
XIII Congreso Argentino de Acústica
VII Jornadas de Acústica, Electroacústica y Áreas Vinculadas
Buenos Aires, 29 y 30 de octubre de 2015

AdAA2015-A003

**Modificación de timbres armónicos utilizando
las soluciones de la ecuación de la onda**

Julián A. Tachella^(a),
Antonio, Quintero-Rincón^(b)

(a) Departamento de Ingeniería Electrónica, Instituto Tecnológico de Buenos Aires. Eduardo Madero 399, Ciudad Autónoma de Buenos Aires, Argentina. E-mail: jtachell@itba.edu.ar

(b) Departamento de Bioingeniería, Instituto Tecnológico de Buenos Aires. Eduardo Madero 399, Ciudad Autónoma de Buenos Aires, Argentina. E-mail: aquinter@itba.edu.ar

Abstract

A new digital timbre transformation that modifies harmonic sounds to inharmonic ones changing the frequency of the subharmonics is proposed. In order to perform the modifications, two digital signal processing techniques are discussed. The first one is a real time oriented algorithm and the second one is an analysis and resynthesis process. The results obtained using these techniques suggest that is possible to transform the timbre without having spectral leakage or significant loss in time resolution.

Resumen

En este trabajo se propone una nueva transformación digital de timbres musicales que convierte timbres armónicos en inarmónicos, alterando la frecuencia de los parciales secundarios del timbre original. Dicha transformación utiliza como punto de partida las soluciones de la ecuación de la onda. Además, se plantean dos técnicas de procesamiento de audio digital para realizar la transformación, la primera orientada a una modificación de señales de audio en tiempo real y la segunda como un proceso de análisis de una muestra y posterior síntesis modificando los parámetros deseados. Los resultados obtenidos sugieren que es posible realizar la transformación tanto en tiempo real como por resíntesis evitando derrame espectral (spectral leakage) entre subarmónicos y sin pérdidas significativas en la resolución temporal.

1 Introducción

Diversas técnicas digitales han sido utilizadas en las últimas décadas para modificar el timbre de instrumentos armónicos, especialmente para guitarras eléctricas (Zölzer, 2002). La mayoría se basa en sistemas lineales y causales (tiempo variantes o tiempo invariantes). En los últimos años han aparecido nuevas modificaciones basadas en sistemas adaptativos no lineales o no causales, que poseen cierta latencia, como armonizadores y *pitch shifters* (Verfaille, 2006). Sin embargo, en todos los casos se mantiene la estructura armónica del timbre original, es decir la posición de los parciales como múltiplos de la frecuencia fundamental. Dicha relación entre parciales puede ser interpretada por medio de la ecuación de la onda en una sola cuerda, ya que la solución para un recinto unidimensional finito indica que los armónicos secundarios aparecen en múltiplos de la frecuencia fundamental (Fletcher, 1998).

En recintos de dos y tres dimensiones la ecuación de la onda tiene otras bases de soluciones que poseen parciales no igualmente espaciados, y por consiguiente distintos timbres, lo cual es el caso de los instrumentos inarmónicos como un tambor o una campana. Siguiendo esta idea se creó una nueva transformación de timbres de instrumentos armónicos a inarmónicos, modulando la frecuencia de los parciales secundarios, hacia posiciones dictadas por las soluciones de la ecuación de la onda para geometrías en más de una dimensión.

La transformación del timbre se puede realizar de dos formas: como un proceso de análisis y resíntesis de las muestras de un timbre real, o como una transformación en tiempo real de la señal proveniente de un instrumento musical. Ambas serán estudiadas aplicando técnicas actuales de procesamiento de señales de audio en las secciones 2.3 y 2.4. La sección 2.5 describe la metodología utilizada y finalmente en las secciones 3 y 3.1 se muestran los resultados y la discusión de los mismos.

2 Modificación de timbres armónicos

En este trabajo, el timbre musical es representado con un modelo determinístico de sinusoidales, más un modelo estocástico de ruido (Serra, 1990). Esta caracterización simple del timbre permite enfocar el estudio en la frecuencia de los armónicos, con el objetivo de realizar las modulaciones necesarias.

