal temporal**
Ap31	$50,20758753$	$0,033951866$
Ap1	$25,93242793$	$0,016624388$
Ap2	$70,1861597$	$0,051257296$
Ap3	$22,82927748$	$0,015587795$
Ap4	$377,1034478$	$0,263801082$
Ap6	$396,153991$	$0,281371629$
Ap7	$315,2629212$	$0,220758856$
Ap8	$325,4534424$	$0,231731885$
Ap10	$193,4273108$	$0,137130563$
Ap11	$86,42098381$	$0,060742979$
Ap12	$147,0639351$	$0,105507432$
Ap13	$318,5241698$	$0,225570463$
Ap15	$201,8097715$	$0,141945753$
Ap16	$86,50362947$	$0,053511909$
Ap17	$92,61525798$	$0,065296831$
Ap18	$147,3312816$	$0,106692915$
Ap19	$222,5599982$	$0,159997895$
Ap20	$129,1646698$	$0,091008591$
Ap21	$141,976681$	$0,106063362$
Ap22	$254,8244538$	$0,17808672$
Ap23	$197,5811957$	$0,140368496$
Ap24	$343,4308735$	$0,247046196$
Ap25	$312,3319339$	$0,219464352$
Ap26	$338,9496961$	$0,239026237$
Ap27	$193,461887$	$0,137797832$
Ap28	$326,3446582$	$0,236185274$
Ap29	$203,9224736$	$0,145837918$
Ap30	$220,6479093$	$0,16561354$

APÉNDICE D

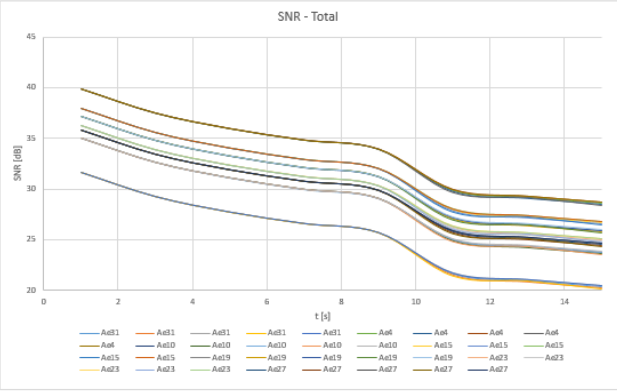
VARIACIÓN DE LOS AUDIOS SECRETOS

Con el objetivo de respaldar los resultados expuestos en la sección 4.5, en donde se aborda la caracterización de los audios portada como causa de los resultados obtenidos en las pruebas de validación (ver Sección 4.4), se plantea la variación de cinco audios secretos a incrustar manteniendo la cantidad de información constante en siete audios portada de distintos géneros musicales (*Ap31*, *Ap4*, *Ap10*, *Ap15*, *Ap19*, *Ap23*, *Ap27*) en el *método uno y método uno temporal* respectivamente, permitiendo así evaluar la influencia del audio secreto en la imperceptibilidad del audio *stego*.

En la Figura D.1 se tienen los valores de SNR obtenidos para cada audio *stego* generado a partir de cada uno de los siete audios portada y cinco audios secretos, asumiendo un valor de $n_{LSB} = 1$; a partir de esta gráfica es posible evidenciar que los valores de SNR se mantienen constantes cuando se conserva el mismo tipo de audio portada, independientemente del audio secreto incrustado, presentándose así como líneas superpuestas indicando que no hay mayores variaciones en los resultados obtenidos. Aunque en ambos dominios este comportamiento se mantiene, cabe mencionar que la decadencia en los resultados de SNR en el *método uno temporal* son más pronunciados debido al desempeño de este método en comparación con el *método uno*.



