Nuevas bases para el procesamiento de música en el dominio de tiempo-frecuencia

Autor:
Juan Manuel Vuletich
Email: jmvuletich@sinectis.com.ar
L.U.745/92

Directora:
Dra. Ana M. C. Ruedin
Email: anita@dc.uba.ar

Año 2005

Departamento de Computación

Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires

ABSTRACT **3**

RESUMEN **3**

INTRODUCCIÓN **3**

Señal 3
Análisis 4
Síntesis 4
Análisis / Resíntesis 4
Escala musical 4
Localización de señales en tiempo y en frecuencia 5
Teselado del plano de tiempo - frecuencia 6
Objetivo del presente trabajo 6
Herramientas utilizadas 7

TÉCNICAS CONVENCIONALES PARA EL DOMINIO DE TIEMPO - FRECUENCIA **8**

Maneras de representar señales 8
Transformada de Fourier y transformada discreta de Fourier 8
Wigner - Ville distribution 10
Transformada de Fourier con ventana o transformada de Gabor 10
Transformada wavelet continua y su discretización 11
Transformada wavelet discreta diádica 11
Wavelets M-ádicas 13
Bases optimizadas para cada señal 14
"Mosaicos arbitrarios del plano de tiempo - frecuencia usando bases locales" (Bernardini / Kovacevic) 14
Resumen de Técnicas Convencionales 15

UNA NUEVA WAVELET DISCRETA PARA MÚSICA **16**

Mosaico del plano de tiempo - frecuencia 17
Descripción *17*
Parámetros del mosaico *18*
Construcción del mosaico *18*
Funciones elementales 20
Ajuste de los elementos de la base al mosaico del plano *20*
Wavelet *20*
Función de Escala *22*
"Función de Escala Espejada" *23*
Propiedades algebraicas de las bases 25
Correlación entre los elementos *25*
Ortogonalización de los elementos de distintas bandas *26*
Ortogonalización de los elementos de una misma banda *26*
Ortogonalización contra desplazamientos a distancias impares *27*
Aproximación del mosaico por racionales *28*
Ortogonalización contra desplazamientos a distancias pares *28*
Construcción y ortogonalización final de la base 30

RESULTADOS OBTENIDOS **32**

CONCLUSIONES **35**

Trabajos Futuros 35

GLOSARIO **35**

BIBLIOGRAFÍA **36**

Abstract

Conventional techniques for signal analysis and processing in the time-frequency domain are not well adapted to digital processing of music signals. This restricts the features and quality of applications. A novel family of wavelet-like bases allows a tiling of the time-frequency plane that is better adapted to the musical scale. This will allow performance enhancements in all kinds of digital audio applications, for example, pitch detectors, sound identification, musical instruments and effects processors.

Resumen

Las técnicas convencionales para análisis y procesamiento de señales en el dominio de tiempo - frecuencia no se adaptan bien al procesamiento digital de señales de audio, en particular de música. Esto limita las posibilidades y el desempeño de las aplicaciones. Una novedosa familia de pseudo wavelets permite un mosaico del plano de tiempo - frecuencia mejor adaptado a las características de la escala musical. Esto permitirá mejorar el desempeño de toda clase de aplicaciones de audio, pudiendo aplicarse por ejemplo a convertidores de audio a MIDI (pitch detectors), identificadores de sonidos, construcción de instrumentos musicales y, procesadores para estudios de grabación.

Introducción

En el área del procesamiento de señales discretas se destaca el procesamiento de sonidos musicales, por tener la música propiedades particulares, distintas de las de otros tipos de señales (p. ej. imágenes). El procesamiento de música tiene importantes aplicaciones prácticas, que van desde la producción musical (herramientas para músicos e ingenieros de grabación), hasta productos para consumidores (equipos de música, software y aplicaciones para distribución y reproducción de música en computadoras personales).

Dentro del amplio campo del procesamiento de sonido y música, este trabajo se centra en técnicas para transformar las señales, expresándolas en bases con ciertas propiedades deseadas. Nos interesa representar señales musicales de forma tal que cada coeficiente exprese el contenido energético de la señal en un intervalo pequeño de tiempo y de frecuencias. Esto permite modificar la señal, por ejemplo realzando o reduciendo ciertos componentes (p. ej. notas musicales), en cierto instante.

Comencemos definiendo algunos términos, y su significado (en el contexto de este trabajo).

