

Instituto Tecnológico de Buenos Aires

22.05 ANÁLISIS DE SEÑALES Y SISTEMAS DIGITALES

Trabajo práctico N°2

Grupo 3

MECHOULAM, Alan	58438
LAMBERTUCCI, Guido Enrique	58009
RODRIGUEZ TURCO, Martín Sebastian	56629
LONDERO BONAPARTE, Tomás Guillermo	58150

Profesores

Jacoby, Daniel Andres
Belaustegui Goitia, Carlos F.
Iribarren, Rodrigo Iñaki

Presentado: 15/05/20

Índice

1. Síntesis aditiva de sonidos	2
1.1. Grupos armónicos	2
1.2. Ruido	2
1.3. Ventana temporal	3
1.4. Caída de armónicos en frecuencia	3
2. Síntesis de sonidos mediante frecuencia modulada	4
2.1. Introducción al modelo	4
2.2. Instrumentos de viento	5
2.2.1. Clarinete	5
2.2.2. Trombón	6
2.2.3. Campana	7
2.2.4. Trompeta	8
2.3. Conclusiones	9
3. Síntesis de sonidos mediante modelos físicos	10
3.1. Karplus-Strong básico	10
3.1.1. Análisis teórico	10
3.1.2. Análisis singularidades	11
3.1.3. Sintonización de frecuencia	12
3.1.4. Tipos de ruido	12
3.1.5. Estabilidad	16
3.2. Mejora propuesta	16
3.2.1. Sintonización de frecuencia	16
3.2.2. Continuidad del sonido	17
3.2.3. Caja de resonancia	18
3.3. Karplus-Strong percusión	18
3.4. Espectrogramas	18
4. Síntesis por muestras	19
4.1. Time Stretching	19
4.2. TD-PSOLA	20
4.3. Estimación de los Pitch-Marks	22
5. Efectos de Audio	23
6. Programas implementados	24
6.1. FFT	24
6.2. Programa Principal	24

1. Síntesis aditiva de sonidos

La síntesis aditiva consiste en formar el sonido de un instrumento sumando todas las señales sinusoidales por las cuales este está formado. Si bien existen muchas nomenclaturas que refieren a lo mismo, a lo largo de esta implementación se utilizarán las denominaciones *primer armónico* como la componente de menor frecuencia de un sonido; *segundo, tercero, ..., n armónico* como las componentes de frecuencia múltiplos del primer armónico del sonido; y *sobretonos* como cualquier componente de frecuencia del sonido mayor al primer armónico que no es múltiplo de este.

1.1. Grupos armónicos

Se sintetizó en una primera instancia al órgano. El órgano es un instrumento muy versátil con un amplio rango de sonidos diferentes, dado que estos están frecuentemente formados por miles de tubos metálicos, abiertos o cerrados. Estos tubos suelen estar agrupados en rangos de manera tal que estos simulen otros instrumentos de viento, como la flauta, el clarinete, el bassoon, el cor anglais, y muchos más. Se puede observar que el órgano es un sintetizador aditivo natural, dado que cada uno de los tubos posee un único tono y timbre, junto a sus armónicos. Es por esta razón que se decidió sintetizar a este instrumento.

Sin embargo, un órgano de hoy en día posee alrededor de 20000 tubos y 300 rangos distintos. Como sería muy difícil sintetizar a cada tubo por separado, se decidió sintetizar al instrumento de una manera versátil, emulando al instrumento. Se agruparon a los armónicos y sobretonos del instrumento dada una sola nota de frecuencia f_0 en los siguientes grupos:

Grupo armónico	Frecuencias
Fundamental	$\sum_{i=1}^{\infty} f_0 + 2 \cdot i \cdot f_0$
Quinta	$\sum_{i=1}^{+\infty} 3f_0 + 6 \cdot i \cdot f_0$
Primaria	$\sum_{i=1}^{+\infty} 2f_0 + 4 \cdot i \cdot f_0$
Octava	$\sum_{i=1}^{+\infty} 4f_0 + 8 \cdot i \cdot f_0$
Duodécima	$\sum_{i=1}^{+\infty} 6f_0 + 12 \cdot i \cdot f_0$
Decimoquinta	$\sum_{i=1}^{+\infty} 8f_0 + 15 \cdot i \cdot f_0$
Decimoséptima	$\sum_{i=1}^{+\infty} 9f_0 + 17 \cdot i \cdot f_0$
Decimonovena	$\sum_{i=1}^{+\infty} 20f_0 + 19 \cdot i \cdot f_0$
Tercera mayor	$\sum_{i=1}^{+\infty} f_0 \cdot 2^{\frac{4}{12}} + 2 \cdot i \cdot f_0$
Cuarta perfecta	$\sum_{i=1}^{+\infty} f_0 \cdot 2^{\frac{5}{12}} + 2 \cdot i \cdot f_0$
Quinta perfecta	$\sum_{i=1}^{+\infty} f_0 \cdot 2^{\frac{7}{12}} + 2 \cdot i \cdot f_0$

Asimismo, se permite configurar la cantidad de armónicos que se añaden del grupo fundamental, y la cantidad de armónicos que se añaden del resto. A partir de estos parámetros, se sintetizaron dos configuraciones distintas, las de flauta, y las de órgano entero, mostrado a continuación:

Grupo armónico	Proporción	Grupo armónico	Proporción
Fundamental	0.3	Fundamental	0.15
Quinta	0	Quinta	0.11
Primaria	0.65	Primaria	0.17
Octava	0	Octava	0.13
Duodécima	0	Duodécima	0.11
Decimoquinta	0	Decimoquinta	0.095
Decimoséptima	0	Decimoséptima	0.095
Decimonovena	0	Decimonovena	0.13
Tercera mayor	0.03	Tercera mayor	0.005
Cuarta perfecta	0	Cuarta perfecta	0.003
Quinta perfecta	0.02	Quinta perfecta	0.002

1.2. Ruido

De vital importancia en la sintetización de instrumentos a la hora de lograr un mayor realismo, y más aun así en instrumentos de viento es el ruido. Se utilizó para esto ruido binario, el cual es superpuesto a cada nota generada por el sintetizador, logrando así un mayor realismo.

1.3. Ventana temporal

Se basó la ventana temporal utilizada en esta implementación en la clásica ventana ADSR. Sin embargo, al sintetizarse un instrumento de viento, se fijó el parámetro de delay como nulo y el de sustain como unitario. Luego, en vez de realizar la ventana a trozos linealmente, se utilizaron subidas y bajadas exponenciales con una oscilación en el periodo de sustain, el cual se utiliza para simular la técnica de *vibrato* en los instrumentos de vientos generada por tanto vibraciones de la fuente de aire como en la embocadura del instrumento, como se observa en la Figura (1). La frecuencia o incluso existencia de estas vibraciones dependen de la longitud de la nota T . Luego, los parámetros que se le pide al usuario son los de A para *Attack*, R para *Release* y H para *Oscillation*, quedando finalmente definida la ventana mediante la fórmula a trozos:

$$\begin{cases} -e^{-t \cdot (\frac{10F}{A \cdot T})} + 1 & 0 \leq t \leq \frac{A \cdot T}{F} \\ 1 + \left(\frac{2}{1 + e^{-T \cdot (t - \frac{A \cdot T}{F})}} - 1 \right) \cdot H \cdot \sin \left(2\pi \cdot \left(t - \frac{A \cdot T}{F} \right) \frac{4T}{F \cdot (1 - R - A)} \right) & \frac{A \cdot T}{F} \leq t \leq T \cdot (1 - \frac{R}{F}) \\ -e^{t \cdot (\frac{10F}{R \cdot T})} + 1 & T \cdot (1 - \frac{R}{F}) \leq t \leq T \end{cases} \quad (1)$$

donde

$$F = \frac{2}{1 + e^{-\frac{f}{10 \cdot f_0}}} \cdot (V + 0.5) \quad (2)$$

Siendo V la velocity de la nota normalizada de 0 a 1, f_0 la frecuencia fundamental de la nota y f la frecuencia del armónico al cual se le aplica la ventana. Se puede observar que la ventana crece y decrece más rápido proporcional a la razón entre la frecuencia del armónico y la frecuencia fundamental de la nota, mientras que la ventana crece y decrece más rápido inversamente proporcional a la velocity de la nota. Esto quiere decir que los sonidos más graves reaccionarán más lento, mientras que los agudos se percibirán más rápido. Además, si una nota se toca con una velocity baja, habrá un transitorio más largo entre la máxima y mínima amplitud de la nota.

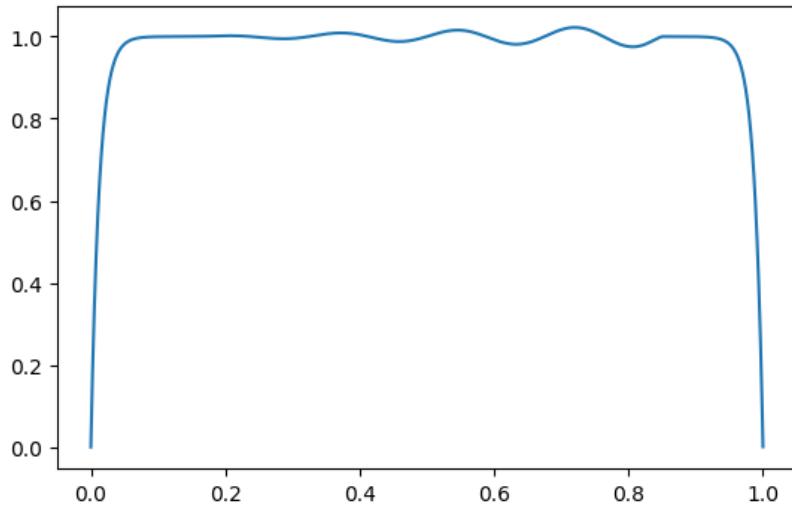


Figura 1: Ventana temporal para $T = 1$, $A = 0.1$, $R = 0.1$, $H = 0.08$ y $F = 1$.

Por último, se utilizó una función sigmoidea modificada, la cual mapea valores dentro del intervalo $[0, +\infty)$ al intervalo $[0, 1]$, con el fin de que las oscilaciones en la fase de sustain aparezcan de manera incremental, como suele suceder naturalmente con el vibrato.

1.4. Caída de armónicos en frecuencia

Al añadir un grupo de armónicos al instrumento, se puede pensar como que se agrega una nueva frecuencia fundamental junto a sus armónicos, sin embargo, estos armónicos no tienen la misma amplitud que su fundamental,

sino que decrecen linealmente en una escala logarítmica. La pendiente de esta envolvente lineal fue hallada observando el espectro de varias notas de instrumentos reales.

Además, esta pendiente decae con la frecuencia, lo que fue experimentalmente hallado como una mejora.

2. Síntesis de sonidos mediante frecuencia modulada

2.1. Introducción al modelo

El siguiente método de síntesis se basa en modelar sonidos mediante señales moduladas en frecuencia. Es por eso que, dada la siguiente ecuación

$$x(t) = A(t) \cos(2\pi f_c + I(t) \cos(2\pi f_m t + \phi_m) + \phi_c) = A(t) \sin(2\pi f_c + I(t) \sin(2\pi f_m t)) \quad (3)$$

con $\phi_m = \phi_c = -\frac{\pi}{2}$, se busca elegir $A(t)$, $I(t)$, f_c y f_m adecuados para poder simular adecuadamente el sonido del instrumento deseado.

De la Ecuación (3) se observa que $I(t)$ es el índice de modulación, mientras que f_c y f_m son la frecuencia de la portadora y la moduladora, respectivamente. Cuando $I(t)$ es positivo, se observan frecuencias que se encuentran por encima y por debajo de la portadora en intervalos dados por la moduladora. La cantidad de frecuencias laterales que se observa está relacionada con dicho índice, es por ello que a mayor $I(t)$, mayor cantidad de picos en frecuencia. A su vez, la frecuencia central decrece con el aumento previamente mencionado ¹.

