

Banco de Síntesis de Formas de Onda para Audio

(Julio 2010)

Nazly Renata Yaya Pinilla-yayisrenatis@hotmail.com

Resumen—Este artículo proporciona la información más importante acerca de la herramienta implementada, en esta se pueden visualizar las características de la señal de audio en tiempo y su análisis en frecuencia y luego mediante la síntesis granular se podrán obtener formas de onda nuevas para alimentar el banco de síntesis.

Palabras clave—Transformada corta de Fourier, Transformada discreta de Fourier, Ventaneos, Síntesis granular.

I. INTRODUCTION

LA mayoría de software y herramientas de procesamiento de señales de audio están dirigidas hacia un mercado de audio profesional, lo que quiere decir que no se pueden adquirir en todos los sectores de la industria, aunque existen herramientas como los computadores que poseen una tarjeta de sonido interna encargada del análisis y síntesis del sonido reconstruyendo los tonos, los usuarios no pueden observar el procesamiento que realiza la tarjeta.

La herramienta que se desarrollará, permite al usuario observar las características de la señal de audio en tiempo y su análisis en frecuencia, además se podrán generar formas de onda diferentes a la original mediante la síntesis. Al realizar una herramienta como esta se estaría implementando el sistema que se encuentra en las tarjetas de sonido pero de una forma más asequible para el público, donde no se reconstruye la señal que se tiene guardada, sino que se genera una nueva a partir del archivo de audio.

El banco de síntesis de formas de onda para audio, está dirigido a estudiantes, docentes, ingenieros o conocedores de la música que quieren sintetizar tonos o adquirir tonos nuevos para ser utilizados en otras herramientas que se encargan de integrar los para formar melodías.

II. MARCO REFERENCIAL

A. Análisis del sonido

Los sonidos son señales de tiempo variable en el mundo real y de hecho todo su significado se relaciona con la variabilidad en el tiempo. Por lo tanto, es interesante desarrollar técnicas de análisis de sonido que permitan observar algunas de las características principales del sonido en tiempo, para facilitar así tareas como la comprensión, comparación, modificación, y síntesis. Por eso, es importante que los criterios para elegir los parámetros del análisis, tales como la longitud y tipo de ventana sean observados detalladamente.[1]

B. Transformada de Fourier a corto plazo

La transformada de Fourier a corto plazo (STFT) es un método utilizado para el análisis, síntesis y modificación de la señal, esta se ha utilizado durante muchos años y sus usos son numerosos. [2]

La transformada de Fourier a corto plazo se puede ver como una forma para representar una señal en el dominio del tiempo y la frecuencia. La idea básica para aplicar la STFT consiste en realizar la transformada de Fourier en una porción limitada de la señal, después cambiando de puesto a otra porción de la señal se realiza el mismo procedimiento, entonces la señal es reconstruida por los valores de la transformada de Fourier obtenida en las diversas partes de la señal. [2]

Cuando los valores de la transformada de Fourier se expresan en coordenadas polares, La transformada de Fourier a corto plazo es llamada “Codificador vocal de la fase”. [2]

1) Análisis mediante la transformada corta de Fourier

En usos estándar, el análisis de la transformada de Fourier a corto plazo se realiza a una tarifa constante: los instantes de tiempo del análisis t_a^u regularmente se espacian, es decir, $t_a^u = uR$ donde R es un incremento fijo del número entero que controla la tarifa del análisis. Sin embargo, en las escalas de tono y en las modificaciones de escala de tiempo, es más fácil

utilizar en la síntesis instantes de tiempo regularmente espaciados y no instantes de tiempo uniformes. En la denominada banda de paso, la transformada de Fourier a corto plazo $X(t_a^u, \Omega_k)$ es definido por:

$$X(t_a^u, \Omega_k) = \sum_{n=-\infty}^{\infty} h(n)x(t_a^u + n) \exp(-j\Omega_k n)$$