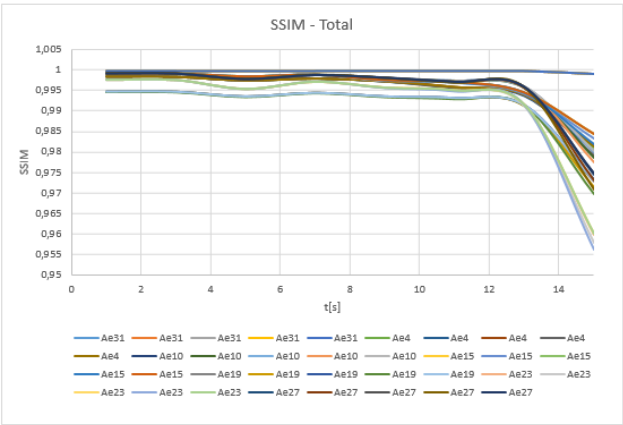
(a) *SNR método uno*



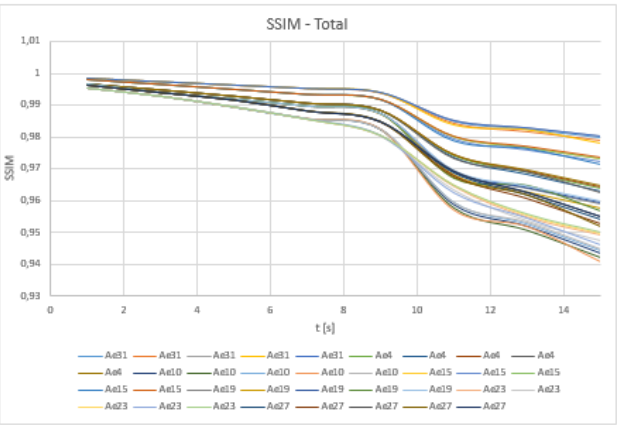
(b) *SNR método uno temporal*

Figura D.1: *Variación de audio secreto. Métrica de evaluación SNR de audios stego.*

En cuanto a los resultados correspondientes a las métricas de evaluación SSIM y PEAQ-ODG, (ver Figuras D.2 y D.3), se obtuvo un comportamiento similar al del SNR, ya que los resultados en ambos casos se mantienen superpuestos mientras el audio portada se mantenga fijo, lo cual permite validar la premisa de que la variación del audio secreto no representa cambios importantes en la imperceptibilidad del audio *stego*; asimismo es notorio el bajo desempeño del *método uno temporal* en comparación al *método uno*, ya que los valores de PEAQ-ODG en este caso fueron mucho menores alcanzando así la escala de *muy molesto* (-4).

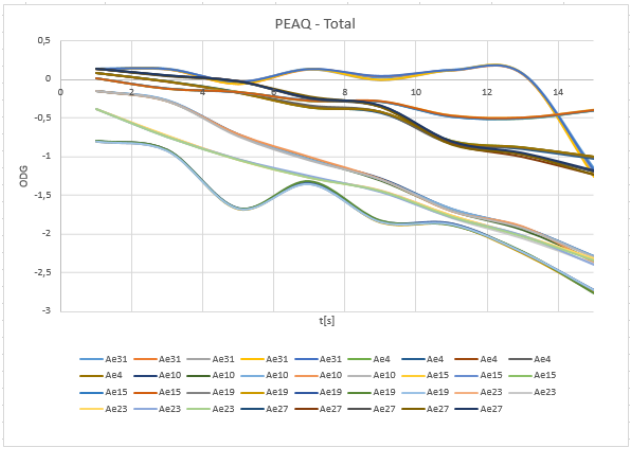


(a) *SSIM método uno*

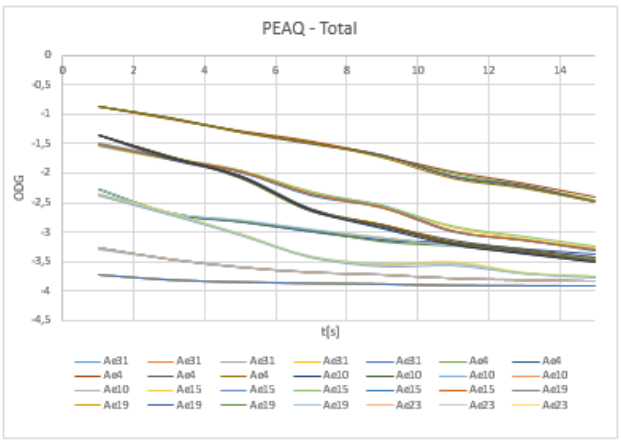


(b) *SSIM método uno temporal*

Figura D.2: *Variación de audio secreto. Métrica de evaluación SSIM de audios stego.*



(a) *PEAQ-ODG método uno*



(b) *PEAQ-ODG método uno temporal*

Figura D.3: *Variación de audio secreto. Métrica de evaluación PEAQ-ODG de audios stego.*

APÉNDICE E

CLAVE

Para que el proceso esteganográfico sea exitoso es necesario que el emisor pueda generar el archivo *stego* y que sólo el receptor deseado pueda recuperar el mensaje secreto incrustado en el archivo *stego*. Para que el receptor pueda recuperar la información secreta sin generar pérdidas o distorsión del mensaje es necesario que reconozca los parámetros usados por el transmisor a la hora de generar el archivo *stego*; para ello, algunos parámetros invariables pueden ser preestablecidos tanto en el transmisor como en el receptor; sin embargo, hay otros parámetros que dependen de las dimensiones del mensaje secreto y el método de incrustación seleccionado (parámetros variables).

Esta información variable se suele utilizar en la esteganografía para incluir un proceso de cifrado y de esta manera aumentar la seguridad en la transmisión de información, de esta manera, la información variable se convierte en una clave o llave que permite descifrar el proceso que se debe llevar a cabo en el receptor para recuperar el mensaje secreto.