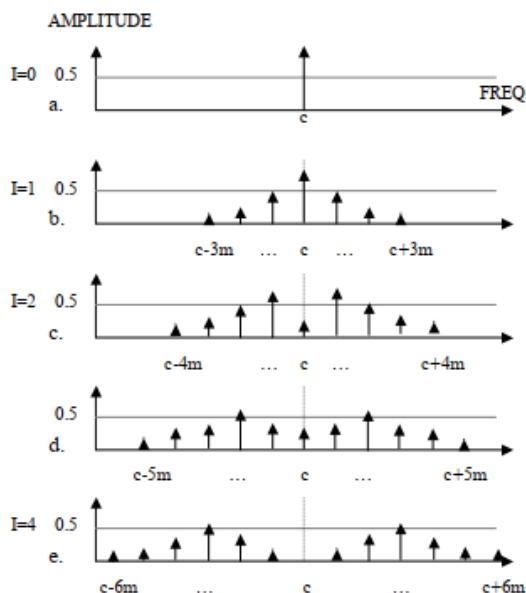


Figura 2: Aumento de $I(t)$ y como afecta las frecuencias laterales.

Las amplitudes de la portadora y las laterales son determinadas por las funciones de Bessel de primer y “n-esimo” orden $J_1(I(t))$ y $J_n(I(t))$, expresando $x(t)$ de la forma:

$$x(t) = A \sum_0^n J_k(I(t)) [\sin(\omega_c + k\omega_m)t - \sin(\omega_c - k\omega_m)t] \quad (4)$$

La principal ventaja de la síntesis mediante F.M. consiste justamente en el hecho de que se puede expresar con facilidad la relación entre la portadora y la moduladora, lo que produce las frecuencias laterales mencionadas previamente. Estas recaen en el lado negativo de frecuencias, reflejándose en el dominio positivo del espectro.

¹John M. Chowning, [The Synthesis of Complex Audio Spectra by Means of Frequency Modulation](#). 1973.

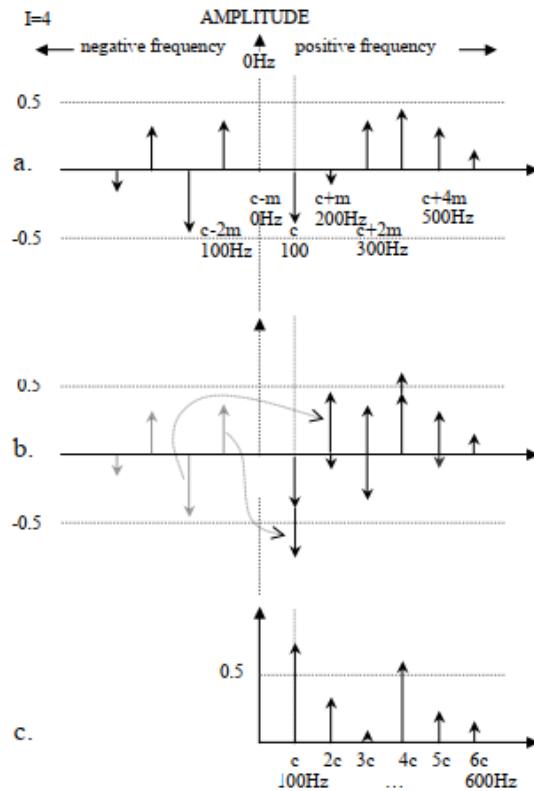


Figura 3: Representación gráfica de la proyección de frecuencias negativas.

Por otro lado, se puede expresar la relación de la frecuencia portadora y moduladora de la forma

$$\frac{f_c}{f_m} = \frac{N_1}{N_2} \quad (5)$$

lo que permite obtener una expresión para la frecuencia central

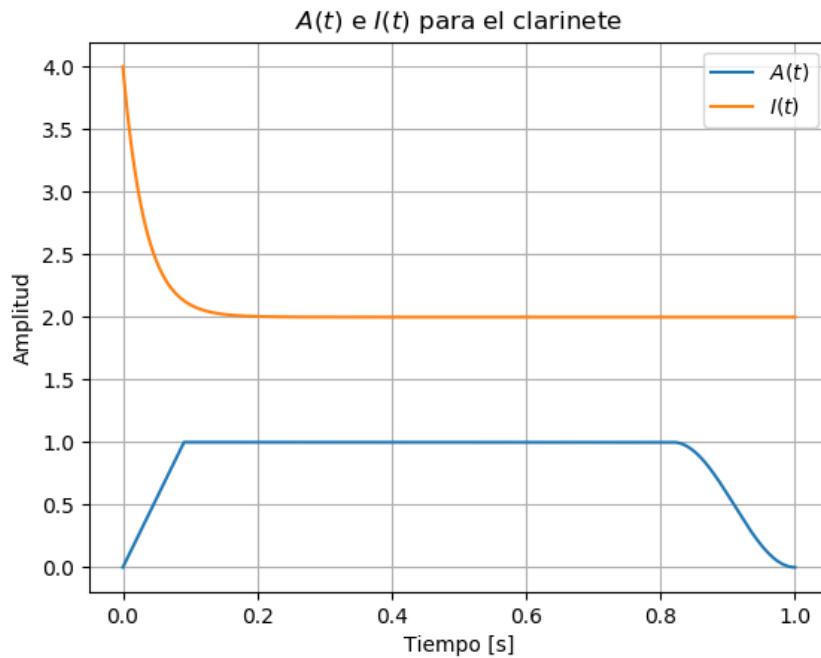
$$f_o = \frac{f_c}{N_1} = \frac{f_m}{N_2} \quad (6)$$

2.2. Instrumentos de viento

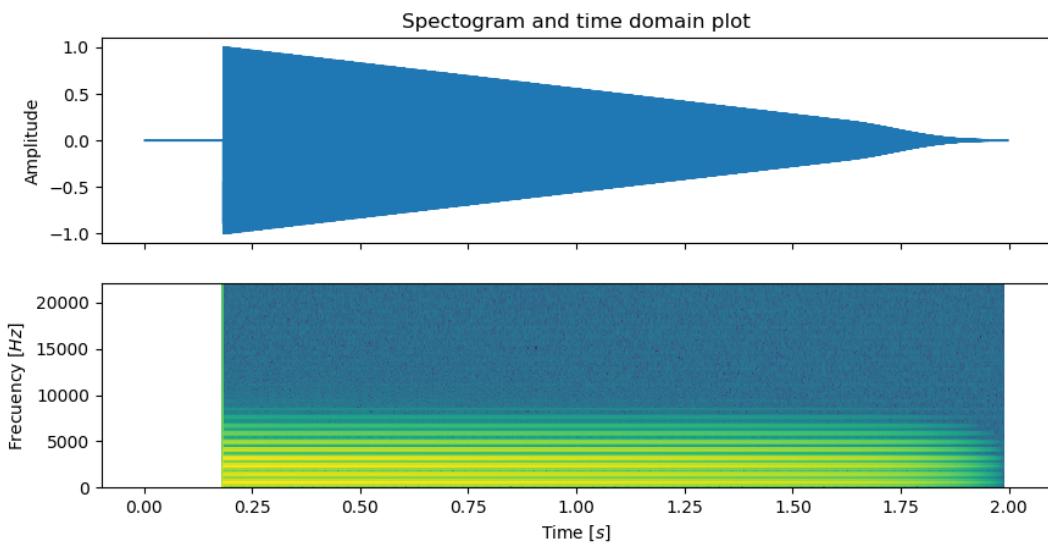
Este método puede ser empleado para sintetizar instrumentos de viento, fijando los parámetros $A(t)$, $I(t)$, f_c y f_m de una forma adecuada. Existen modelos dados para ciertos instrumentos, los cuales pueden ser modificados para conseguir un sonido más fidedigno, siendo el principal parámetro la relación presentada en la Ecuación (5).

2.2.1. Clarinete

Para al síntesis del clarinete se utilizó una relación $3f_c = 2f_m$. Además, las funciones $A(t)$ e $I(t)$ para una señal de una duración de 1 segundo son las presentadas a continuación.

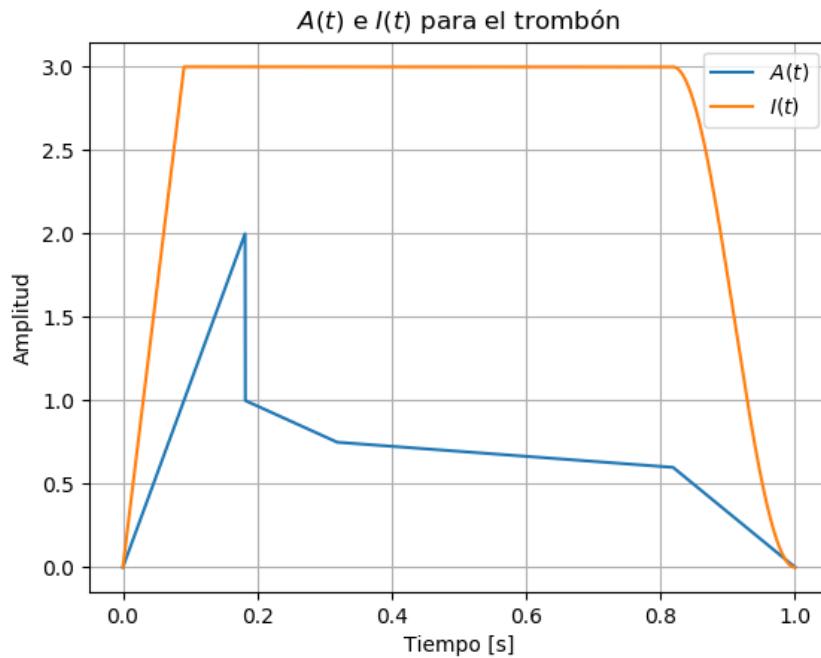
Figura 4: $A(t)$ e $I(t)$ para un clarinete.

Luego, se obtuvo un espectrograma para una nota a 440 Hz de dicho instrumento. Para dicho análisis se valió de una ventana de Hanning con 256 puntos para la $NFTT$ y una superposición de 128.

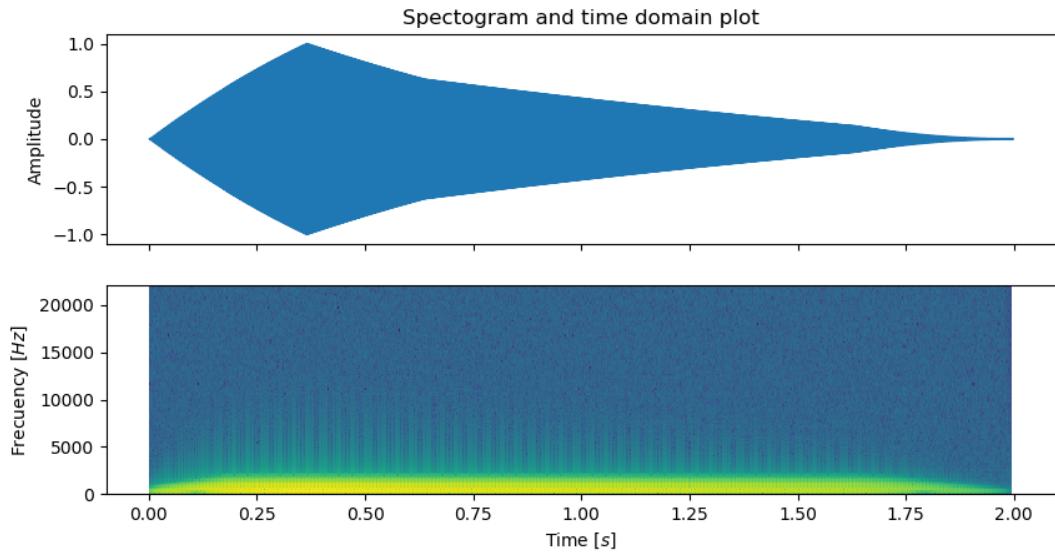
Figura 5: Espectrograma de una nota de clarinete a 440 Hz .

2.2.2. Trombón

De manera similar al caso anterior, para al síntesis del trombón se utilizó una relación $f_c = f_m$. Además, las funciones $A(t)$ e $I(t)$ para una señal de una duración de 1 segundo son las presentadas a continuación.