2.1 Soluciones de la ecuación de la onda

La ecuación de la onda (1), donde v_p es una constante real positiva que indica la velocidad de propagación de la onda y \mathbf{u} es la solución buscada, puede resolverse mediante el método de separación de variables (Haberman, 2012):

$$\nabla^2 \mathbf{u} = \frac{1}{v_p^2} \frac{\partial^2 \mathbf{u}}{\partial t^2} \quad (1)$$

Para el caso estudiado de recintos finitos, la separación de variables lleva a una ecuación diferencial ordinaria de la forma de (3) para la parte temporal de \mathbf{u} , y un problema en derivadas parciales (2) para la parte espacial.

$$\nabla^2 \mathbf{u} = -\lambda \mathbf{u}, \quad (2)$$

$$-\lambda T(t) = \frac{1}{v_p^2} \frac{d^2 T(t)}{dt^2}, \quad (3)$$

La ecuación (3), para el caso de λ real y positivo, posee una base de soluciones sinusoidales de frecuencia $f = \sqrt{\lambda} v_p$, que dictan la posición de los parciales del timbre estudiado. Es por esto que la determinación del valor de λ es fundamental a la hora de clasificar distintos timbres y, en el caso presentado, de realizar nuevas transformaciones entre timbres.

La resolución de la parte espacial (2) será distinta según el tipo de coordenadas utilizadas y las condiciones de contorno impuestas. Al separar cada una de las componentes espaciales se obtendrán ecuaciones del tipo (4),

$$\frac{d}{dx} \left[p(x) \frac{d\phi(x)}{dx} \right] - q(x)\phi(x) + \lambda \sigma(x)\phi(x) = 0 \quad (4)$$

donde $\phi(x)$ es la incógnita y las funciones $p(x)$, $q(x)$ y $\sigma(x)$ dependerán de la simetría del problema, ya sea cartesiana, cilíndrica, esférica u otras.

La ecuación diferencial ordinaria (4) es un problema de Sturm-Liouville para un intervalo finito y cerrado $[a, b]$ y representa un problema de autovalores, donde las funciones espaciales $\phi(x)$ son las autofunciones y λ los autovalores correspondientes. Éstos dependerán de la simetría del problema y las condiciones de contorno. Si el problema de Sturm-Liouville es regular, los autovalores serán reales y formarán una secuencia infinita $\lambda_1 < \lambda_2 < \dots < \lambda_n$ con un mínimo λ_1 , tal que las autofunciones $\phi_n(x)$ que le corresponden a cada autovalor, forman un conjunto completo para toda función en el intervalo cerrado $[a, b]$.

Para el caso particular de los instrumentos musicales, se utilizan condiciones de frontera nulas (Condiciones de Dirichlet) para modelizar que el material donde se desarrolla la onda permanece fijo en sus extremos. Estas condiciones, tanto para los casos regular y singular, relacionan el valor de los autovalores con los ceros de las funciones $\phi_n(x)$. Por consiguiente, se pueden hallar los conjuntos $\{\phi_n(x), \lambda_n\}$ para distintas simetrías, con el fin de utilizarlos en la modificación de timbres reales.

Por ejemplo, al resolver la ecuación de la onda en un recinto unidimensional, como una cuerda, las funciones $\phi_n(x)$ son sinusoidales y los ceros igualmente espaciados, característicos de los timbres armónicos. En cambio, al utilizar una simetría cilíndrica en coordenadas polares como lo es un tambor, las $\phi_n(x)$ son las Funciones de Bessel $J_n(x)$ y la posición de los parciales viene dada por los ceros de dichas funciones.