Donde $h(n)$ es la ventana del análisis y $\Omega_k = \frac{2\pi k}{N}$. Se debe

tener en cuenta que la transformada de Fourier es en forma discreta siendo el tiempo y la frecuencia discretos. La transformada de Fourier es calculada en N puntos, N debe ser generalmente más larga que la longitud de la ventana del análisis $h(n)$. [1]

La transformada discreta de Fourier es una función periódica de una variable continua, puede ser invertido por medio de una integral computada durante un período.

$$w(m-n)y(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} Y_m(w)e^{jwn} dw$$

2) Visualización en una ventana [3]

Para analizar las frecuencias presentes en la señal es mejor hacerlo por tramos, ya que de esta forma es más sencillo encontrar la información de frecuencia en un tiempo determinado, para realizar este proceso se hace un ventaneo, es decir se aplica sobre una parte la señal la función ventaneo. Existen diferentes tipos de ventanas cada una posee parámetros diferentes y permiten mejor rendimiento en determinadas frecuencias.

La magnitud del espectro de cada ventana se caracteriza por tener un gran lóbulo principal centrado en $w = 0$ seguido por una serie de lóbulos laterales con amplitudes disminuidas. Dos parámetros que predicen el rendimiento de la ventana son el ancho del lóbulo principal y el nivel relativo del lóbulo lateral que tienen que ver con el diseño del filtro.

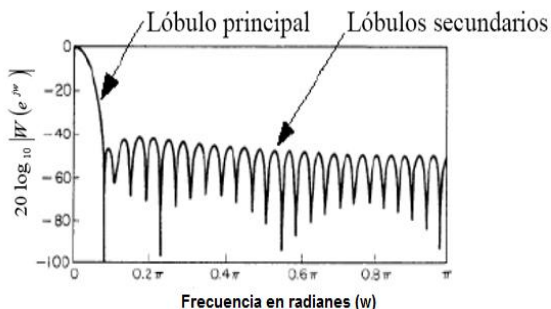


Figura 1. Características del espectro de la ventana

El ancho del lóbulo principal Δ_{ML} es la distancia más cercana entre los cruces por cero en ambos lados del lóbulo principal y el nivel relativo de los lóbulos laterales A_{sl} es la diferencia en dB entre las amplitudes del lóbulo lateral más grande y el lóbulo principal.

Para asegurar una transición rápida entre la banda útil y la banda de parada, la ventana debe tener un ancho pequeño del lóbulo principal, y para reducir la banda útil y las ondulaciones en la banda de parada, el área debajo de los lóbulos laterales debe ser menor.

El ancho del lóbulo principal (Δw) de la ventana determina la resolución en frecuencia. Una buena atenuación de los lóbulos laterales frente al principal evita la distorsión en la forma y envolvente del espectro.

Para que exista una compensación entre la anchura del lóbulo principal y el nivel del lóbulo lateral que pueden ser manipulados eligiendo o diseñando una ventana apropiada. Las ventanas que aplican diferentes funciones para lograr el suavizado deseado son Hamming, Blackman, Hanning, Kaiser, Barlett entre otras.

C. Síntesis Granular [4]

En un fenómeno acústico el sonido está compuesto por miles de partículas o granos que se propagan en el tiempo, el grano dura entre uno y cien milisegundos. La representación granular se utiliza para apreciar fenómenos sonoros con un alto nivel de complejidad, ya que los granos son unidades de energía unidas por tiempo y frecuencia

La noción de granulación fue propuesta por Dennis Gabor, en la teoría se propuso que mediante la representación granular se podía describir un sonido.

En 1960 el compositor Iannis Xenakis fue el primero en explicar la teoría acerca de la composición con granos de sonido y realizó la siguiente observación: "Todos los sonidos se conciben como un ensamble de un gran número de sonidos elementales que están dispuestos en el tiempo.", luego en 1974, Curtis Roads inicio la implementación de síntesis granular mediante un algoritmo para su aplicación.

