TECNIACUSTICA
BARCELONA • 1996

Aplicación de técnicas de estimación espectral superresolutivas a subbandas armónicas para la síntesis de sonidos percutivos

Joan Claudi Socoró , Mª Eugènia Santamaría, Elisa Martínez , Xavier Jové

Ingenieria La Salle. Universidad Ramón Llull.
Departamento de Comunicaciones y Teoría de la Señal.
Pº Bonanova, 8. Barcelona 08022, ESPAÑA. E-mail ee03168@els.url.es

INTRODUCCIÓN

Una de las técnicas de síntesis de sonido propuestas durante los últimos años por *Jean Laroche* y *Jean-Louis Meillier* aplicada a la síntesis de sonidos percutivos parece ser un buen método para conseguir una buena compresión de los parámetros de síntesis sin degradar demasiado la calidad de los resultados obtenidos en comparación con las formas de onda originales. Dicha técnica se basa en un modelo digital inspirado en el proceso físico que se produce en la generación de un sonido percutivo como puede ser el del piano, el de una guitarra pinzada o el de un contrabajo. La mayoría de las técnicas actualmente utilizadas en los sintetizadores profesionales recurren a la reproducción de la parte aleatoria del sonido registrada digitalmente debido a la difícil generación de ésta por medio de algoritmos, mientras que la parte armónica es fácil de reproducir por medio de osciladores o de formas de onda básicas. Por otro lado, el almacenamiento de una señal entera requiere una cantidad de memoria apreciablemente alta, y su reproducción directa siempre estará sometida al típico efecto no deseado: la dependencia tonal, tímbrica y temporal. Un aumento de la velocidad de reproducción de las muestras supone un aumento tonal, pero a la vez un timbre más agudo y una reducción de su duración.

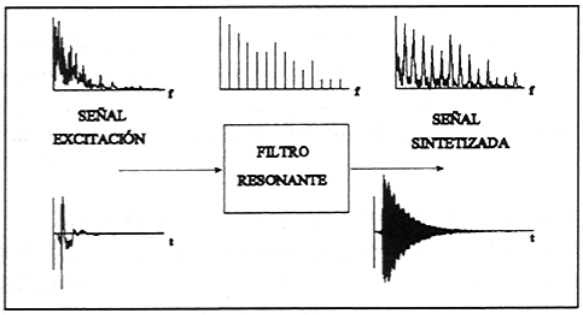

Figura 1: Modelo Filtro-Excitación *para la generación de sonidos percutivos. Arriba los espectros de una marimba, abajo las señales temporales.*

La técnica que se ha utilizado se basa en un modelo tipo excitación-filtro. La excitación pretende reproducir la parte del ataque o de la percusión propia del sonido, mientras que el filtro introduce las resonancias propias de todo sonido armónico con parte percutiva. La bondad del modelo está estrechamente ligada al proceso de estimación de las frecuencias propias de la parte resonante, con lo cual la resolución de dichas frecuencias es un factor muy importante a tener en cuenta.

En éste artículo se presenta una variante de técnicas de análisis espectral superresolutivas para el cálculo de dichas frecuencias y que se puede generalizar a la aplicación global de técnicas de análisis espectral para la detección de sinusoides atenuadas con ruido.

MODELOS FILTRO/EXCITACIÓN Y MÚLTIPLES FILTROS/EXCITACIÓN

El modelo propuesto en [1] se basa en la analogía mecánica de la generación de un sonido percutivo. En primer lugar, la vibración se inicia mediante una transferencia instantánea de energía a un cuerpo resonante, la cual se suministra a través de un excitador (como puede ser el martillo en el piano) que percute sobre éste de forma brusca. Éste excitador se modela mediante una *señal de excitación* de corta duración y de contenido fundamentalmente ruidoso (de espectro relativamente ancho y sin resonancias importantes). Por otro lado el cuerpo resonante posee los modos propios que dan el contenido armónico al sonido y los cuáles son excitados por medio del excitador físico. Dicho cuerpo se asocia a un fil-