2.2 Trasposición de Picos Espectrales

El sonido de un instrumento armónico se desea modificar de modo que se conserven ciertas características de la señal original, en busca de que el nuevo sonido este fuertemente influenciado por el sonido sin procesar. Debido a que la frecuencia fundamental de la muestra procesada dicta la nota musical que fue interpretada, y en la transformación propuesta se busca mantener la misma, la frecuencia de la fundamental y su envolvente de amplitud y fase se mantienen inalteradas. En cambio, los parciales superiores, múltiplos de la fundamental, se modulan hacia la frecuencia deseada más cercana, pero manteniendo su envolvente espectral de amplitud y fase.

Las modificaciones espectrales de una señal de audio pueden ser realizadas tanto en el dominio del tiempo como en el de la frecuencia. Aunque las técnicas en el dominio del tiempo han tenido más éxito debido a su menor costo computacional (Charpentier, 1986), no son útiles

para el fin perseguido, ya que realizan un escalamiento por un factor constante del espectro de la se\u00f1al que no permite modificar cada arm\u00f3nico por separado. De las t\u00e9cnicas utilizadas en el espectro de la frecuencia, la m\u00e1s acorde para el objetivo perseguido es la del *Phase Vocoder* (Flanagan, 1966), que mediante la representaci\u00f3n de la STFT (Short Time Fourier Transform) permite analizar la envolvente espectral de los distintos parciales individualmente.

Los arm\u00f3nicos de la se\u00f1al original forman el conjunto de frecuencias (5) que deben cambiarse por el conjunto de frecuencias deseadas, *normalizadas respecto de la fundamental*, formando el conjunto (6), donde z_n corresponde al n -\u00e9simo parcial.

$$\{f_n = nf_0, 2 \leq n < n_0\} \quad (5)$$

$$\{f_n = \frac{z_n}{z_1}, 2 \leq n < n_1\} \quad (6)$$

La trasposici\u00f3n de frecuencias consiste en asignar los elementos pertenecientes al conjunto (5) a elementos del conjunto (6). Como el valor de n_0 est\u00e1 fijado por la cantidad de parciales presentes en la se\u00f1al a modificar y n_1 es la cantidad arbitraria de parciales a considerar, en general se elige $n_1 < n_0$. La asignaci\u00f3n puede realizarse siguiendo diversos criterios, cada uno de ellos resultando en un sonido distinto, pero respetando la premisa de modificar lo menos posible el timbre original, se modulan las f_n de entrada de modo que se correspondan con la f_n de menor distancia en m\u00f3dulo.

2.3 Sistema en tiempo real

Este nuevo modelo se basa en la modificaci\u00f3n de la se\u00f1al proveniente de un instrumento en tiempo real por medio de un *Phase Vocoder*. Aunque la estructura cl\u00e1sica del *Phase Vocoder* no permite una trasposici\u00f3n de frecuencias en tiempo real debido a su alto costo computacional, la variaci\u00f3n propuesta por (Laroche, 1999) logra realizar la modulaci\u00f3n con un costo considerablemente menor, ya que evita calcular la fase de cada l\u00ednea espectral como en el esquema tradicional. El algoritmo desarrollado en este trabajo utiliza esta idea para lograr la trasposici\u00f3n de los picos espectrales deseada: Se toma la STFT de un bloque de la se\u00f1al de entrada, multiplicado por una ventana de Hann para disminuir la fuga espectral en la mayor medida posible (Harris, 1978), y en cada una de ellas se buscan los m\u00e1ximos locales que corresponden a los distintos parciales. Los m\u00e1ximos locales que se encuentran por debajo de un valor umbral de amplitud m\u00ednima audible por el o\u00eddo humano deben ser filtrados. Cabe aclarar que el valor de amplitud del m\u00e1ximo local debe ser normalizado dividi\u00e9ndolo por el largo de la ventana, previo a la comparaci\u00f3n. El umbral puede optimizarse seg\u00fan el instrumento analizado, tom\u00e1ndose como valor promedio -60 dB (Serra, 1990).