La técnica de generación de sonido llamada síntesis granular, consiste en dividir un sonido en pequeños segmentos, para luego mediante las características de la señal envolvente extraer los granos generando así nuevos sonidos.

La duración de un grano puede ser constante, aleatoria o puede variar dependiendo de la frecuencia, esto significa que se pueden asignar duraciones cortas a los granos. Una envolvente de amplitud da la forma a cada grano de sonido, esta envolvente

puede variar en diferentes implementaciones de una curva de tipo campana, una señal Gaussian, o una ventana como la Hamming y Hanning que son las más utilizadas en el caso de un archivo de audio.

III. DESARROLLO

A. Esquema general de la herramienta

La herramienta está compuesta por dos partes: una de análisis de la señal de audio y la otra la síntesis de la señal.

En la primera parte se realiza el análisis de la señal mediante la transformada corta de Fourier. En esta se toman tramos de la señal y realizar la transformada discreta de Fourier a cada uno, los tramos son tomados mediante la multiplicación con una ventana observándose así el análisis de la señal en frecuencia.

La segunda parte es la síntesis. Allí mediante las características que se le asignen a la ventana envolvente de los cuantos (fragmentos de sonidos cortos) se generan diferentes formas de onda, luego de obtener la nueva señal en tiempo se realiza el análisis de Fourier para observar la señal sintetizada en frecuencia.

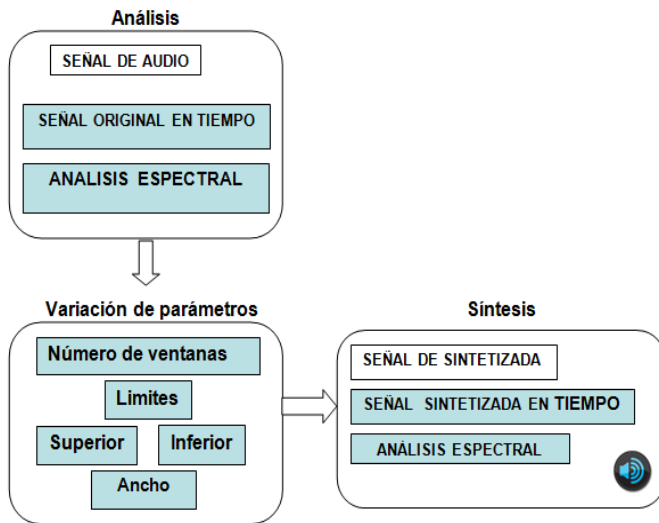


Figura 2. Esquema Herramienta

B. Análisis de la señal original

Para analizar señales de audio es necesario acudir al dominio del tiempo o al dominio de la frecuencia, en el dominio del tiempo se puede observar la variación de la amplitud con respecto al tiempo, en este se observa una representación natural de la señal siendo difícil determinar sus componentes fundamentales razón por la cual se debe obtener la señal en el dominio de la frecuencia en la cual se representa la variación de la amplitud en función de la frecuencia apreciándose los componentes de la señal.

En la figura 12, se puede observar la señal de audio con su respectivo análisis espectral en MATLAB.