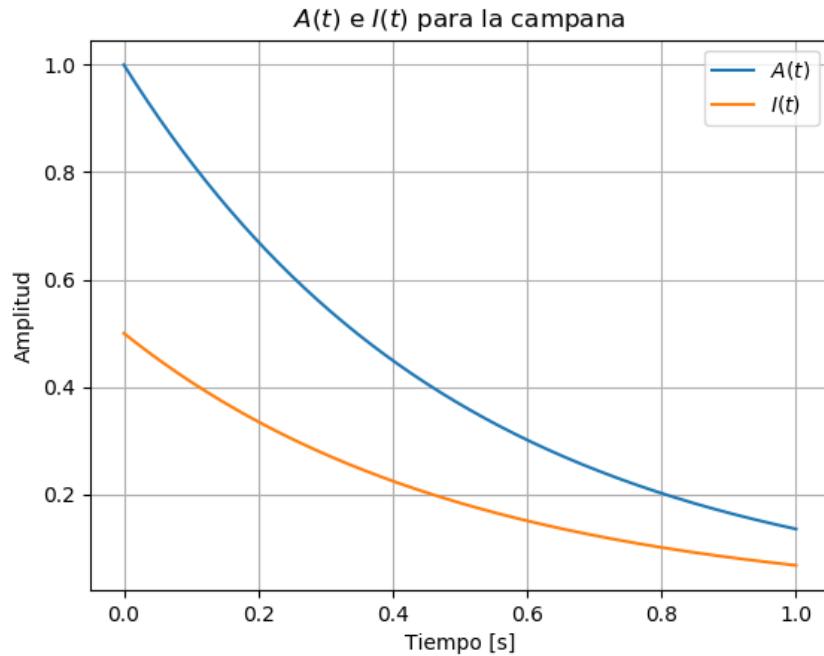
Figura 6: $A(t)$ e $I(t)$ para un trombón.

Después, se obtuvo un espectrograma para una nota a 440 Hz de dicho instrumento. Para dicho análisis se valió de una ventana de Hanning con 256 puntos para la $NFFT$ y una superposición de 128.

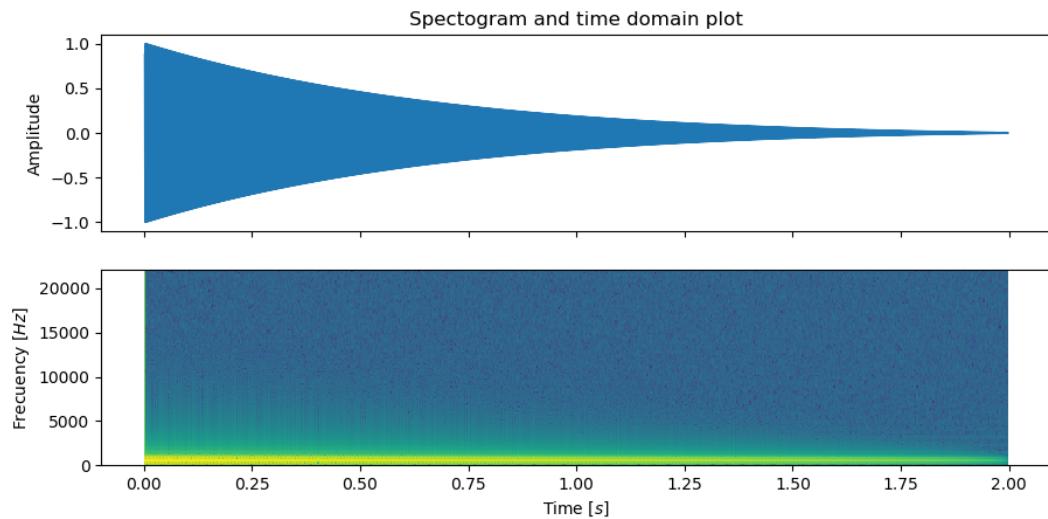
Figura 7: Espectrograma de una nota de trombón a 440 Hz .

2.2.3. Campana

Analogamente, para la campana se empleó una relación $f_c = 2f_m$. Además, las funciones $A(t)$ e $I(t)$ para una señal de una duración de 1 segundo son las presentadas a continuación.

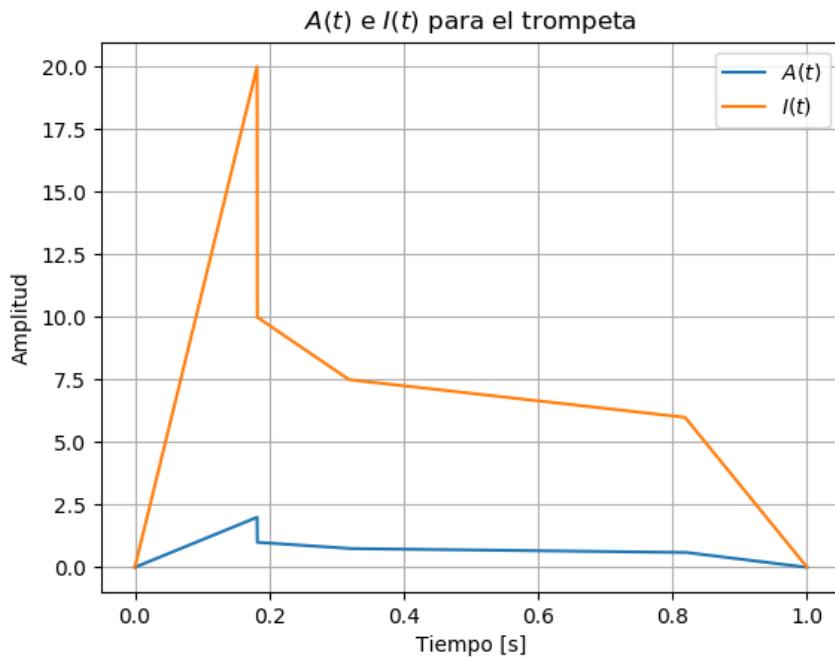
Figura 8: $A(t)$ e $I(t)$ para una campana.

Luego, se desarrolló un espectrograma para una nota a 440 Hz de dicho instrumento. Para dicho análisis se validó de una ventana de Hanning con 256 puntos para la $NFTT$ y una superposición de 128.

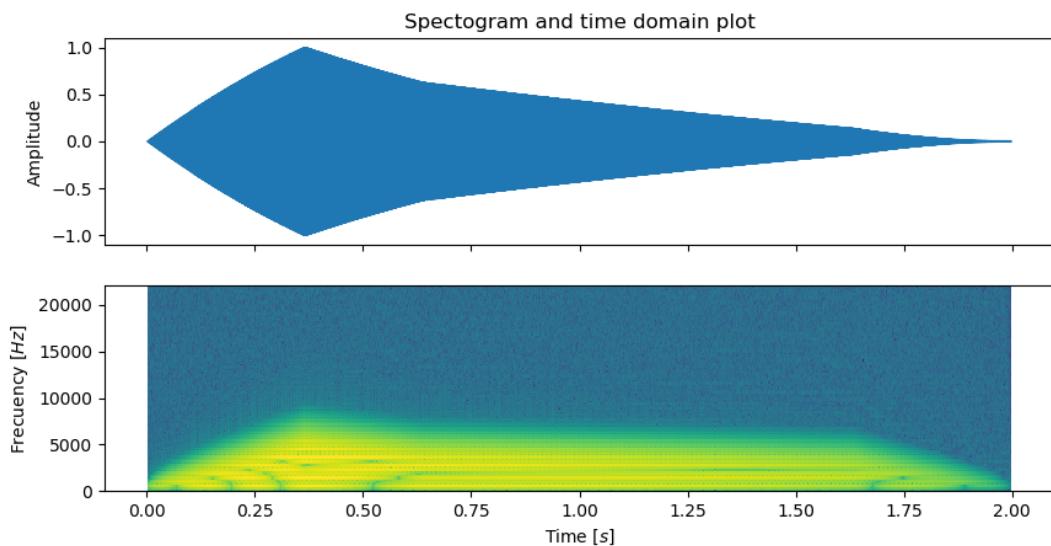
Figura 9: Espectrograma de una nota de campana a 440 Hz .

2.2.4. Trompeta

Finalmente, para sintetizar una trompeta se utilizó la misma relación que para el trombón, $f_c = f_m$. Además, las funciones $A(t)$ e $I(t)$ para una señal de una duración de 1 segundo son las presentadas a continuación.

Figura 10: $A(t)$ e $I(t)$ para una trompeta.

Luego, se obtuvo un espectrograma para una nota a 440 Hz de dicho instrumento. Para dicho análisis se valió de una ventana de Hanning con 256 puntos para la $NFFT$ y una superposición de 128.

Figura 11: Espectrograma de una nota de trompeta a 440 Hz .

2.3. Conclusiones

No existe un acercamiento analítico para determinar el mejor conjunto de parámetros para la síntesis de un instrumento mediante el método de F.M., lo que dificulta la selección de estos ². Es por ello que existen bases de como tratar cada instrumento, las cuales pueden ser libremente modificadas, dependiendo del gusto de cada persona, justificando la existencia variaciones de cada “plantilla”. A pesar de ello, dichos cambios deben hacerse con cierto

²Andrew Horner y James W. Beauchamp, [Instrument Modeling and Synthesis](#). 2009.

criterio, ya que se desea mantener un sonido similar al real, y provocar cambios sin tener conocimiento pueden generar sonidos muy distantes a los deseados.

3. Síntesis de sonidos mediante modelos físicos

En esta sección se analiza el método de síntesis basado en el modelado físico, propuesto por Karplus-Strong.

3.1. Karplus-Strong básico

El modelo básico de Karplus-Strong consiste filtrar una forma de onda a través de una linea de retardo. Gracias a esto se logra simular el sonido de una cuerda de guitarra.

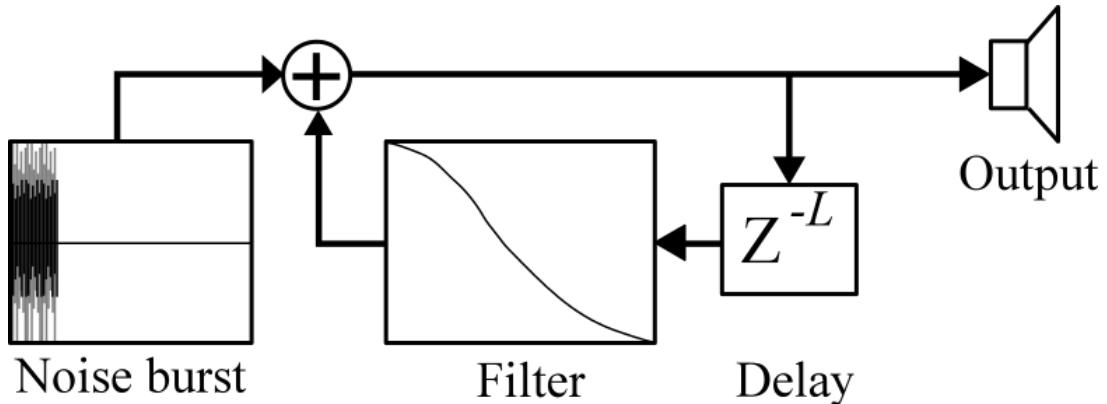


Figura 12: Modelo clásico Karplus-Strong.

3.1.1. Análisis teórico

Este algoritmo se puede describir por su diagrama en bloques como se ve a continuación.

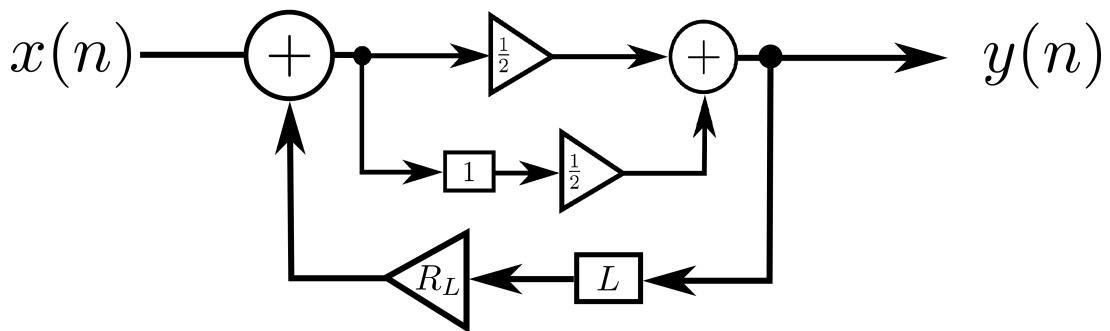


Figura 13: Algoritmo Karplus-Strong.