La trasposici\u00f3n de frecuencias se logra copiando el contenido de una cantidad de l\u00edneas espectrales alrededor del m\u00e1ximo hacia la posici\u00f3n de frecuencia deseada, multiplicando cada l\u00ednea espectral por el factor complejo (7) donde Δf es el corrimiento en frecuencia realizado, normalizado respecto de la frecuencia de muestreo de la se\u00f1al f_s . Dicho factor corrige la fase del nuevo parcial, evitando errores de fase que son desagradables al o\u00eddo humano. El espectro de amplitud resultante de intercambiar las l\u00edneas espectrales se puede observar en la figura 1.

$$Z = e^{j2\pi \frac{\Delta f}{f_s}} \quad (7)$$

El largo de la ventana utilizada, que se denominará con la letra W , determina la cantidad de líneas espectrales de la FFT realizada y de esta forma la resolución en frecuencia $f_{bin} = \frac{f_s}{W}$.

Esta resolución limitará la precisión a la hora de encontrar los picos de la señal de entrada, de reasignarlos a las frecuencias deseadas y de elegir cuántos bins alrededor del máximo serán copiados. Este fenómeno se vuelve crítico en notas de frecuencia fundamental relativamente baja, ya que los picos de los parciales se encuentran más cercanos entre sí y la trasposición se realiza con mayor error. Aunque es deseable un W lo suficientemente grande como para tener una resolución espectral alta, no se puede hacer arbitrariamente grande debido a que se perdería resolución temporal, suavizando el efecto del ataque de la señal.

Cabe resaltar que el algoritmo sólo funciona correctamente si se conoce la frecuencia fundamental de la señal analizada, por lo que es necesario identificarla automáticamente en vistas a una aplicación en tiempo real. Dicha tarea se puede realizar de dos formas, en serie o en paralelo, según se estime la frecuencia fundamental como parte del algoritmo de traslación o fuera del mismo. La realización serie consiste en calcular la frecuencia fundamental utilizando los datos espectrales obtenidos por el *Phase Vocoder*, pesando los candidatos a frecuencia fundamental según la relación entre la amplitud espectral de los parciales $F(nf_0)$ y la amplitud espectral de las líneas espectrales entre los parciales $F((n+1/2)f_0)$ (Sun, 2002):

$$peso = \frac{\sum F(nf_0)}{\sum F((n + \frac{1}{2})f_0)} \quad (8)$$

Otras técnicas basadas en el *Phase Vocoder* (Keiler, 2002) refinan la estimación utilizando el cambio de fase entre distintas ventanas. Aunque evitan aumentar la cantidad de cálculo, estos algoritmos tienen una precisión baja, ya que son muy propensos a detectar parciales superiores como frecuencia fundamental. Por esto también se propone la realización en paralelo, que utiliza el algoritmo de búsqueda de la frecuencia fundamental (McLeod, 2005) utilizando la WSNAC (Windowed Special Normalized Autocorrelation), que posee una tasa de error menor a las técnicas que utilizan el *Phase Vocoder*. En esta implementación el algoritmo de estimación de fundamental sintoniza al algoritmo de modificación de timbres para que realice las modulaciones en las líneas espectrales correctas.

2.4 Sistema de análisis y síntesis

Como alternativa al sistema de tiempo real, se puede usar el sistema de análisis y resíntesis, donde se puede realizar la transformación del timbre en dos etapas, una de análisis y otra de síntesis. En este sistema se procesa offline la muestra de una nota de un timbre armónico para estimar el modelo determinístico de sinusoidales más ruido, siguiendo el procedimiento de Serra (1990), basado en el *Phase Vocoder*. De esta forma se obtienen los valores temporales de amplitud y fase de los distintos parciales, según lo muestra la Figura 3. Luego, para la síntesis se conserva la senoidal correspondiente al armónico fundamental y se modifica la frecuencia del resto de los parciales, sin modificar su envolvente de amplitud y sin considerar la fase medida.

Este sistema posee como ventaja que los parciales pueden ser resintetizados con mayor precisión a la frecuencia deseada, ya que no se posee la limitación que impone la cuantización de las líneas espectrales, y puede ser implementado en sintetizadores guardando los datos de las envolventes de amplitud varias notas de la escala.