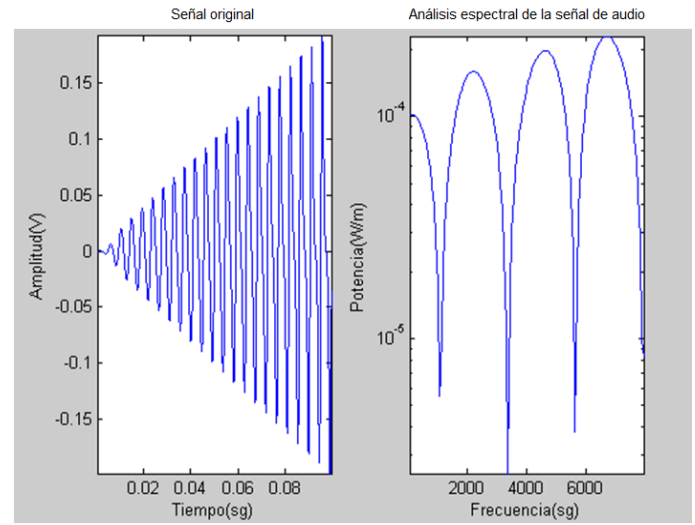


Figura 3. Señal en el dominio del tiempo y de la frecuencia

Una de las herramientas matemáticas más importantes para llevar a cabo la transformación de señales del dominio del tiempo al dominio de la frecuencia es la Transformada de Fourier.

Para el desarrollo del proyecto se realizó el análisis de la señal de audio mediante la transformada corta de Fourier, este proceso consiste en descomponer la señal en una serie de segmentos y analizarlos independientemente. Para realizar la descomposición es necesario establecer la ventana que permita tomar las muestras de la señal para luego aplicar la transformada discreta de Fourier.

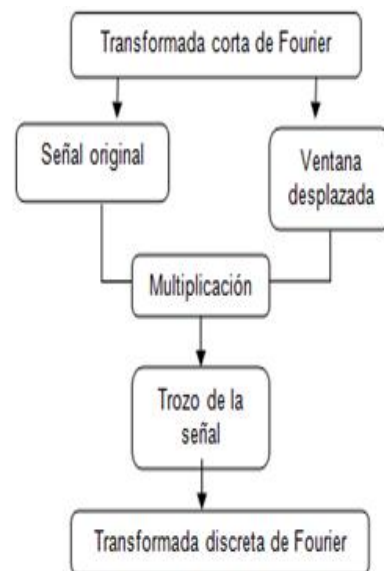


Figura 3. Diagrama bloques Transformada corta de Fourier

1) Características del ventaneo

Para poder identificar la ventana más adecuada para el análisis de la señal de audio es necesario observar las características de cada una de ellas. [5]

Los principales parámetros que caracterizan una ventana temporal son los siguientes:

- El ancho de banda a 3dB es la diferencia entre las frecuencias en la que al aplicar la ventana hay una atenuación que se mantiene igual o inferior a 3dB comparada con la frecuencia central del pico máximo. (Habilidad que tiene la ventana para separar componentes de igual amplitud).
- Selectividad: indica la habilidad que tiene la ventana para separar componentes diferentes. (el parámetro es el factor de forma que se obtiene de la relación entre el ancho de banda a 60dB y a 3dB).
- Altura del lóbulo lateral principal, determina la resolución en frecuencia.
- Atenuación de los lóbulos laterales evita la distorsión en la forma y envolvente del espectro.
- Magnitud del rizado en la banda de paso, esta define la capacidad de la ventana para representar la magnitud real.

Tabla. Características fijas de algunas funciones de las ventanas

Tipo de ventana	Ancho del lóbulo principal Δ_{Mz}	Nivel relativo del lóbulo lateral A_{L_i}	Atenuación mínima de la banda de parada	Transición ancho de banda Δ_w
Rectangular	$4\pi/(2M+1)$	13.3dB	20.9dB	$0.92\pi/M$
Hanning	$8\pi/(2M+1)$	31.5dB	43.9dB	$3.11\pi/M$
Hamming	$8\pi/(2M+1)$	42.7dB	54.5dB	$3.32\pi/M$
Blackman	$12\pi/(2M+1)$	58.1dB	75.3dB	$5.56\pi/M$

Se puede decir que la mayoría de las ventanas permiten realizar un análisis de las señales de audio pero por ejemplo en la práctica la ventana rectangular introduce algunos errores en los componentes en frecuencia, la ventana Blackman posee

buena resolución en frecuencia pero no muy buena en amplitud comúnmente es utilizada para medir distorsiones, las que tienen mayor utilidad en este campo son la Hamming y la Hanning, la Hamming posee una buena resolución en frecuencia como en amplitud y la Hanning trata de suavizar los efectos en los bordes de la señal.