tro digital que posea los polos a las frecuencias propias de éste y que estén cerca del circulo unidad, a la entrada del cual se suministra la *señal de excitación*. De este modo, la señal a la salida del filtro contendrá tanto la parte percutiva de corta duración como la parte resonante (con decaimiento exponencial, característico de todo cuerpo libre al cual se le ha suministrado una condición de velocidad inicial), imitando casi perfectamente la naturalidad del sonido (ver Fig.1). Separando de éste modo la parte estocástica de la parte determinística, podemos aprovechar la similitud de la parte percutiva de notas cercanas para generar, a partir de una única excitación, diversas notas de un mismo instrumento con un filtro diferente para cada nota, que es lo que se conoce como el modelo *Múltiples filtros/ Excitación* (ver Fig.2). La excitación común se debe calcular bajo un criterio de máxima similitud de las síntesis de cada nota [1].

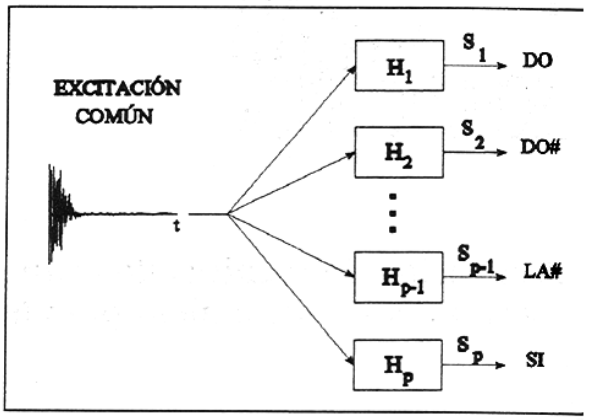

Figura 2: Modelo Múltiples filtros/Excitación.

El filtro resonante que se propone y que parece tener un mejor comportamiento para el objetivo que se persigue es un modelo de secciones de cosenos en paralelo [1]. Fijada la configuración o tipología del filtro, los parámetros que determinan su comportamiento son los polos, que dependen directamente de la estimación de las resonancias en la parte determinística de la señal musical. La *señal de excitación* se calcula mediante la técnica del filtrado inverso, con lo que, si la estimación de las resonancias es suficientemente buena, la señal que se obtiene de filtrar el sonido percutivo mediante el filtro inverso deja de tener contenido armónico y se ajusta bastante bien al modelo propuesto. La elección de la configuración del filtro se justifica en la necesidad de utilizar filtros cuyos inversos tengan un comportamiento bien condicionado, lo cual se traduce en que no existan valles muy profundos en su la función de transferencia [8].

Así pues, el modelo filtro-excitación de varias notas permite calcular una sola excitación y varios filtros que generen la escala musical de todo un instrumento musical percutivo. De este modo, la información total se reduce a una única señal (de menor duración) y a unos cuantos coeficientes (que no dejan de representar una mínima parte de la información necesaria para la síntesis de varias notas de un instrumento percutivo).

ESTIMACIÓN DE LOS POLOS DE SEÑAL

Uno de los factores más importantes es el proceso de estimación de las frecuencias y de los factores de atenuación de las resonancias de la parte determinística del sonido, que configuran la distribución de polos de señal. No hay que olvidar que sumada a la parte determinística tenemos también la influencia de la parte percutiva, la cual se puede entender como una componente ruidosa que afectará a la estimación y tenderá a distorsionarla. Por este motivo, el uso de técnicas de estimación espectral superresolutivas se centra sobretodo en aquellas basadas en la descomposición de valores singulares y autovalores de matrices de señal, con el objetivo de minimizar la influencia del ruido en el cálculo (*Prony* [4] [6], *Kumaresan-Tufts* [7], *Matrix-pencil* [2][5]).