2.5 Metodología

Si se trabaja con una frecuencia de muestreo de 44,1 kHz, se elige un largo de ventana de 2048 muestras, lo cual fija una resolución temporal de 46,4 ms y una resolución espectral de 21,5 Hz, que es aceptable para notas superiores a los 196 Hz, o su equivalente musical Sol₃. Se toman 2 o 3 líneas espectrales alrededor del pico significando un ancho de banda máximo de 64,5 Hz alrededor de cada pico. De acuerdo a la técnica de *overlapp and add*, se suma cada bloque modificado a la salida utilizando un solapamiento del 75 %, para lograr una alta correlación entre bloques de salida y mejorar el resultado auditivo. Un solapamiento del 50 % reduce el costo computacional, pero la calidad de los resultados disminuye notoriamente. Además, se toman bloques de entrada solapados en el mismo factor en la entrada y la salida, ya que no se busca realizar escalamientos temporales a la señal de entrada.

Para el caso del sistema en tiempo real, antes de realizar la transformación es necesario corroborar que la reasignación de frecuencias de parciales secundarios mantenga una distancia mínima entre parciales mayor a la cantidad de líneas espectrales trasladadas por pico, de lo contrario se superpondrán los picos espectrales y el resultado diferirá del esperado.

3 Resultados

Se evaluaron los dos métodos propuestos con la grabación de un piano acústico tocando la nota Sol₃, que corresponde a una frecuencia fundamental de 196 Hz, utilizando los parámetros mencionados en la sección anterior. En la figura 1 se puede observar la modulación de los parciales para una ventana dada, en el sistema de tiempo real. Nótese que el parcial correspondiente a la frecuencia fundamental es igual para la salida y para la entrada, mientras que el resto de los parciales fueron modulados utilizando los ceros de las Funciones de Bessel. Además, es importante remarcar que la distancia entre los mínimos locales más cercanos de dos picos espectrales modulados vecinos es mayor a una línea espectral, evitando solapamientos indeseados y por ende fuga espectral.

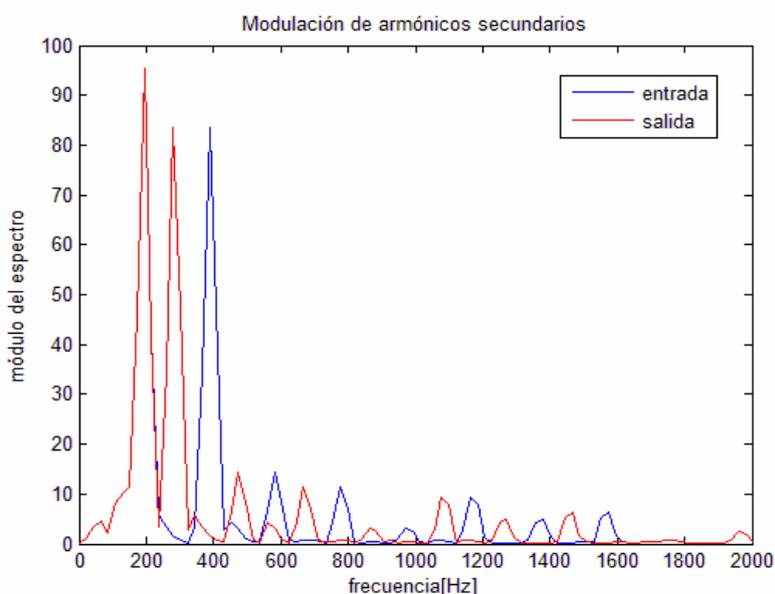


Figura 1. Trasposición de picos espectrales correspondientes a los parciales secundarios

En la figura 2 se observa la envolvente temporal de entrada y salida en el sistema de tiempo real, donde la salida no pierde la característica de ataque de la muestra original, a pesar de que algunos picos en la entrada se ven suavizados por el efecto del *Phase Vocoder*.