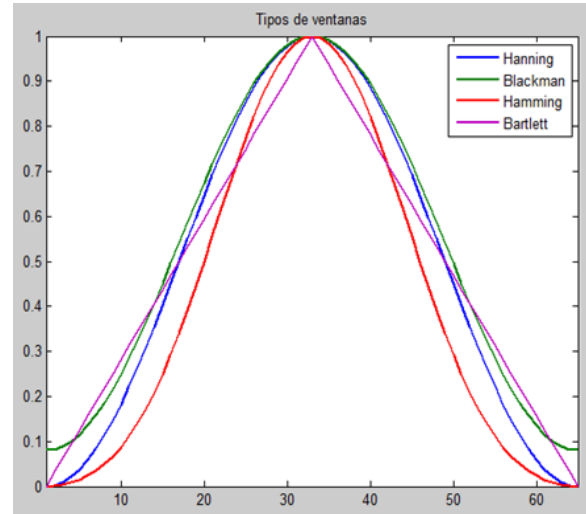


Figura 4. Tipos de ventanas

En MATLAB se encuentran preestablecidas cada una de las ventanas, haciendo más preciso el cálculo matemático de cada una de ellas igualmente se puede manipular la longitud a implementar.

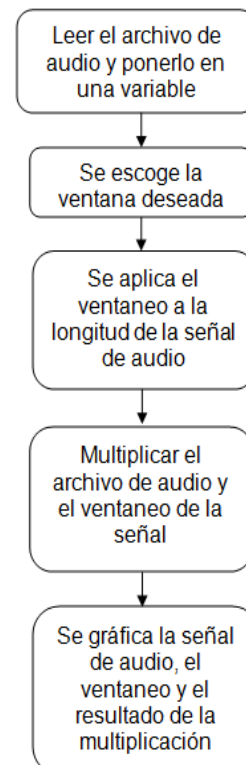


Figura 5. Diagrama implementación ventanas

2) Transformada discreta de Fourier [6]

La transformada de Fourier permite descomponer la señal en frecuencias, la función definida en el tiempo $f(t)$ es transformada al dominio de la frecuencia $F(\omega)$ siendo esta la transformada de Fourier. Fourier especificó que es posible expresar las señales como la suma de senos y cosenos con diferentes frecuencias y amplitudes para establecer la función original.

En los procesos de adquisición de datos el análisis no se realiza en forma continua sino que se llevan a cabo mediante técnicas digitales donde es necesario realizar un muestreo en intervalos de tiempo, por esto, es preciso aplicar la transformada de Fourier a la señal discreta y realizar el análisis espectral pero mediante la transformada discreta de Fourier.

Se puede notar que el número de operaciones que se deben realizar en la transformada discreta de Fourier demandan mucho tiempo y recursos del sistema, por eso, en 1965 Cooley y Tukey desarrollaron un nuevo algoritmo que reduce en gran medida el número de operaciones, la transformada rápida de Fourier se utiliza para implementar la transformada discreta de Fourier (DFT) de forma rápida y eficiente, ya que para realizar la DFT se requieren (N^2) operaciones aritméticas y en el algoritmo de la transformada rápida de Fourier (FFT) se obtiene el mismo resultado con solo $(N * \log_2 N)$ simplificando así las operaciones y haciendo más eficiente el proceso.