Señal

Este trabajo trata únicamente sobre señales unidimensionales discretas que corresponden a segmentos de música digitalizada. Si las señales a utilizar están muestreadas con frecuencia de muestreo f_s y no contienen energía en frecuencias superiores a $f_s/2$, entonces el teorema de Nyquist (Nyquist 1928, Shannon 1949) garantiza entonces que las señales originales pueden ser reconstruidas perfectamente. Como se verá más adelante, una señal de duración finita no puede cumplir la hipótesis del teorema. Aparece entonces un fenómeno llamado aliasing al reconstruir, pero si f_s es suficientemente alta, el aliasing puede ser ignorado. Esto ocurre en el compact disc y en todos los formatos de audio digital: las grabaciones tienen duración finita.

Análisis

Analizar una señal consiste en aplicar un algoritmo que extrae información en forma de parámetros, que resultan útiles para describirla, o conocer algún aspecto de ella. Por ejemplo, un vúmetro (que muestra la intensidad de la señal a lo largo del tiempo), o un analizador de espectro (que muestra la evolución del contenido frecuencial a lo largo del tiempo).

Síntesis

Sintetizar una señal consiste en generar una señal a partir de ciertos parámetros. Por ejemplo, un sintetizador de música genera sonidos en función de parámetros como la nota a tocar, intensidad, timbre, etc.

Análisis / Resíntesis

Análisis / Resíntesis, o Procesamiento en el Dominio Transformado es aplicar una transformada invertible, para extraer parámetros, operar sobre ellos y sintetizar una nueva señal relacionada con la original (pero probablemente no idéntica).

Ejemplos de esto son la ecualización, el procesamiento de rango dinámico por bandas (como el sistema Dolby), las técnicas usadas por los ingenieros de grabación para mezclar y hacer los ajustes finales para discos (masterización), o la compresión de datos. Trabajar en el espacio transformado permite elegir una base de representación de la señal en la que puedan obtenerse el efecto buscado manipulando sólo unos pocos coeficientes; facilitándose saber cuáles son los coeficientes relevantes, y cómo modificarlos. Expresar la señal en un espacio de este tipo permite algunas de las siguientes aplicaciones:

- Eliminar coeficientes, para eliminar ruido.
- Multiplicarlos por factores, o curvas en el tiempo o en la frecuencia (para realzar o eliminar ciertas componentes).
- Modificar coeficientes (que quizás eran cero) para enriquecer el timbre.
- Reagruparlos o descomponerlos (para separar sonidos y procesarlos por separado).
- Cuantizar y/o eliminar coeficientes, para ahorrar espacio de almacenamiento (compresión de datos).

Escala musical

La escala musical cromática usada por toda la música occidental es descendiente de la escala pitagórica de los antiguos griegos. La construcción de la escala se basa en varias características del sistema auditivo humano que se dan al combinar sonidos de altura definida.

La primera es la forma en que nuestro oído interpreta la diferencia de altura entre dos sonidos escuchados uno a continuación del otro. Por ejemplo, la diferencia de altura que percibimos entre un sonido de 100Hz y uno de 200Hz es la misma que la que percibimos entre un sonido de 400Hz y uno de 800Hz. (Hz es la abreviatura de Hertz, la unidad estándar de frecuencia. Un Hz equivale a un ciclo por segundo.) Esta relación se da entre dos sonidos que tengan uno el doble de frecuencia que el otro, y se la llama octava.

Por otra parte, si dos (o más) sonidos suenan al unísono, decimos que forman un acorde. Si las frecuencias de los sonidos son próximas, o una es próxima a un múltiplo entero de la otra, aparecen los llamados “batidos”, aparentes fluctuaciones en la intensidad del sonido percibido, que se deben a un fenómeno de interferencia entre los dos sonidos originales. Esto ocurre cuando la diferencia de frecuencias está entre 1 y 15Hz. Este efecto lo aprovecha el

afinador de pianos, que afina cada cuerda basándose en una recién afinada, buscando hacer desaparecer los batidos cuando suenan juntas.

Si las frecuencias de los sonidos que forman el acorde son próximas, pero no tanto (su diferencia es mayor a 20Hz), aparece la disonancia. La frecuencia de aparición de los batidos es la diferencia de frecuencia entre los sonidos. Cuando esta diferencia es mayor a 20Hz, su frecuencia de aparición es audible, y los batidos se funden generando un nuevo timbre. Entonces escuchamos una sensación de aspereza. Es fácil sentir la disonancia simplemente presionando varias teclas vecinas de un piano con el puño.

Finalmente, cuando las razones entre las frecuencias de los sonidos del acorde son fracciones con numerador y denominador pequeños, escuchamos una consonancia. Las oscilaciones de los distintos sonidos coinciden cada pocos ciclos, formando una señal periódica. Los sonidos se funden en uno sólo, y el resultado es agradable al oído. Ejemplos de esto son los intervalos más usados en la música: la tercera mayor (5/4), la cuarta perfecta (4/3) y la quinta perfecta (3/2).