De este diagrama en bloques se puede obtener la ecuación en diferencias:

$$y(n) = \frac{1}{2} \cdot x(n) + \frac{1}{2} \cdot x(n-1) + \frac{1}{2} \cdot R_L \cdot y(n-L) + \frac{1}{2} \cdot R_L \cdot y(n-L-1) \quad (7)$$

A partir de esta expresión, se calcula su transformada Z y despeja para obtener la transferencia:

$$H(z) = \frac{\frac{1}{2} \cdot z^{L+1} + \frac{1}{2} \cdot z^L}{z^{L+1} - \frac{R_L}{2} \cdot z - \frac{R_L}{2}} \quad (8)$$

Vale la pena mencionar que de la Ecuación (7) es una ecuación en diferencias que cuenta como condiciones iniciales la “wavetable” suministrada por el ruido.

3.1.2. Análisis singularidades

Se observa que la Ecuación (8) cuenta con $L + 1$ polos y $L + 1$ ceros (de los cuales L de esos se encuentran en el origen). A continuación se muestra un diagrama de polos y ceros del sistema:

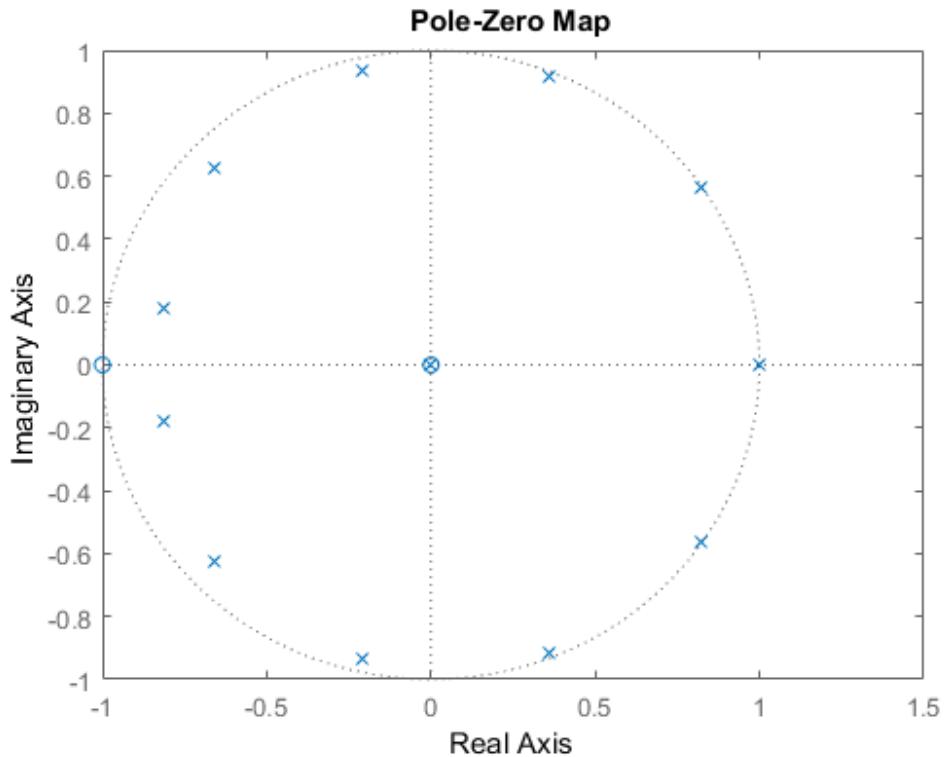


Figura 14: Diagrama de polos y ceros.

Adicionalmente, se graficó el diagrama de Bode del sistema.

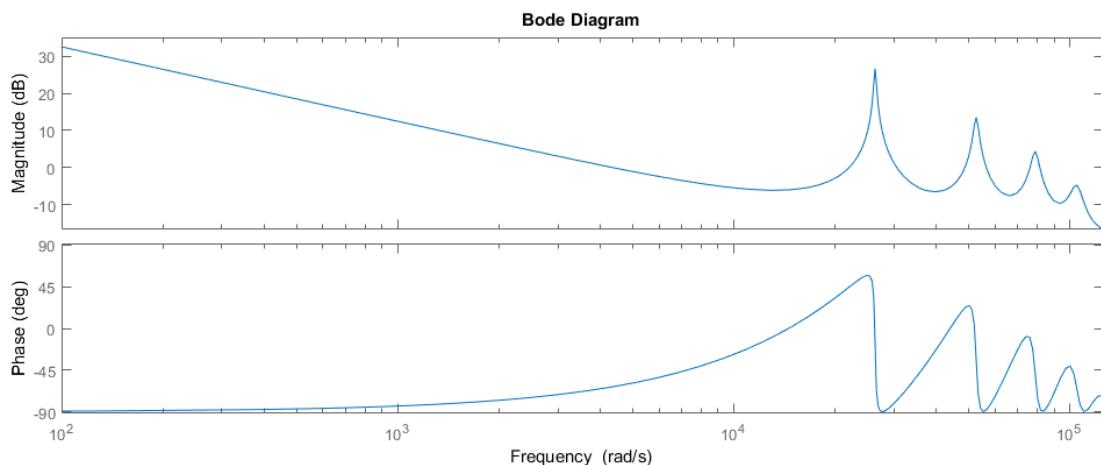


Figura 15: Diagrama de Bode.

Considerando que los parámetros eran: $f_s = 44.1 \text{ kHz}$, $L = 10$ y $R_L = 1$

3.1.3. Sintonización de frecuencia

En cuanto a la elección de una frecuencia de oscilación se puede observar en los últimos gráficos que existe un valor de frecuencia, la cual tiene mayor probabilidad de cumplir el criterio de Barkhausen, la cual corresponde a $f_r = \frac{f_s}{\sqrt{L+0.5}}$. Esto se debe a que el sistema es la superposición de una linea de retraso L junto con otro sistema de retraso $L + 1$. La señal al recorrer el lazo lo hace cada $\frac{L+L+1}{2}$ cambiando esto por frecuencia se obtiene:

$$f_r = \frac{f_s}{L + 0.5} \quad (9)$$

3.1.4. Tipos de ruido

Se propuso excitar el sistema con distintos tipos de ruido de entrada, siendo estos:

- Ruido Gaussiano
- Ruido Uniforme
- Ruido Binario

Primero se aplicó ruido gaussiano de longitud $L = 50$, y se obtuvo la siguiente salida:

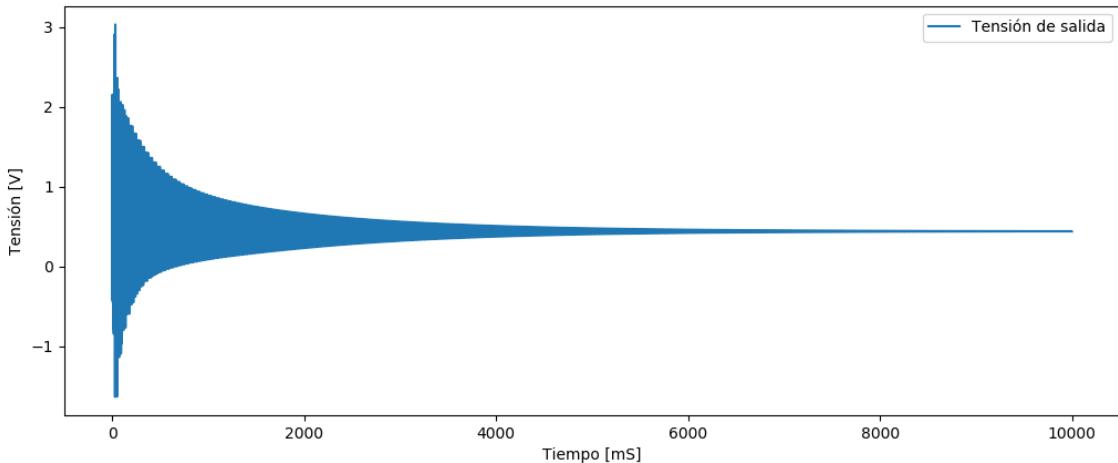


Figura 16: Respuesta a ruido gaussiano, $L = 50$.

Se puede apreciar que la respuesta a este tipo de ruido, parece tener cierta simetría respecto al eje. Algo notable de mencionar es como el sistema se estabiliza para un valor levemente superior a 0. Adicionalmente se realizó un detalle en la imagen:

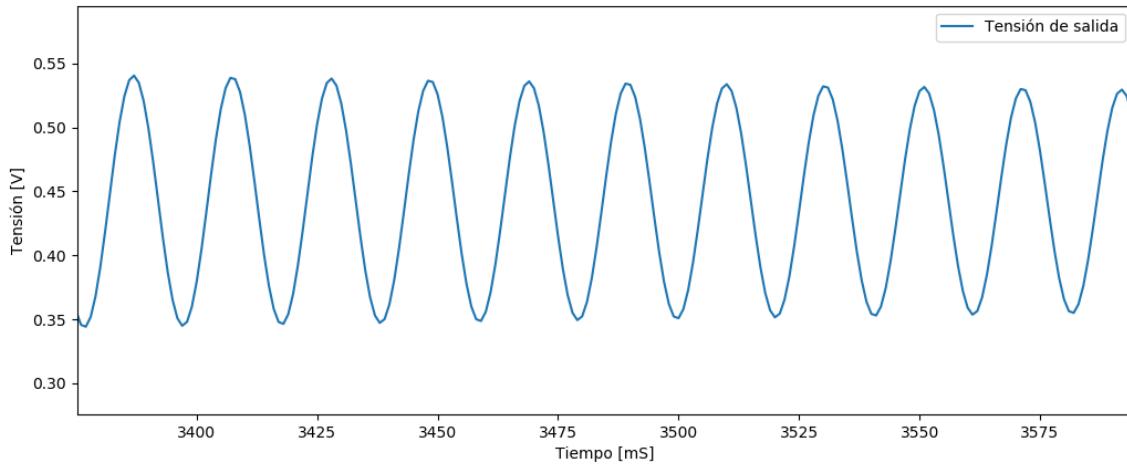


Figura 17: Respuesta a ruido gaussiano detalle, $L = 50$.

En esta, se puede apreciar que el sistema efectivamente se encuentra oscilando, mientras siendo atenuado por una envolvente.

Luego se excitó el sistema con ruido uniforme, obteniendo la siguiente respuesta:

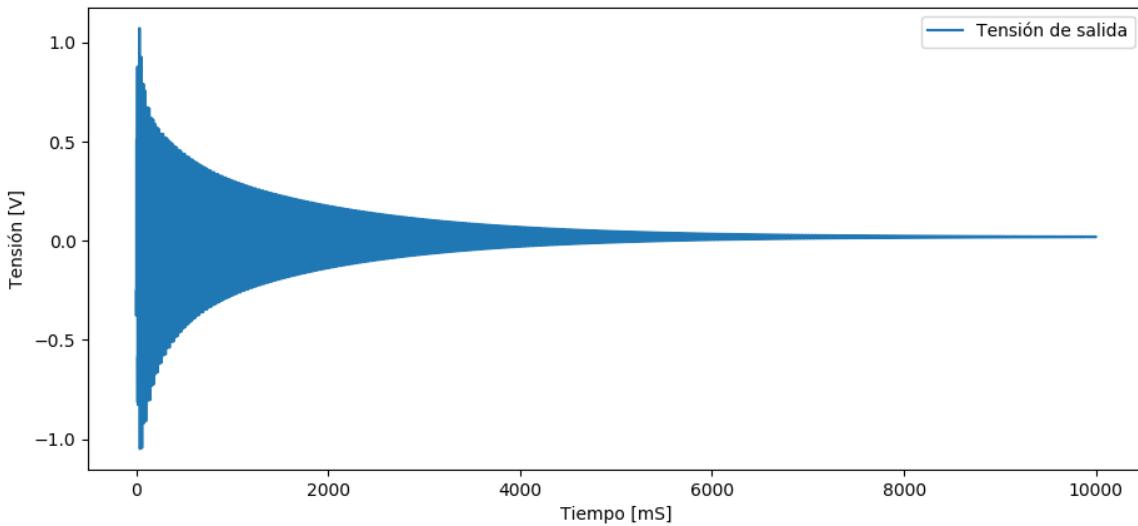


Figura 18: Respuesta a ruido uniforme, $L = 50$.