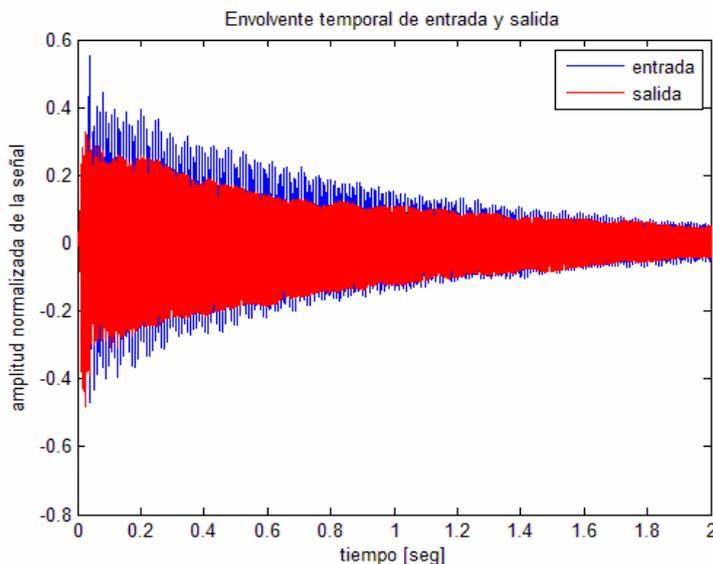


Figura 2. Envolvente temporal de amplitud a la entrada y salida del sistema en tiempo real.

En la figura 3 se muestran los distintos parciales detectados de la señal original utilizando el método de análisis y síntesis. Las frecuencias detectadas a la entrada se encuentran en múltiplos de 196 Hz, mientras que en la salida los parciales detectados son resintetizados con la misma amplitud, pero distinta frecuencia, salvo por el parcial fundamental que no es modificado.

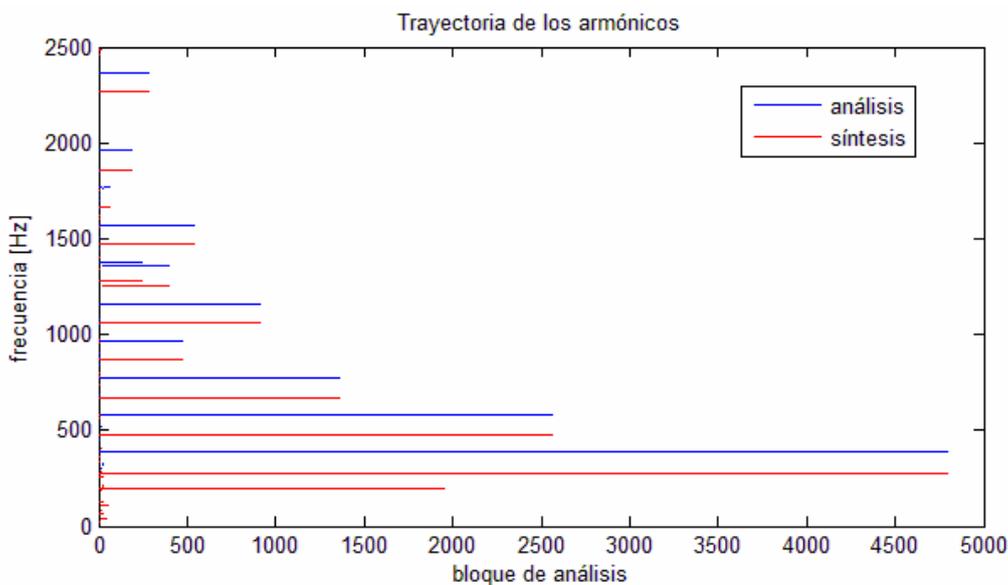


Figura 3. Trayectoria de los parciales analizados y sintetizados, para un piano tocando la nota Sol₃.