El primer punto antes expuesto nos muestra que la progresión de las frecuencias de las notas no es lineal sino logarítmica. Buscando frecuencias útiles para generar consonancias, la "escala de entonación justa mayor" tenía 7 notas por octava, construidas como fracciones simples de una nota fundamental (arbitraria) con la siguiente secuencia: 1, 9/8, 5/4, 4/3, 3/2, 5/3, 15/8, 2. Esta escala evolucionó posteriormente hasta la escala cromática temperada actual, difundida por J. S. Bach en el siglo XVII, y que tiene 12 notas o semitonos por octava, con las siguientes razones entre la frecuencia de cada nota y una nota base arbitraria: 1, $(\sqrt[12]{2})^2$, $(\sqrt[12]{2})^3$, $(\sqrt[12]{2})^4$, $(\sqrt[12]{2})^5$, $(\sqrt[12]{2})^6$, $(\sqrt[12]{2})^7$, $(\sqrt[12]{2})^8$, $(\sqrt[12]{2})^9$, $(\sqrt[12]{2})^{10}$, $(\sqrt[12]{2})^{11}$, 2. Estas aproximaciones irracionales de las fracciones simples generan batidos, pero las aproximaciones son suficientemente buenas como para que el oído medio no los detecte. La razón para abandonar la "entonación justa" (fracciones simples) y reemplazarla por aproximaciones irracionales fue poder construir instrumentos de afinación fija (por ejemplo de teclado, como el clave o el piano) que pudieran tocar en escalas basadas en cualquiera de las notas sin necesidad de introducir notas con frecuencias distintas para cada escala. El afinador de pianos equilibra los batidos en todos los pares de notas, y de esa manera construye la escala temperada. (Para más detalles, se aconseja consultar [OLS/67], [NUÑ/92], o cualquier buen libro de Teoría Musical o Ingeniería de Sonido.)

Localización de señales en tiempo y en frecuencia

Las señales pueden representarse de diversas maneras, considerándolas vectores y utilizando distintas bases de espacios vectoriales. La representación temporal usual usa la base canónica. Otra representación fundamental es la frecuencial, obtenida por medio de la transformada de Fourier, y que utiliza una base formada por senos y cosenos de distintas frecuencias.

El principio de incertidumbre de Heisenberg es un teorema sobre ciertos pares de operadores matemáticos. En mecánica cuántica se aplica a la posición y momento de cualquier partícula. En procesamiento de señales se aplica a la representación temporal y frecuencial de cualquier señal. Ambos son casos particulares de pares de operadores que cumplen las hipótesis del teorema. En mecánica cuántica, significa que no es posible determinar simultáneamente la posición y velocidad (o posición y energía) de una partícula. En procesamiento de señales, el principio de incertidumbre establece una cota a la localización en el tiempo y en la frecuencia de cualquier señal. Esto significa que si se busca concentrar la mayor parte de la energía de la señal en un intervalo lo más reducido posible de tiempo y en un intervalo lo más reducido posible de frecuencia, una mejora en un dominio implica una pérdida en el otro.

Adicionalmente, tener soporte compacto en uno de los dominios (tiempo o frecuencia) implica tener soporte infinito en el otro. O sea que no es posible construir una señal que sea distinta de cero sólo en un intervalo de tiempo, y que su contenido frecuencial sea distinto de cero sólo en un intervalo de frecuencias. Esta propiedad se aplica tanto a señales continuas como discretas. Para más detalles véase [STR/97a], p.67 y p.432.

Teselado del plano de tiempo - frecuencia

Es usual trabajar con el plano de tiempo - frecuencia, que expresa el tiempo en el eje horizontal y las frecuencias en el vertical. Como se trata con señales digitalizadas, el intervalo de frecuencias a considerar es acotado. Normalmente se habla del plano de tiempo - frecuencia, aunque en realidad se trabaja con señales muestradas y con un subconjunto del plano (una banda horizontal que abarca las frecuencias entre 0 y $f_s/2$). Se hace entonces una partición de este plano en pequeñas porciones, todas de igual área, llamada teselado, embaldosado o mosaico (en inglés "tiling"). A cada teselado corresponde una (o más) bases o formas de representar señales. Y a cada baldosa corresponde un coeficiente en esa representación de una señal. En todos los casos, para una cierta tasa de muestreo, la cantidad de baldosas por unidad de tiempo es la misma. Esto permite mantener constante la cantidad de coeficientes a utilizar para representar las señales, requisito para la existencia de bases. Cuando se expresa una señal en una base cuyos elementos se adaptan a un mosaico del plano, se dice que se la expresa en el dominio de tiempo - frecuencia. Muchas formas usuales de representar las señales pueden verse como mosaicos del plano de tiempo - frecuencia. Algunas de éstas son la representación temporal, la transformada discreta de Fourier, la transformada de Fourier con ventana (o transformada de Gabor) y la transformada wavelet discreta.