Se puede apreciar que la respuesta al ruido uniforme, cuenta con una simetría respecto al eje mas notable que el gaussiano, y un tiempo de decay más lento. Adicionalmente el valor al que tiende es mucho mas cercano a cero.

Luego, de la misma forma que para el gaussiano, se realizó un detalle en la figura:

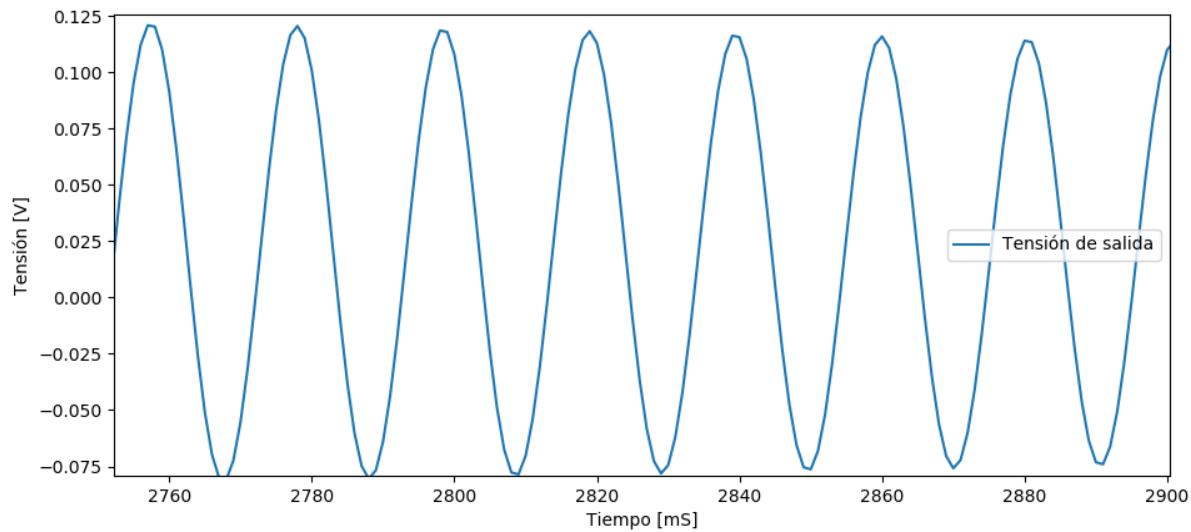


Figura 19: Respuesta a ruido uniforme detalle, $L = 50$.

Finalmente se ingresó al sistema con ruido Binario aleatorio.

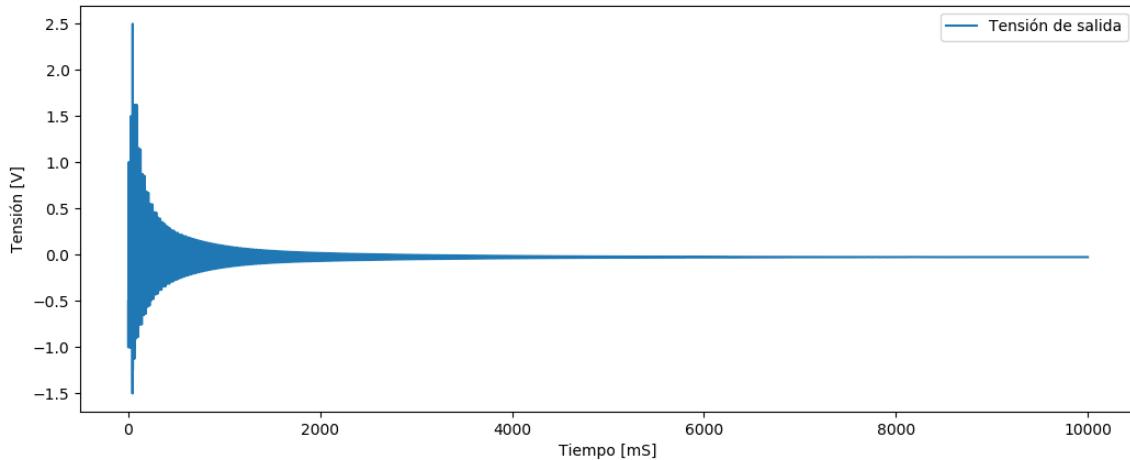
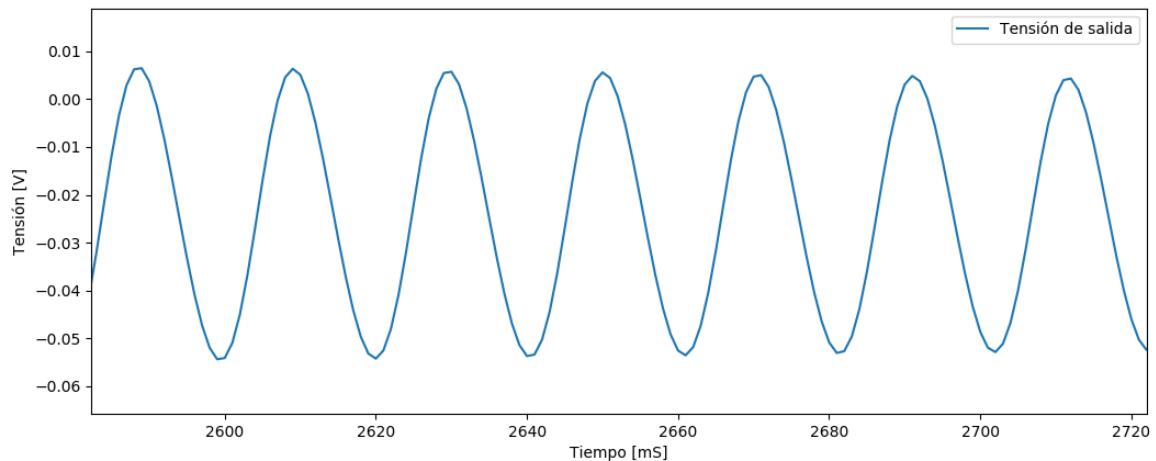
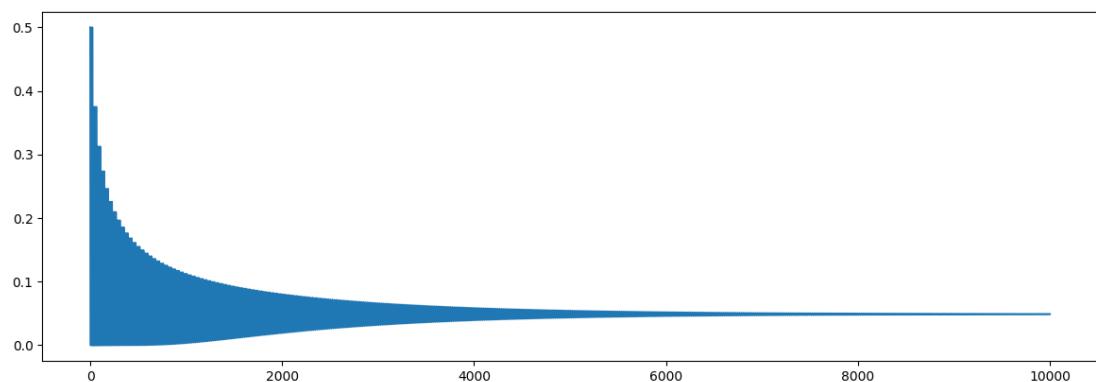


Figura 20: Respuesta a ruido Binario, $L=50$.

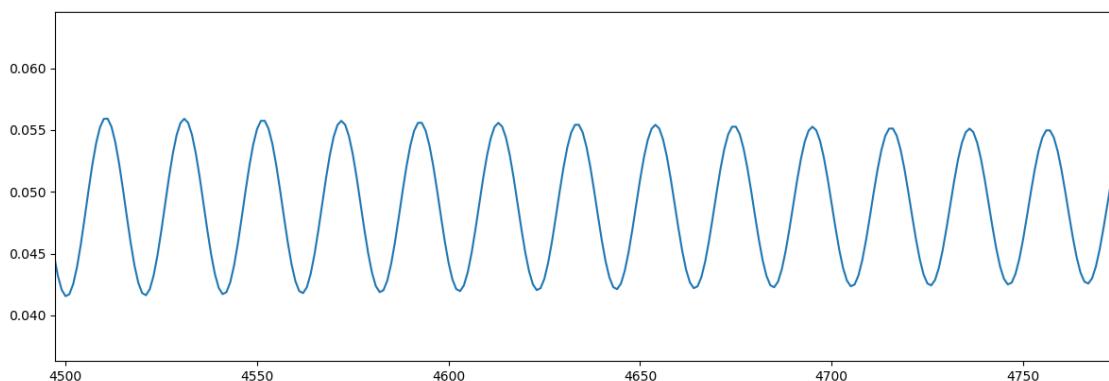
Se puede observar en la Figura (20) que existe una envolvente mucho más severa que las anteriores. Ademas, a medida que el tiempo aumenta, el valor de tensión tiende cercano a cero, lo cual es un efecto deseado.

Figura 21: Respuesta a ruido Binario en detalle , $L = 50$.

Adicionalmente se computó la respuesta impulsiva del sistema, siendo esta:

Figura 22: Respuesta al Impulso, $L = 50$.

Vale la pena mencionar que la salida nunca es negativa, esto se debe a que cuando la excitación es únicamente positiva, dado que la realimentación es positiva y que no existe ninguna inversión de fase, la salida siempre resulta positiva. Así también se muestra un detalle de la respuesta.

Figura 23: Respuesta al Impulso detalle, $L = 50$.

También cabe destacar que la “velocity” fue introducida no solo al final de la sintetización como modulador de volumen, sino también en el ruido del comienzo, dado que dicha propiedad, en la guitarra, simboliza la intensidad con la cual fue tocada la cuerda.

3.1.5. Estabilidad

La estabilidad del sistema es determinada por la Ecuación (8). Se puede observar que si RL es mayor o igual a la unidad, el sistema es inestable. Si bien, teóricamente es cierto, en la realidad se encuentra que si se da dicha condición, no solo no se provoca inestabilidad, sino que es preferible este valor dado, ya que logra extender las oscilaciones un mayor tiempo.

3.2. Mejora propuesta

El sistema anterior cuenta con algunas limitaciones. Un claro ejemplo de esto es la frecuencia, a cual para valores pequeños de L , la diferencia entre $f_r(L)$ y $f_r(L + 1)$ resulta grande, dejando una banda de frecuencias sin poder ser sintonizadas, como ejemplifica la siguiente imagen.

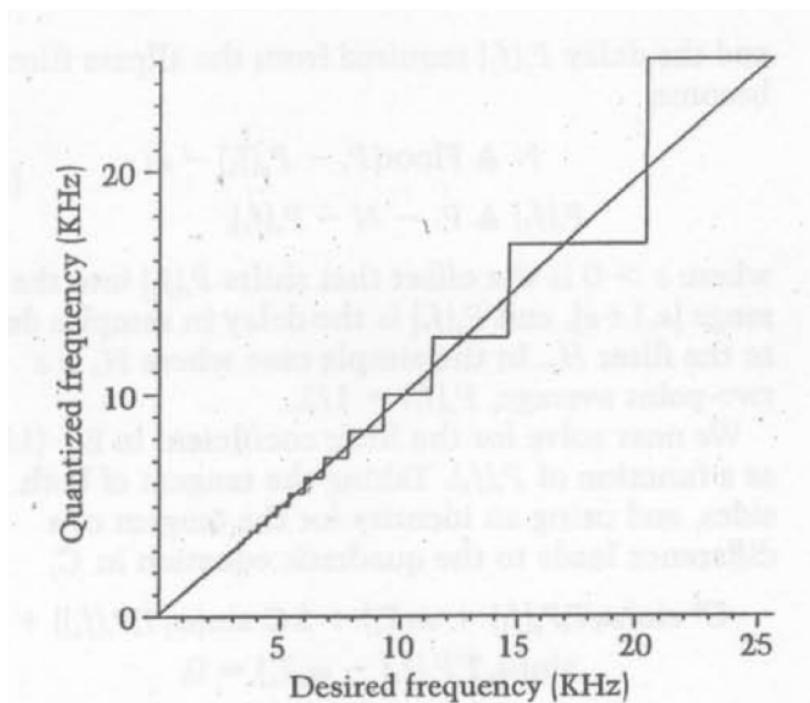


Figura 24: Dificultad en sintonización de frecuencia.