Para la esteganografía de audio se reconocen tres opciones de enviar la información necesaria para el proceso de extracción del audio secreto, es decir, la clave:

1. **Parámetros presentes en el audio *stego*:** los parámetros usados en el transmisor equivalen a información, la cual puede expresarse de forma binaria. Esta información se incrusta en el audio portada en una ubicación pre-acordada entre el transmisor y receptor. Para extraer la información secreta, el receptor recupera los parámetros variables directamente del audio *stego*, es decir, la información necesaria para ejecutar la técnica esteganográfica se encuentra centralizada (parámetros y archivo *stego*).

   Los parámetros variables pueden ser pocos o muchos, dependiendo del método usado y la dimensión del mensaje secreto, lo cual afecta la cantidad de carga útil (información secreta) que realmente se puede incrustar en el audio portada, lo cual degrada la capacidad de incrustación de la técnica esteganográfica; no obstante, tiene como ventaja que evita que se pueda interceptar el mensaje independiente que se tendría que usar en otro caso para transmitir la clave y que es, potencialmente, un riesgo en la seguridad.

2. **Parámetros almacenados en un archivo de clave:** El transmisor adiciona los parámetros necesarios para la recuperación del mensaje secreto en un archivo de clave; esto implica que el transmisor brinda dos salidas: el audio *stego* y el archivo de clave [39]. El archivo de clave se envía al receptor antes del audio *stego*, lo cual permite que éste último cuente con los parámetros adecuados para obtener el mensaje secreto. El usar un archivo clave, garantiza que la capacidad de incrustación del audio portada sea usada totalmente en carga útil; además, el archivo de clave puede someterse a un cifrado para garantizar la seguridad de la información que transporta.

3. **Parámetros almacenados en un audio *stego* adicional:** Para evitar alterar la capacidad de incrustación del audio portada en el cual se adiciona la información secreta, es posible crear un audio *stego* adicional que es el encargado de llevar los parámetros al receptor, como por ejemplo [40], en donde implementan y envían dos audios *stego* previos para sincronizar los parámestros entre el transmisor y receptor antes de enviar el audio *stego* que contiene el mensaje secreto. Es relevante mencionar que, tanto el transmisor como el receptor, deben saber de antemano la cantidad de parámetros incrustados y su localización dentro de los dos primeros audios *stego*.

Considerando que la incrustación de los parámetros variables en audios *stego* no es viable dado que varían en cantidad según el método implementado y que la incrustación directa de estos parámetros en el audio *stego*, además de disminuir el espacio de carga útil, facilita la extracción de la información secreta por parte de un observador no deseado, que en el peor de los escenarios puede descubrir que el archivo corresponde a un audio *stego*, hacen que en el contexto de este trabajo de grado se considere que el archivo clave es la opción más adecuada, ya que es posible incluir el número de parámetros necesarios, se puede envíar en un instante aleatorio (diferente a la hora de envío del audio *stego*), por un medio diferente (una red y/o plataforma distinta), y sin afectar la capacidad de incrustación (carga útil) del audio portada elegido.

Los archivos clave se implementan en los *métodos uno, tres* y *cuatro* a partir del Capítulo 4; dadas las diferencias entre estos métodos, la información almacenada en dichos archivos varía, tal y como se expone a continuación.

- **Archivo de clave para el *método uno*:** contiene el número de LSB alterados (n_{LSB}), el número de muestras de audio secreto incrustadas (*Ls*) y el valor máximo y mínimo del audio secreto para garantizar una adecuada decodificación del audio secreto al establecer correctamente las regiones de cuantificación.

- **Archivo de clave para el *método tres*:** este archivo contiene los límites superiores e inferiores de las bandas usadas para la incrustación (*Bn*), el número de LSB alterados en las bandas (n_{LSB}), el número de bits incrustados (*Ts*) y el valor máximo y mínimo del audio secreto.

- **Archivo de clave para el *método cuatro*:** almacena los límites superiores e inferiores de cada una de las bandas para la incrustación (*Bn*), el número de LSB modificados en las bandas (n_{LSB}), el número de bits incrustados (*Ts*) y los valores máximo y mínimo del audio secreto.

Dado que realizar el cifrado del archivo clave excede el alcance de este trabajo de grado, se recurre a herramientas disponibles en línea como *CiberProtector* [41] para este proceso, de esta forma se realizaron pruebas para validar el funcionamiento de la clave cifrada, la cual, emplea *AES* 256 bits y *bcript* como métodos de cifrado, de está forma si un atacante pudiera obtener la clave secreta, sería inentendible y por tanto inaccesible para él. Ésta junto con el audio *stego* en formato *WAV*, permiten en todos los casos que el receptor deseado pueda recuperar el audio secreto.