3.1 Discusión y trabajo futuro

Se ha presentado una nueva transformación de timbres, incluyendo sus motivaciones teóricas en la ecuación de la onda y se han propuesto dos implementaciones utilizando técnicas de procesamiento de audio, una como una modificación en tiempo real, y la segunda como un sistema de análisis y resíntesis. Los resultados obtenidos en la primera, sugieren que la transformación es viable en tiempo real, permitiendo su implementación como un nuevo efecto digital, mientras que la segunda ofrece mayor sensibilidad en la modificación de las frecuencias, evitando derrame espectral (spectral leakage) entre subarmónicos y es especialmente apta para sintetizadores.

Esta modificación puede también ser utilizada desde un punto de vista psicoacústico, ya que el método propuesto permite evaluar el efecto del corrimiento de los parciales secundarios en el oyente, aplicación con mucha potencialidad en la composición musical.

Aunque la transformación se aplicó sobre muestras de piano y guitarra, también puede emplearse en cualquier instrumento armónico, desde violines hasta contrabajos. Además, las transformaciones posibles son muy diversas, como puede ser cualquier conjunto de autovalores λ_n obtenido mediante la ecuación de la onda, o conjuntos de frecuencias que simulen los parciales de sonidos inarmónicos reales, o cualquier otro conjunto de frecuencias elegidas arbitrariamente. También queda abierta la investigación de la transformación inversa a la presentada, modificar timbres inarmónicos a timbres armónicos.

Debido a que este tipo de transformación no es usual en el procesamiento de señales de audio, se buscaron los métodos actuales que mejor ajustaran a la necesidad planteada, dejando abierta la investigación de técnicas que realicen la misma tarea más eficientemente, como puede ser el caso de utilizar un muestreo no uniforme del espectro para centrar el análisis alrededor de los picos espectrales deseados.

Referencias

- Charpentier, F.; Stella, M.;(1986) "Diphone synthesis using an overlap-add technique for speech waveforms concatenation". Acoustics, Speech, and Signal Processing, IEEE International Conference, ICASSP'86.
- Flanagan, J.L.; Golden R.M; (1966) "Phase Vocoder", Bell Syst Tech. J., vol 45, pp.1493-1509.
- Fletcher, Neville H; Rossing, Thomas D; (1998) "The physics of musical instruments", Springer 2da edición.
- Haberman, Richard; (2012). "Applied partial differential equations with Fourier Series and Boundary Problems". Pearson 5ta edición, Upper Saddle River, Estados Unidos.
- Harris, Frederic J.; (1978) "On the use of windows for harmonic analysis with the discrete Fourier Transform. Proc. of the IEEE, Vol 66(No 1).
- Keiler, Florian; Marchand, Sylvian (2002) "Survey on extraction of sinusoids in stationary sounds" Proc. of the 5th Int. Conference on Digital Audio Effects (DAFx-02), Hamburgo, Alemania.
- Laroche, Jean; Dolson, Mark; (1999) "New Phase-Vocoder Techniques for pitch shifting, harmonizing and other exotic effects", Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp 91-94.
- McLeod, Philip;Wyvill, Geoff; (2005) "A smarter way to find pitch". Proc. of International Computer Music Conference, ICMC.
- Serra, Xavier; Smith, Julius; (1990). "Spectral Modeling synthesis: A sound analysis/synthesis based on a sinusoidal representation,". Computer Music J., vol. 14, no. 4, pp. 12-24.

- Sun, Xuejing; (2002) "Pitch determination and voice quality analysis using subharmonic-to-harmonic ratio" Proc of IEEE International Conference on Acoustics, Speech, and Signal Processing, Orlando, Florida.
- Szego, Gabor; (1955). "Orthogonal Polynomials". American Mathematical Society, 2da edición.
- Verfaille, Vincent; Arfib, Daniel; (2006) "Adaptative Digital Audio Effects (A-DAFx): A New Class of Sound Transformations"
- Zettl, A.;(2005) "Sturm–Liouville Theory", American Mathematical Society.
- Zölzer, Udo; (2002) "DAFX—Digital Audio Effects". New York: Wiley