En muchos casos, el plano queda dividido en un conjunto de franjas horizontales, cada una de una baldosa de ancho, y muchas baldosas de largo. Estas franjas se denominan "bandas de análisis".

Como ya se dijo, es teóricamente imposible obtener una función con soporte compacto a la vez en el tiempo y en la frecuencia. Esto significa que los únicos mosaicos del plano, para los cuales es posible obtener bases que se ajusten exactamente a ellos, son aquellos que dividan al plano sólo en tiempo (como la representación temporal) o sólo en frecuencia (transformada de Fourier y del coseno).

En todas las representaciones que pretenden localización temporal y frecuencial, en realidad cada elemento de la base ocupa muchas (de hecho, infinitas) baldosas, y únicamente están centrados en una baldosa en la cual concentran la mayor parte de su energía. Además, al mejorar la concentración de la energía en un dominio se la empeora en el otro. Esto significa que podemos elegir respetar mejor la localización en un dominio, sin modificar el mosaico, pero afectando la localización en el otro dominio.

Objetivo del presente trabajo

En general, las técnicas más modernas y exitosas para representar señales en el dominio de tiempo - frecuencia utilizan wavelets. Por ejemplo, en Procesamiento de Imágenes, las bases wavelets discretas (diádicas) han sido utilizadas con gran éxito en análisis (detección de bordes y objetos, clasificación de texturas, etc.) y en procesamiento (por ejemplo en compresión, superando a la DCT, o transformada discreta del coseno utilizada por el estándar JPEG).

Sin embargo, en señales musicales, y de sonido; las bases construidas con wavelets discretas no han tenido el mismo éxito. Para análisis (donde se estudian características de la señal, pero no se intenta reconstruirla), la transformada wavelet continua discretizada da los mejores resultados. Sin embargo, para el procesamiento, las wavelets discretas no han resultado una mejora frente a técnicas más antiguas. Por ejemplo, la técnica más exitosa de compresión de audio (el formato MP3) utiliza la MDCT, o transformada discreta del coseno modificada; y la mayoría de las aplicaciones de producción de música usan variaciones de la STFT. Los intentos por utilizar wavelets discretas en lugar de ellas no han rendido frutos. Se cree que esto ocurre porque el ancho de cada banda de análisis no puede ajustarse apropiadamente a las características de la música (v.g. las notas musicales).

El objetivo de este trabajo es presentar bases pseudo wavelet del espacio de señales muestreadas $\ell^2(Z)$ que resultan de utilidad para el tratamiento de sonidos musicales. Los coeficientes para representar una señal en estas bases estarán localizados en el tiempo y en la frecuencia. Esto significa que si se reconstruye la señal, modificando previamente el valor de un coeficiente, entonces los efectos de esta modificación estarán limitados a un intervalo

temporal corto, y sólo a un rango de frecuencias. Se presenta entonces una transformada invertible al plano de tiempo - frecuencia, que se adapta a un mosaico del plano diseñado a partir de la escala musical.

Con relación a las restricciones de las wavelets discretas diádicas convencionales, en [TOR/99], pág. 22, Torresani dice: "The connection between continuous and discrete wavelet systems is not completely understood. ... The multirresolution approach seems to be also extremely constrained by algebraic arguments, which should be developed further." ("La conexión entre los sistemas de wavelets discretos y continuos no es comprendida totalmente. ... El enfoque multirresolución parece además estar extremadamente restringido por los argumentos algebraicos, que deben ser desarrollados aún más."). Y en [DAU/92], pág. 16, Daubechies dice: "Although the constructive method for orthonormal wavelet bases, called multirresolution analysis, can work only if a_0 is rational, it is an open question whether there exist orthonormal wavelet bases (necessarily not associated with a multirresolution analysis), with good time-frequency localization, and with irrational a_0." ("A pesar de que el método constructivo para bases wavelet ortonormales, llamado análisis multirresolución, sólo puede funcionar si a_0 (la razón entre el ancho de 2 bandas vecinas) es racional, es una pregunta abierta si existen bases wavelet ortonormales (necesariamente no asociadas con un análisis multirresolución), con buena localización temporal y frecuencial, y con a_0 irracional."). Resulta asimismo notable que en la portada de [STR/97a] aparece un pentagrama musical con varias notas, como metáfora de la wavelet diádica. Pero de las 12 notas posibles, sólo aparece la nota Do en 5 octavas distintas. O sea, notas cuya frecuencia es $2^i f_0$ para algún f_0 con i entero. ¡Esto sólo ya sugiere la necesidad imperiosa de generalizar la técnica para poder representar también las otras notas de la escala!