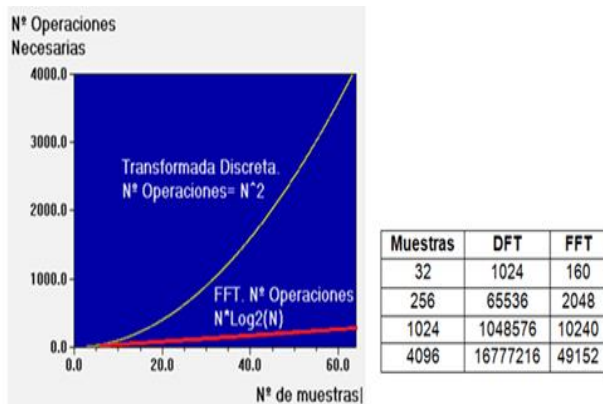


Figura 6. Comparación entre DFT y FFT

Se crea una función en MATLAB donde se van a introducir los parámetros, luego se halla la longitud del vector asignándole una variable, se crea una variable que contiene los números complejos que se van a obtener, luego se hace una división recursiva en donde se aplica la transformada rápida de Fourier.

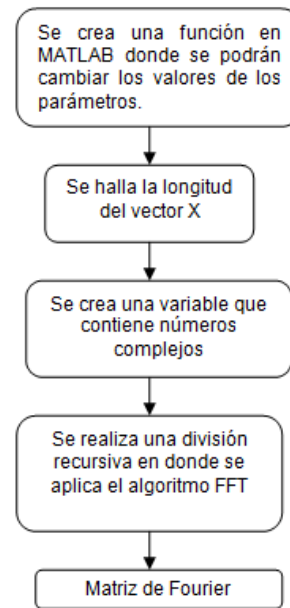


Figura 7. Implementación Transformada discreta de Fourier

C. Síntesis de la señal de audio

La síntesis de la señal de audio se hizo en base a las síntesis granular, está es una técnica utilizada en producción musical, en la cual se toman pequeños fragmentos de la señal mediante una envolvente y luego mediante los fragmentos que se tomaron se obtiene una nueva forma de onda.

En este caso la señal envolvente es una ventana Hamming, ya utilizada en el análisis de la señal, al ir variando el número de ventanas, sus límites superior e inferior y el ancho respectivamente se modifican las características de la señal en el intervalo de tiempo que va desde que se termina un grano hasta que comienza uno nuevo realizando así la síntesis de la señal de audio, adicional a esto parámetros como la frecuencia y la amplitud también varían.

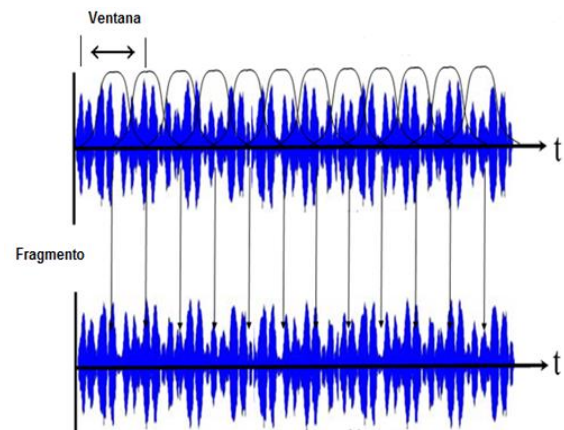


Figura 8. Señal de audio sintetizada mediante granos

IV. DESCRIPCIÓN HERRAMIENTA

La interfaz gráfica se implementó con el objetivo de poder sintetizar la señal de audio mediante las características de la ventana, ya que mediante esta envolvente y teniendo en cuenta los cuantos (granos) que se extraen de la señal original se pueden observar y escuchar los cambios que se presentaron en la señal de audio tanto en frecuencia como en amplitud, además se pueden manipular los programas que ya se habían implementado mediante la interfaz.

Lo primero que se hace es una presentación del proyecto que posee un pequeño manual práctico.

En la siguiente interfaz, se le permite al usuario abrir un archivo de audio el cual puede ser seleccionado, luego se observan las gráficas de la señal original y su análisis en frecuencia, otra de las opciones es poder escuchar el sonido.