Otro defecto, es el final abrupto con el que cuentan algunas notas. Si se pide una duración reducida, las discontinuidades en el sonido provocan un efecto indeseado.

3.2.1. Sintonización de frecuencia

Para definir la frecuencia de resonancia del sistema, dado que la Ecuación (9) no deja grados de libertad, ya que la frecuencia de muestreo es fija, lo único que se puede hacer es realizar una modificación al circuito. El nuevo modelo propuesto es el siguiente:

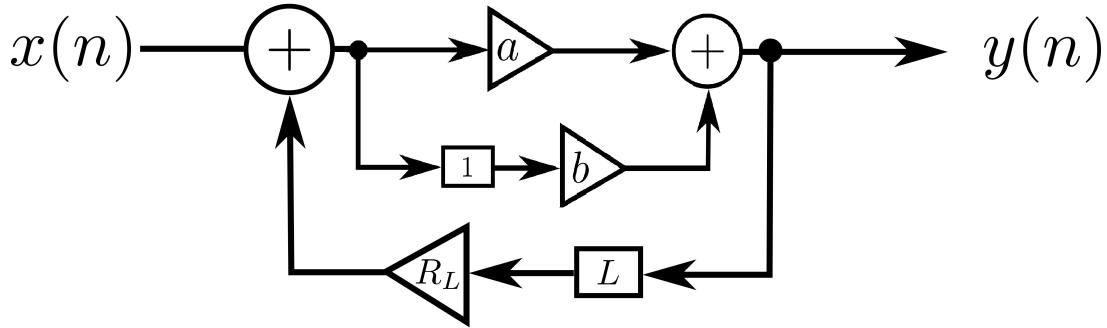


Figura 25: Algoritmo Karplus-Strong mejora propuesta.

Utilizando esta propuesta se puede obtener la ecuación en diferencias del sistema, siendo esta:

$$y(n) = a \cdot x(n) + b \cdot x(n - 1) + a \cdot R_L \cdot y(n - L) + b \cdot R_L \cdot y(n - L - 1) \quad (10)$$

Nótese, que si $a = b = \frac{1}{2}$ el sistema coincide con el original. Utilizando el mismo razonamiento, se llega a que la nueva frecuencia de resonancia es:

$$f_r = \frac{f_s}{L \cdot (a + b) + b} \quad (11)$$

Se debe elegir apropiadamente a y b , teniendo en cuenta que su suma debe de dar 1 para que el sistema sea el promedio ponderado realizado por estos sea realmente un promedio.

Un punto en contra de esta solución es que, dependiendo de los valores de a y b , se toma más en consideración una u otra linea del promedio.

3.2.2. Continuidad del sonido

Se tuvo en cuenta que el sonido sea una función suave, dado que cambios bruscos en el sonido no son placenteros ni esperados en la sintetización que es deseada implementar. Para lidiar con este problema lo que se realizó fue definir un factor de ventana, a partir de la cual se atenuó gradualmente el sonido hasta que sea nulo. Para esto se definió una ventana que tiene valor unitario para valores de tiempo menores al factor de ventana especificado, mientras que para valores superiores, decréce gradualmente con una función cosenoide, barriendo desde 0 a $\frac{\pi}{2}$ como se ilustra a continuación.

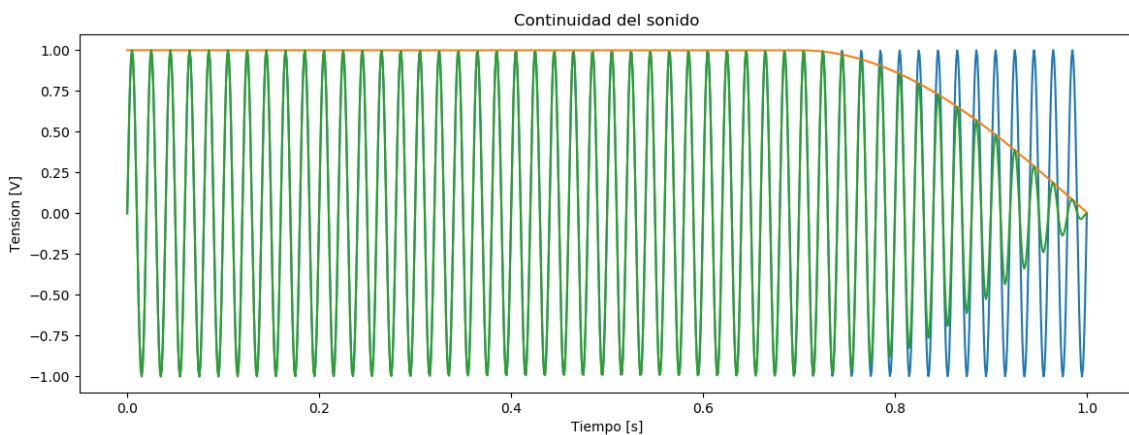


Figura 26: Muestra del factor de ventana.

De esta forma se aumenta la calidad de la síntesis en gran medida si se combina con la idea de que la duración de la nota y la del sonido son cosas distintas.

3.2.3. Caja de resonancia

La caja de resonancia es una parte primordial de la gran mayoría de instrumentos acústicos, principalmente de cuerda y percusión. Tiene la finalidad de amplificar o modular un sonido (en los instrumentos de cuerda generalmente a través de un puente). Los instrumentos que cubren rangos de sonidos más graves, como el contrabajo, el violonchelo o el bombo, necesitan una caja de resonancia bastante mayor que el resto.

La necesidad de utilizar una caja de resonancia en una guitarra se debe justamente a que la vibración de las cuerdas, que generan los frentes de onda esféricos propios del sonido, tienen una gran dificultad para propagarse por el medio (el aire) debido a su reducida amplitud.

La manera en la cual este elemento sopesa dicho problema es capturando parte del frente de onda en su cavidad. Allí las frecuencias producidas por las cuerdas son amplificadas en igual magnitud y luego liberadas al medio con la intensidad suficiente para que pueda propagarse y que sea apreciable el sonido.

Finalmente cabe destacar que en el ámbito de señales realmente no es necesario dicha caja, al igual que tampoco un filtro que la emule, debido a que su función es amplificar, efecto que se puede hacer digitalmente sin ningún problema.

3.3. Karplus-Strong percusión

Se puede realizar una modificación al modelo inicial agregando un factor aleatorio como se observa a continuación:

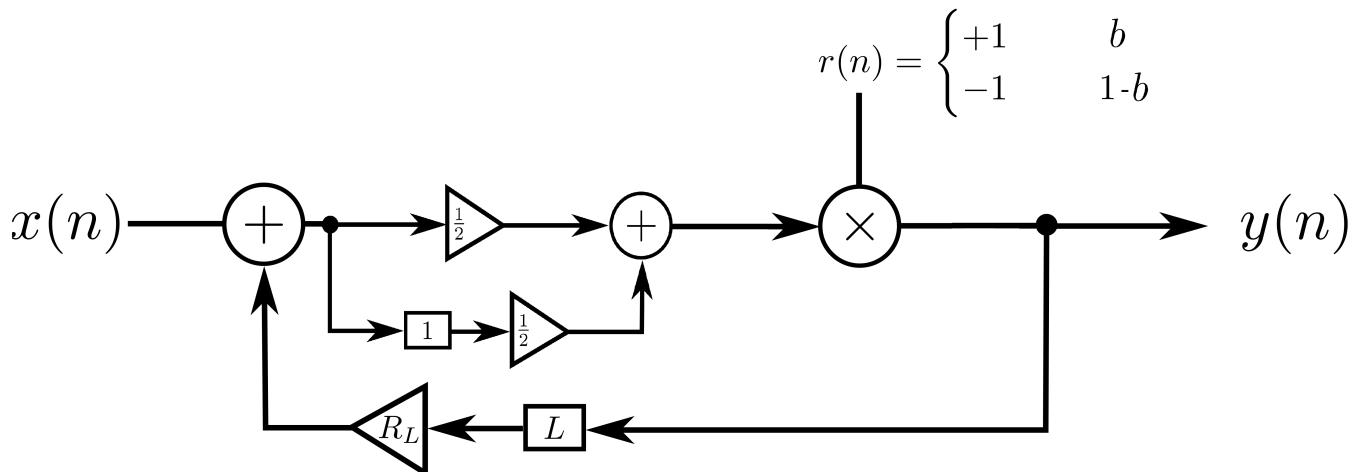


Figura 27: Karplus-Strong percusión.

El valor de b indica el nivel de cambio para el sonido final. Vale la pena notar que si $b = 1$, el algoritmo resulta igual al original. Para obtener sonidos de percusión se utiliza un valor de $b = 0,5$.

Una gran diferencia entre la guitarra sintetizada y un elemento de percusión es que estos últimos no cuentan con “notas”, sino que tienen un sonido característico, para obtener un sonido similar a un elemento de percusión se utilizan valores elevados de L .

3.4. Espectrogramas

Finalmente se realizaron espectrogramas de tanto la guitarra como el “redoblante” obteniendo los siguientes resultados.

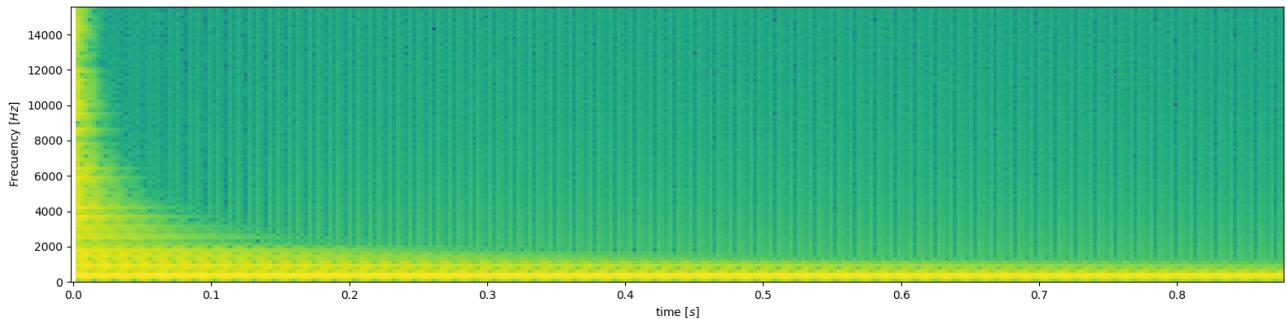


Figura 28: Espectrograma Guitarra.

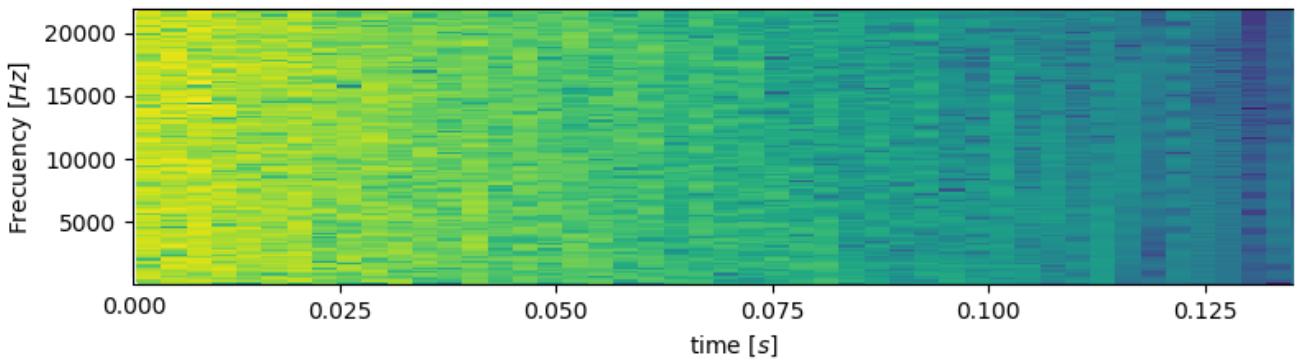


Figura 29: Espectrograma Redoblante.