Se demuestra que, a la práctica, cuando se analizan sonidos con más de 50 componentes sinusoidales, o bien cuando la componente de ruido es demasiado importante, es muy difícil obtener una buena estimación de las frecuencias y de los factores de atenuación de todas las componentes presentes en el sonido. Generalmente, siempre hay alguna de ellas (sobretodo en situaciones en que una componente de baja energía se halla entre componentes de elevada energía) que queda sin detectar, lo cual hace imposible el cálculo de la *excitación*. Por otro lado, la influencia del ruido genera desviaciones en los factores de atenuación que pueden incluso hacerlos negativos, lo que traduciría en un filtro inestable y desvirtuaría totalmente el proceso de síntesis. Otro de los motivos por los que la detección total de las componentes presentes es complicada, es la no perfecta adaptación de la señal al modelo *Filtro-Excitación*, lo que se demuestra en componentes que no poseen un decaimiento exponencial, si no que mantienen una tendencia más bien lineal. En estos casos es necesario realizar un análisis concreto de la componente en cuestión para aproximar su evolución decreciente por el modelo exponencial.

Otro gran inconveniente del análisis directo de la señal musical mediante las técnicas superresolutivas antes mencionadas es que, generalmente, al tratar con señales digitales de relativa larga duración, y al tratarse de técnicas de elevado coste computacional, debemos limitarnos a trabajar con una porción reducida de señal,

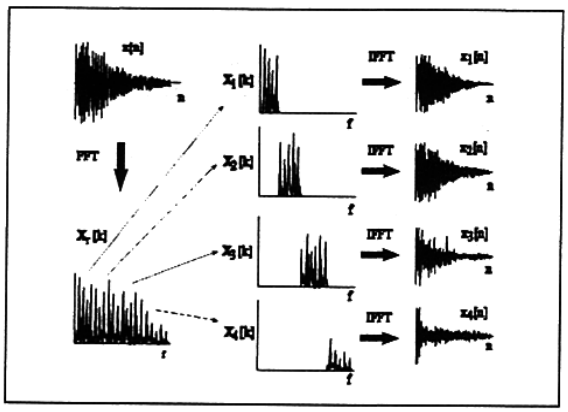

Figura 3: Proceso de descomposición de la señal percutiva en subbandas de frecuencia.

la cual debe contener una mínima componente de ruido y una "buena" representación del decaimiento de cada una de las componentes sinusoidales del sonido. Es por esto que la trama de señal para el proceso de análisis debe ser cuidadosamente escogida y en la mayoría de las veces no se cumple alguno de los requisitos antes mencionados.

Análisis de subbandas de frecuencia

La mejora introducida en el proceso de obtención de los polos de señal se basa en un pre-procesado antes de la aplicación de los algoritmos de estimación espectral. Dicho pre-procesado persigue como principal objetivo conseguir una mejor focalización del análisis a las frecuencias de interés. En primer lugar la señal musical se transforma en frecuencia y a continuación se subdivide en varias subbandas de frecuencia (ver Fig.3), cada una con su correspondiente reflejada. La separación debe hacerse con el criterio de no disociar ninguna de las componentes armónicas en dos, ya que de lo contrario el análisis de dicha componente seria totalmente inactivo. En segundo lugar, se antitransforma cada subbanda y se obtienen las señales temporales asociadas a cada una de ellas. En la Fig.3 se muestra el esquema del pre-procesado donde vemos la partición en cuatro subbandas de frecuencia (donde no se muestran las partes simétricas de cada una). Notar que cada señal contiene unas cuantas resonancias, con lo que la señal temporal asociada a cada subbanda contendrá un numero reducido de componentes. Por otro lado, la longitud de estas señales es menor que la de la señal original, con lo que podemos proceder a un análisis de las componentes de cada una de ellas mediante alguno de los métodos antes mencionados, sin tener que limitarse a un intervalo temporal demasiado reducido. De esta forma la resolución temporal disminuye, sin embargo los coeficientes de atenuación asociados a los módulos de los polos de señal pueden estimarse con mayor precisión, ya que la ventana temporal puede ser ahora mayor. Así, el análisis de señales percutivos con más de 50 componentes es totalmente viable, puesto que podemos escoger el número de bandas adecuado para que cada una contenga menos de este número de componentes.

Posterior al análisis de cada subbanda, es necesaria una etapa de reajuste de los parámetros frecuencia y factor de atenuación de cada sinusoide. Puesto que el proceso explicado implica una expansión del eje frecuencial dentro de cada subbanda, la resolución de las frecuencias es mayor.