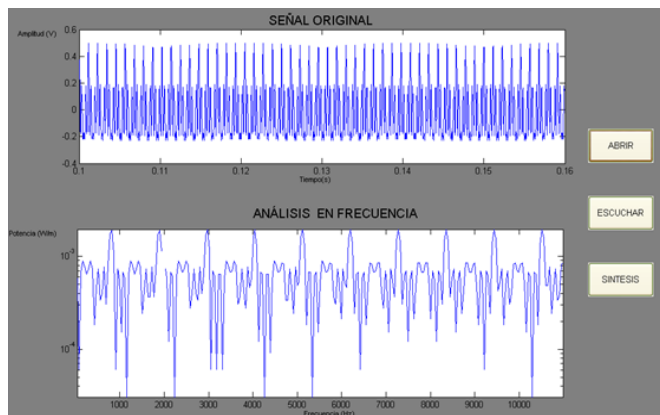


Figura 9. Interfaz gráfica de señal original y análisis en frecuencia

Los archivos de audio que permiten realizar el análisis y síntesis son los que tienen una duración de uno a cinco segundos, ya que si son de mayor duración la capacidad de procesamiento del sistema no permite observar el análisis en frecuencia de la señal y no se puede realizar la síntesis, también es muy importante que se trabajen con señales en formato wav y que la información sea codificada en formato PCM, en caso contrario se genera un error en la información de codificación del archivo y no abrirá.

Es importante tener en cuenta las siguientes características para que el sistema funcione correctamente, el límite superior siempre debe ser mayor que el límite inferior tienen un rango de 0.1 a 0.9 que permite definir la ventana en los rangos que decida el usuario y el ancho debe estar entre 0.1 a 0.8, lo más adecuado es que se escoja un valor intermedio entre los límites inferior y superior creando así la forma de la envolvente que se va mover en el tiempo y va ir extrayendo los fragmentos. El número de ventanas debe mantenerse dentro del rango de 25 a 200, en este rango las señales se sintetizan adecuadamente y permite realizar el análisis en frecuencia a la señal sintetizada.

Al escoger el botón síntesis se abre otra interfaz en donde se cambian los valores de la ventana o envolvente de los fragmentos, según cada uno de estos valores se modificarán las características de la señal original. Se pueden variar valores como número de ventanas, límite superior e inferior y el ancho.



Figura 10. Parámetros de la ventana envolvente

Al colocar continuar sale una ventana con dos gráficas que permiten observar los cambios que tuvo la señal en tiempo y en el análisis en frecuencia, igualmente se tiene la opción de escuchar el tono sintetizado o poderlo guardar para ir almacenando en el banco de tonos.

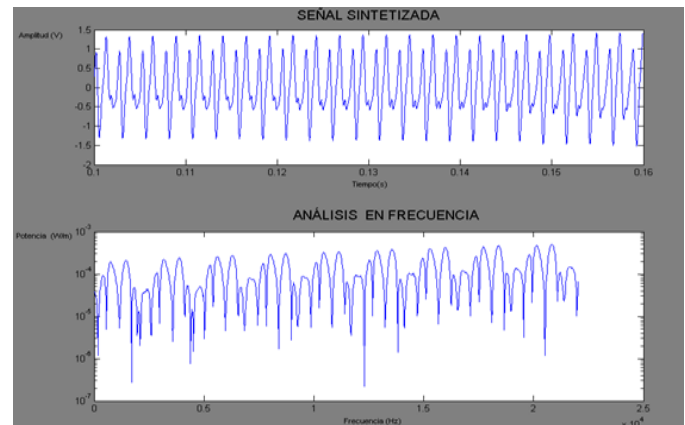


Figura 11. Señal sintetizada y análisis en frecuencia

Por último, se le permite al usuario observar la comparación de las dos señales, allí podrá escuchar los dos sonidos el original y el sintetizado y guardar la señal sintetizada.