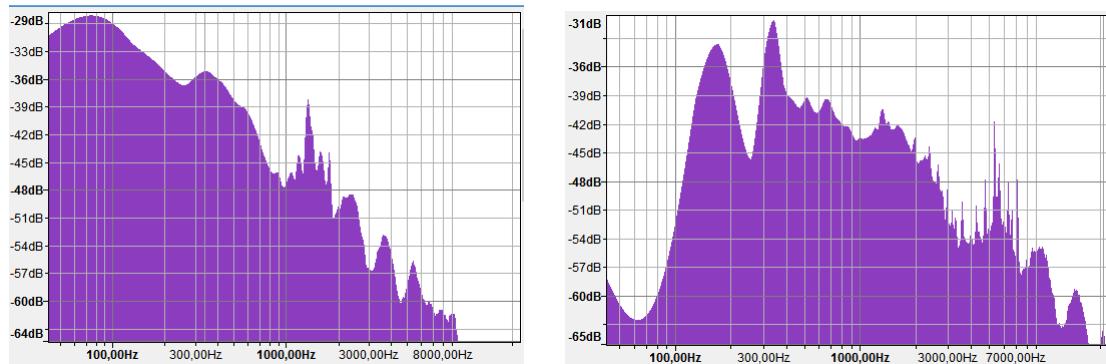
Se puede diferenciar que la guitarra cuenta con mayor densidad armónica de baja frecuencia al igual que se extiende considerablemente mas en el tiempo, mientra que el redoblante cuenta con mayor contenido de alta frecuencia y su duración es mucho menor.

4. Síntesis por muestras

La Síntesis por muestras tiene sus orígenes en los estudios de grabación. Previo a las técnicas de síntesis se debían grabar nuevamente aquellos sonidos que no satisfacían por completo a los músicos o a los ingenieros de sonido, obligándolos a utilizar más horas el estudio de grabación, lo cual llevaba varios gastos asociados y retrasaba la culminación del proyecto.

4.1. Time Stretching

Supongase que se desea grabar un comercial de televisión y por cuestiones legales se debe añadir una pequeñas aclaraciones al final del mismo. Dado que el tiempo de aire es altamente costoso, se tiene que poder incluir toda la información necesaria en un espacio de tiempo muy pequeño. Entonces es posible suponer que basta con grabar el mensaje a velocidad de habla normal y luego reproducirlo más rápido para que ocupe menos tiempo. Es decir reproducir el sonido saltándose segmentos para poder acortar la duración.



(a) Espectro de un discurso hablado a velocidad normal.

(b) Espectro del discurso con una velocidad 4 veces mayor.

Figura 30: Consecuencias de la compresión temporal

Como se puede observar, la Figura (30a) representa el espectro de habla de una persona hablando a una velocidad promedio. Nótese que la zona de mayor contenido armónico se halla aproximadamente entre los 80 Hz y 200 Hz . Sin embargo, si se mira al espectro del mismo discurso pero ahora con 4 veces menos duración, Figura (30b), es posible observar que el espectro no se preservó. De hecho, se puede notar que hay mayor potenciapectral a frecuencias más altas que en el espectro original. Esto se traduce en un sonido chillón que poco recuerda al discurso original. De formaanáloga, si por alguna razón se quisiera aumentar o reducir el pitch de una pista de audio se, debe recurrir al método de acortar la pista o en el caso contrario alargarla. En la siguiente sección se explora como es posible controlar la duración y el pitch de una pista de audio de manera independiente.

4.2. TD-PSOLA

TD-PSOLA son las siglas para Time Domain Pitch-Synchronous Overlap-Add. Los algoritmos basados en **PSOLA** reutilizan pequeños segmentos llamados **short term signals**, que son el resultado de aplicar una ventana de **Hanning**, la cual se extiende hacia los **pitch-marks** vecinos, sobre cada **pitch-mark** (más acerca de **pitch-marks** en la siguiente sección) y solaparlos convenientemente.

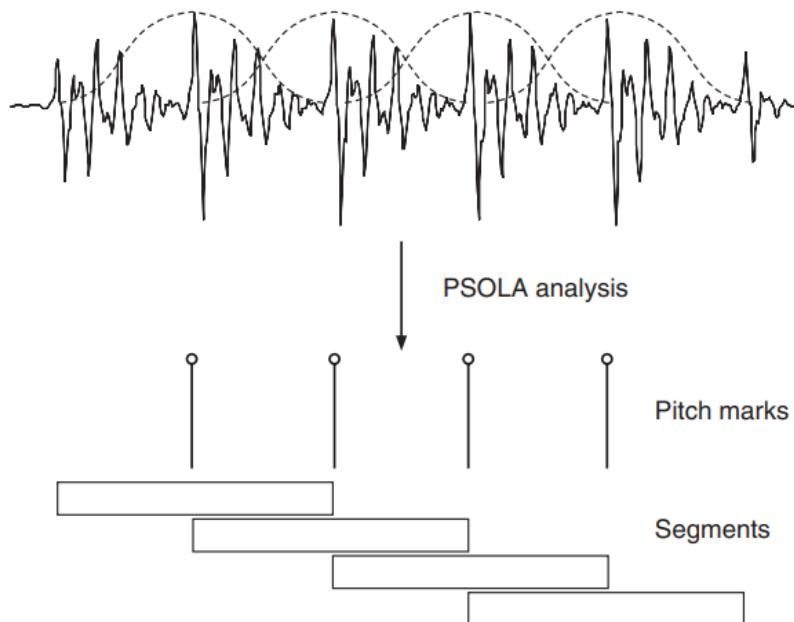


Figura 31: Segmentación del sonido de acuerdo a los pitch-marks.

Para poder conseguir alargar la duración de los sonidos sin alterar el **pitch** del mismo, es necesario conservar la forma de los segmentos que la conforman.

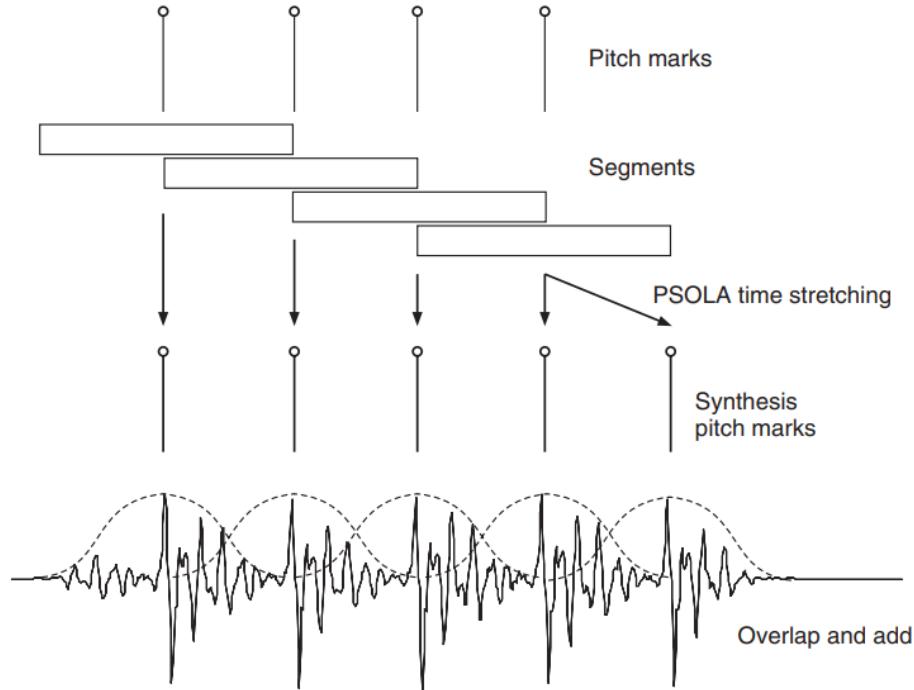


Figura 32: Ilustración del estiramiento temporal usando PSOLA.

Esto se expresa dentro del algoritmo de estiramiento como un factor $\alpha = \frac{d_{nueva}}{d_{original}}$, que permite mapear la posición de los pitch-marks dentro del sonido original hacia el sonido de llegada. Las ventanas de Hanning son de gran utilidad porque permiten transiciones más suaves entre cada segmento.

Para poder tener control sobre el **pitch** del sonido, se debe cambiar la distancia entre los pitch-marks para así afectar a la frecuencia fundamental f_0 del sonido. Análogamente, este escalado en frecuencia viene denominado como un factor $\beta = \frac{f_{nueva}}{f_{original}}$.

Estos dos factores en combinación permiten regular tanto el **pitch** como la duración del sonido.

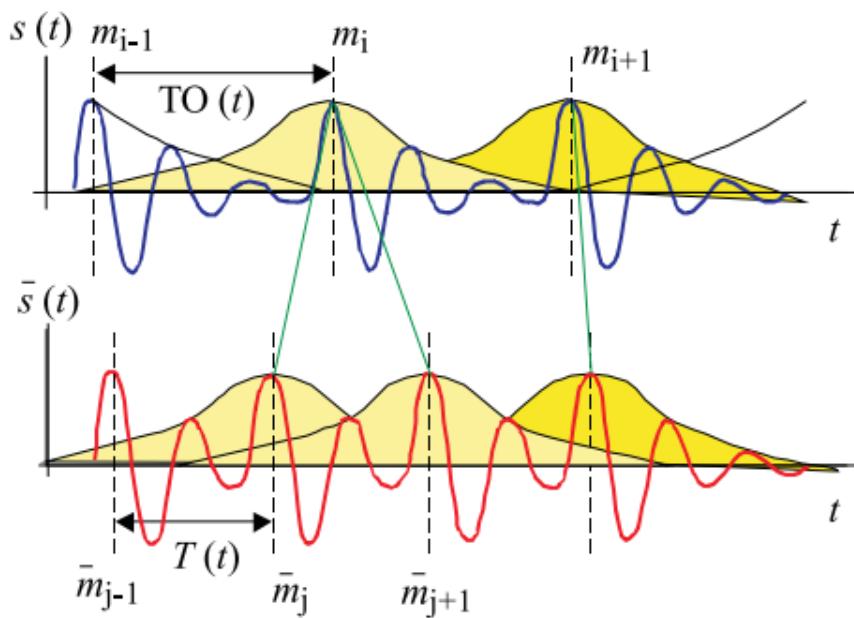


Fig. 2. Example of pitch-shifting and time stretching using PSOLA

Figura 33

4.3. Estimación de los Pitch-Marks

Los **pitch-marks** representan momentos del sonido en el que su amplitud es máxima en un entorno. En realidad este término se emplea cuando se habla de **glottal pulses** dentro del campo de Voice Processing, y hacen referencia a ciertos impulsos de aire generados nuestra habla. Estas posiciones son centrales para la estructura del sonido. Poder localizarlos con precisión es un punto clave para los algoritmos que se basan en ellos para sintetizar nuevos sonidos.

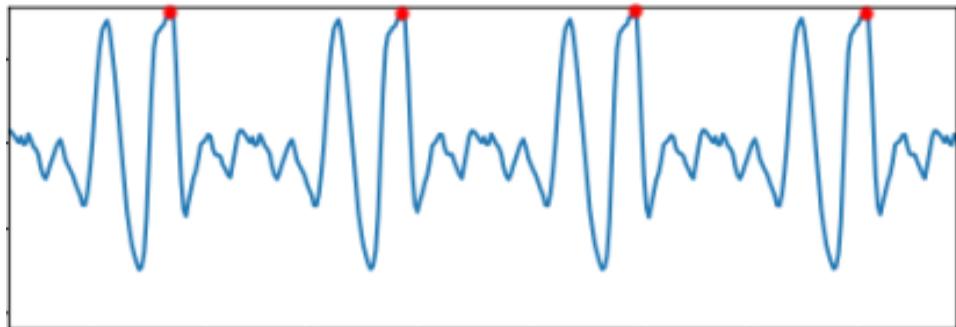


Figura 34: Imagen aumentada de una nota musical.