Análisis de subbandas armónicas

Una particularización de la técnica anterior para sonidos percutivos con contenido determinístico armónico es la de dividir la señal original en tantas subbandas como componentes armónicas tenga ésta. Cada banda quedará centrada a la frecuencia del armónico en qüestión y tendrá por anchura la frecuencia fundamental. De éste modo se aísla totalmente la información de una componente junto al ruido dentro de la banda, consiguiendo localizar siempre sus parámetros frecuencia y atenuación. Al ser las bandas de anchura constante la escala temporal también lo será para cada subseñal, y la resolución obtenida también.

Otra ventaja que ofrece esta variante es que se pueden detectar componentes sub-armónicas como es el caso de sonidos como el piano, donde el decaimiento de las sinusoides no es exponencial, si no que denotan una cierta sub-modulación, la cual puede modelarse por una contribución de sinusoides muy cercanas a la componente armónica principal, la detección de las cuáles es prácticamente inviable si no es a través de este método.

CONCLUSIONES

Para facilitar el análisis de los polos de señal (elección de parámetros, ventana temporal,...), para la síntesis de sonidos percutivos reales que tengan una componente armónica importante, mediante el modelo *Filtro/Excitación* o *Múltiples filtros/Excitación* [1], se propone la subdivisión de la señal en subbandas armónicas. La aplicación de técnicas de análisis espectral superresolutivas a subbandas mejora el proceso de detección de los parámetros de cada sinusoide, y permite focalizar el estudio evitando distorsionar toda la síntesis por una o dos componentes mal detectadas. No obstante, al ser necesario aplicar el método de estimación para cada subbanda, aumenta inevitablemente el coste computacional. Las pruebas realizadas con sonidos de piano, guitarra, bajo, y marimba, demuestran que la variante propuesta es efectiva, permitiendo calcular, en la mayoría de los casos, una excitación válida para regenerar la señal original. La mejora se hace notar en casos como el del piano, en el que la detección de todas las componentes por medio de un análisis directo de toda la señal es prácticamente imposible, sobretodo de las más débiles. También se

hace posible el detectar subcomponentes, que pueden aumentar considerablemente la naturalidad del sonido sintetizado. No obstante es necesario un aprendizaje para el proceso de elección de los diferentes parámetros a considerar en el análisis, como pueden ser: número de sub-armónicos a detectar por cada componente, ventana de análisis, parámetro de resolución obtenida, etc..

REFERENCIAS

[1] Jean Laroche, Jean-Louis Meillier, *"Multichannel Excitation/Filter Modeling of Percussive Sounds with Application to the Piano"*, Trans. on Speech and Audio Processing, vol. 2, pp. 329-344., Abril 1994.

[2] Y. Hua, T. K. Sarkar, *"Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise"*, IEEE Trans. Acoust. Speech Sig. Process., vol. 38, pp. 814-824, 1990.

[3] J. M. Jot, *"An analysis/synthesis approach to real-time artificial reverberation"*, in Proc. IEEE ICASP-92 (San Francisco), Marzo 1992.

[4] S. Lawrence Marple, Jr., *"Prony's Method"*, Digital Spectral Analysis with Applications, Prentice-Hall, 1987.

[5] Jean Laroche, *"The use of the matrix pencil method for the spectrum analysis of musical signals"*, J. Acoust. Soc. Amer.,vol. 94, pp. 1958-1965, Octubre 1993

[6] Jean Laroche, *"A new analysis/synthesis system of musical signals using Prony's method. Application to heavily damped percussive sounds"*, Proc. IEEE ICASP-89, (Glasgow), Mayo 1989, pp. 2053-2056.

[7] Ramdas Kumaresan, Donald W. Tufts, *"Estimating the Parameters of Exponentially Damped Sinusoids and Pole-Zero Modeling in Noise"*, Trans. Acous. Speech Signal Processing , vol. 30, Diciembre 1982, pp. 833-840.

[8] M. P. Ekstrom, *"A spectral characterization of the ill-conditioning in numerical deconvolution"*, IEEE Trans. Audio Electronics., vol. 21, pp. 344-348, Agosto 1973.