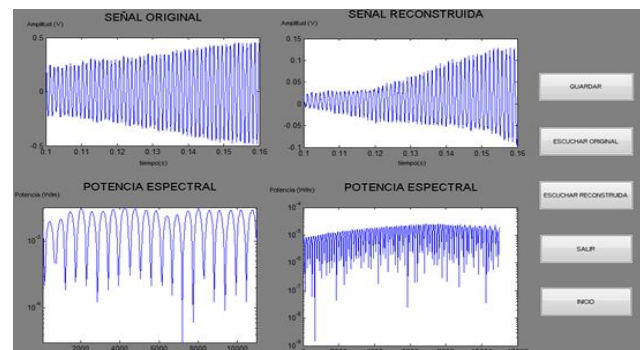


Figura 12. Comparación señal original y sintetizada

V. CONCLUSIONES

- En el estudio de los tipos de síntesis, se observó que la mayoría de ellos generaban tonos mediante la lectura rápida de señales que tenían almacenadas, la única que permitía manipular las características de los archivos de audio era la síntesis granular utilizada para realizar la reconstrucción de la señal mediante los segmentos extraídos de la señal de audio generando de esta manera nuevas formas de onda.
- La síntesis que se realiza con un número de envolventes que están entre un rango de veinte a treinta, no sufre alteraciones significativas en sus características, cambian suavemente, ya que los segmentos que se toman no hacen que la señal envolvente se solape.
- La síntesis que se genera entre los valores de cuarenta y noventa envolventes, permiten que el sonido de la nueva señal genere un tono agudo con respecto a la señal del archivo original.
- La síntesis que se realiza con un número de ventanas entre cien y doscientos permiten obtener sonidos graves, en este caso por tener un mayor número de ventanas se solapan y generan repeticiones en las muestras de sonido.
- Los archivos de audio que permiten realizar el análisis y síntesis son los que tienen una duración de uno a cinco segundos, ya que si son de mayor duración la capacidad de procesamiento del sistema no permite observar el análisis en frecuencia de la señal y no se puede realizar su síntesis
- En MATLAB se debe trabajar con señales en formato wav y su información debe ser codificada en formato PCM, en caso contrario se genera un error en la información de codificación del archivo y no abrirá.
- Se diseñó una interfaz gráfica amigable que le permite al usuario tener una herramienta en donde pueda observar la representación gráfica de una señal de audio en tiempo, además puede controlar el comportamiento de la señal sintetizada permitiéndole escuchar y guardar la nueva forma de onda.
- El sistema diseñado es práctico y de bajo costo, le permite al usuario a partir de una señal inicial, poder sintetizar cuantas veces quiera el tono, adquiriendo variedad de tonos que le posteriormente podrán ser utilizado en otras herramientas para formas melódías.

- [4] D. Prieto and J. Reyes (2001). Fundamentos de Síntesis de audio con granos, 2001 <http://www.maginvent.org/articles/granintro/granintro.html>
- [5] E. Florez, S. Cardona and L. Jordi, Selección de la ventana temporal en la transformada de Fourier en tiempos cortos utilizada en el análisis de señales de vibración para determinar planos en las ruedas de un tren Revista Facultad de ingeniería Universidad de Antioquia (2009).
- [6] J. Cortés, H. Cano and J. Chaves, Del análisis de Fourier a las wavelets-transformada continua wavelet (CWT), (2007).<http://redalyc.uaemex.mx/redalyc/pdf/849/84903723.pdf>

Nazly Renata Yaya Pinilla: Nació en Bogotá, el 26 de febrero de 1987. Su bachillerato lo realizó en el colegio Santa Teresa de Jesús, estudia la fundación universitaria San Martín y está concluyendo la etapa profesional con el proyecto de grado Banco de síntesis de formas de onda para audio.

REFERENCES

- [1] D. Rocchesso, Introduction to digital audio signal processing. Venezia, 2003. Pag 99.
- [2] M. Kahrs and K. Brandenburg, Applications of digital signal processing to audio and acoustics. Kluwer Academic Publishers New York, Boston, Dordrecht, London, Moscow. 2002
- [3] S. K. Mitra, Digital signal processing. McGraw-Hill, Second Edition. New York, 1998