En la Figura (34) se pueden apreciar los pitch-marks estimados para la nota **A4**. En este caso es posible decir, mediante inspección visual, que se la estimación ha tenido un éxito considerable. Sin embargo, estimarlos para una sonido más complejo como puede ser la voz humana o una pista de audio no es una tarea sencilla y se deben tener en cuenta varios factores. Para la estimación de los pitch-marks de las notas se utilizó el programa *Audacity* para visualizar el espectro y obtener la frecuencia fundamental de la nota $f_0 = 395\text{Hz}$. A partir de la frecuencia fundamental y el previo conocimiento del **sample rate** de la muestra se realizan las siguientes hipótesis.

$$P \approx 1/f_0 = 2.53\text{ ms}$$

Es decir que P , el **pitch-period** (tiempo entre pitch-marks) tiene aproximadamente esa duración.

$$M_p = \frac{\text{sampling rate}}{f_0} = \frac{44100}{395} \approx 112 \text{ muestras}$$

Este dato indica que entre cada pitch-mark existen 112 puntos de espacio. Con este dato es posible utilizar el paquete científico *Scipy* y calcular los máximos de la señal que estén equiespaciados por lo menos 112 muestras entre sí.

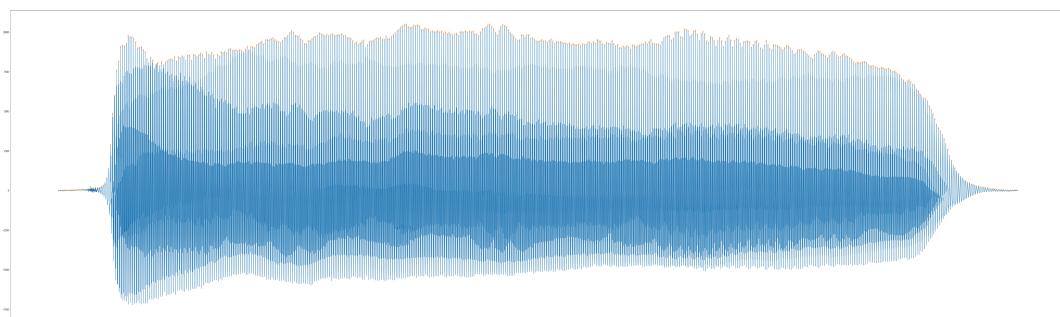


Figura 35: Pitch Marks en una nota A4 de saxofón, $f_0 = 395\text{ Hz}$ (se denota una mejor apreciación aplicando zoom en la parte superior).

Como se puede apreciar en la Figura (36) este método funciona bien para la nota A4. Sin embargo al probar este método sobre la nota A#6 no se obtiene resultados óptimos.

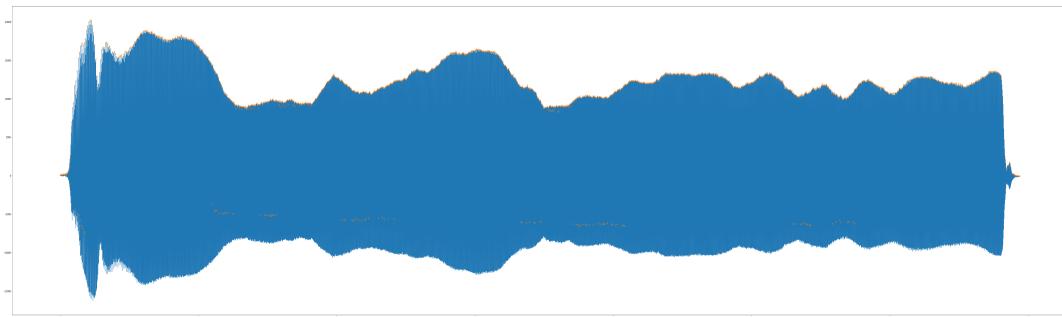


Figura 36: Pitch Marks en una Nnota A6 de saxofón (se denota una mejor apreciación aplicando zoom en la parte superior).

5. Efectos de Audio

Se implementaron dos efectos, eco simple y reverberación, siguiendo el método descrito por M. Schroeder³. Para el eco se utilizó una sola linea de delay mostrada en la Figura (37).

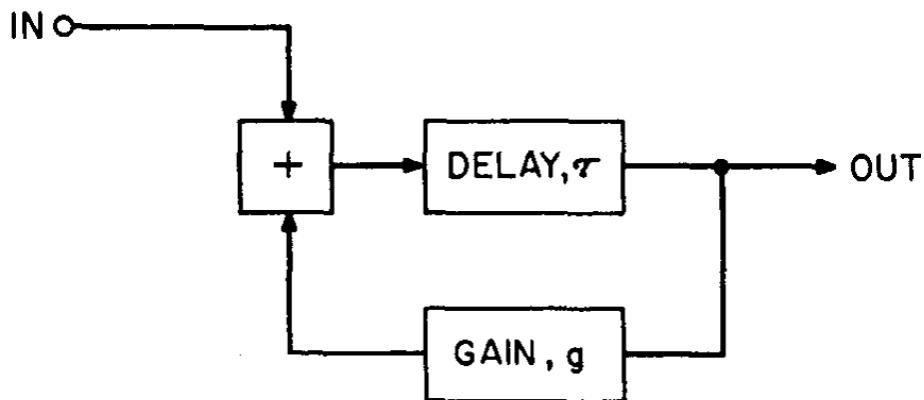


Figura 37: Linea de retardo.

Los parámetros para configurar este efecto son los de delay time τ , el cual denota el retardo de la línea; y decay factor g , el cual describe la ganancia del lazo. Se puede modelizar el tiempo en el cual el sonido decae por 60 dB, T , como

$$T = \frac{3 * \tau}{-\log_{10}|g|} \quad (12)$$

Luego, se utilizó el modelo propuesto por M. Schroeder mostrado en la Figura (38) para implementar la reverberación, con cuatro lineas de retardo en paralelo, seguidas de dos filtros pasa todo con linea de retardo, los cuales aumentan la densidad de los ecos sin afectar la ganancia del sistema en función de la frecuencia.

³M. Schroeder, "Natural Sounding Artificial Reverberation", Journal of the Audio Engineering Society, vol. 10, no. 10, 1962. [Accessed 9 May 2020].

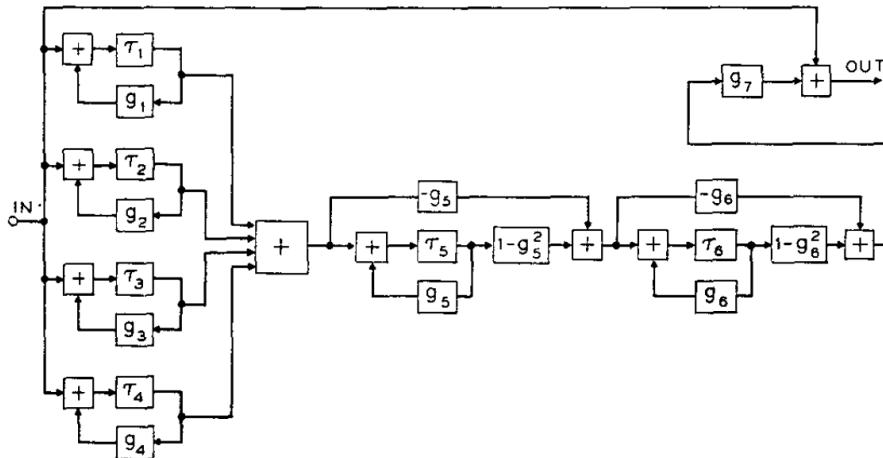


Figura 38: Filtro all-pass con linea de retardo.

Se utilizaron los parámetros sugeridos por Schroeder detallados a continuación⁴

Parámetro	Sugerencia
τ_5	5 ms
τ_6	1.7 ms
g_5	0.7
g_6	0.7
τ_1	101.560 ms
τ_2	113.356 ms
τ_3	122.426 ms
τ_4	131.54 ms

Tabla 1: Parámetros utilizados según las referencias empleadas.

Luego, se ajustaron los valores de g_1 hasta g_4 según (12). Finalmente, para el eco se le pide al usuario que ingrese el delay time y el decay factor, denominados en el programa con las letras T y D . Para la reverberación, se le pide al usuario el tiempo de reverberación, el cual es el tiempo en el que el sonido decae 60 dB, denotado con la letra T ; y el mix factor, detallado en la Figura (38) como g_7 , denotado en el programa con la letra M .

6. Programas implementados

6.1. FFT

Se implementó la FFT utilizando el algoritmo de Cooley-Tukey de manera recursiva. Se probó con diversas entradas reales aleatorias, de tamaño 4096, con una media temporal de 40 μs .

6.2. Programa Principal

Se desarrolló un programa, el cual permite agregar midis de cualquier duración, para luego sintetizar cada track deseado en dicho archivo. Permite seleccionar diversos instrumentos, varios de ellos con parámetros modificables. Además, es posible compilar dichos tracks en un archivo del tipo wav y hasta generar un archivo de “preview” de un track dado, es decir una pequeña muestra de como se escucha un track en particular.

También es posible elegir efectos a aplicar, tanto para el archivo wav final, como para cada track en particular.

⁴John M. Chowning, *The Synthesis of Complex Audio Spectra by Means of Frequency Modulation*. 1973.

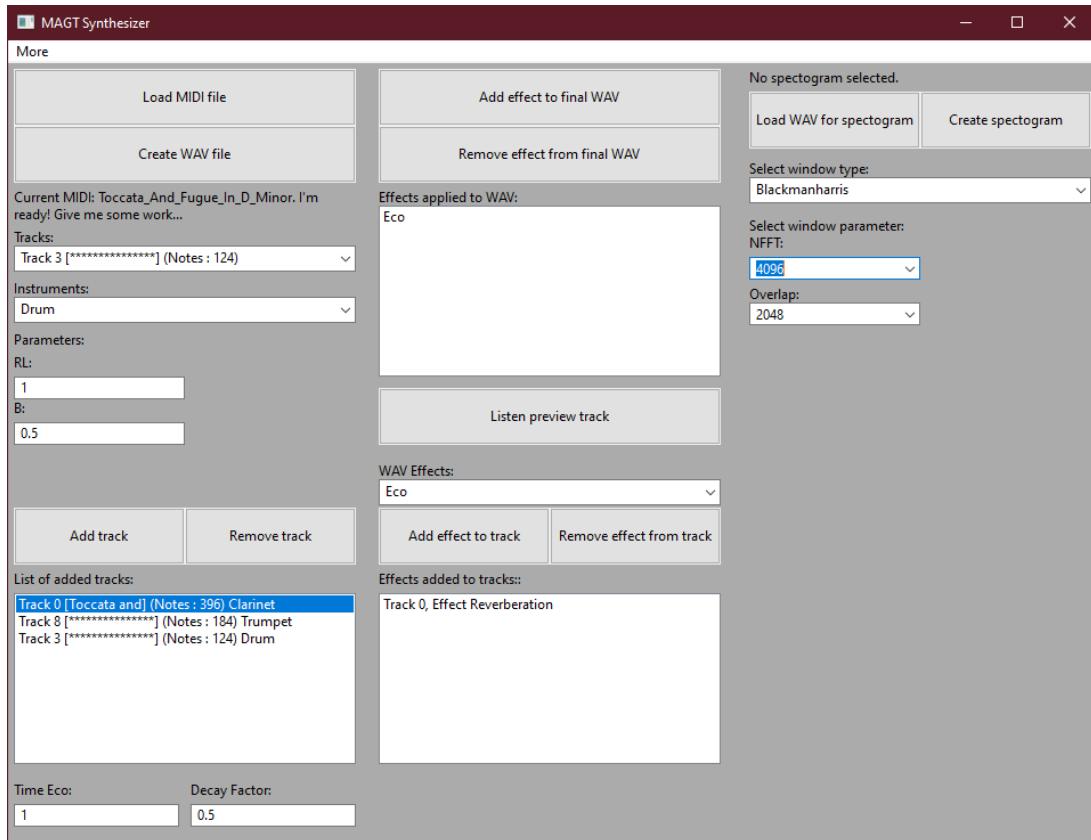


Figura 39: Captura de la GUI implementada.

Finalmente se destaca que es posible realizar espectrogramas de cualquier archivo wav, sea generado por este programa o no, pudiendo elegir el tipo de ventana, la cantidad de puntos de la FFT y el factor de overlap.

Cabe destacar que el programa fue desarrollado en su totalidad en C++, a excepción de los gráficos del espectrograma, los cuales son generados mediante un llamado a un código en Python. El front-end se implementó mediante el uso de la librería WxWidgets.