Estando de acuerdo con estos autores en la necesidad de nuevos enfoques más generales, este trabajo no usa el método clásico para construir wavelets discretas (el llamado "análisis multirresolución", o MRA), sino que se trabaja directamente con bases explícitas representadas como vectores columna en matrices, buscando nuevas bases de $\ell^2(Z)$, con las características deseadas (buena localización frecuencial, y $a_0 = \sqrt[12]{2}$). El costo es un consumo mayor de memoria y procesador en las computadoras (por almacenar la matriz completa, y resolver el sistema de ecuaciones lineales asociado), pero la ventaja es una mayor libertad para buscar las bases deseadas. Como resultado se construyen bases que no son estrictamente wavelets (o sea que no son traslaciones y dilataciones de una única función básica). El resultado de este trabajo es una nueva familia de pseudo wavelets discretas que prometen ser más apropiadas para aplicaciones musicales.

Herramientas utilizadas

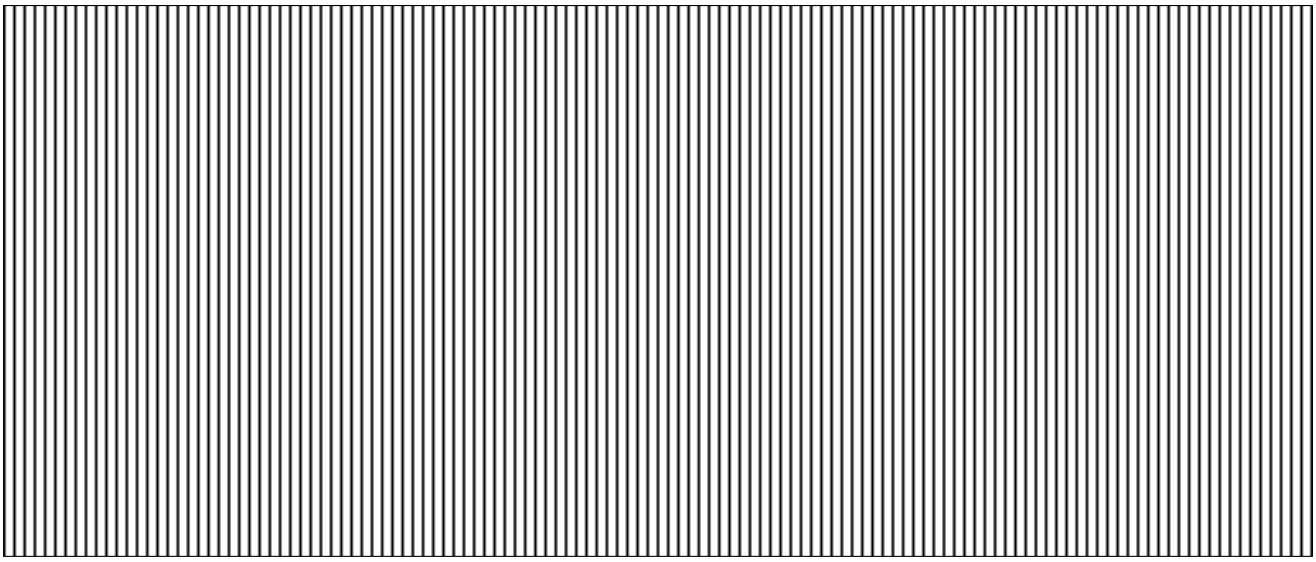
Para llevar a cabo este trabajo, se utilizó el ambiente de programación Squeak Smalltalk (www.squeak.org). Es un ambiente orientado a objetos puro, extensible y de fuente abierto (open source). Se implementó un conjunto de clases para tratamiento de matrices y bases de R^N. También se implementó el soporte necesario para graficar funciones y vectores, y estudiar sus propiedades.

Técnicas convencionales para el dominio de tiempo - frecuencia

Maneras de representar señales

Como ya se dijo, las señales a representar son sucesiones a_k con $\sum a_k^2$ finito, y el espacio al cual pertenecen se denomina $\ell^2(Z)$. Las diferentes formas en que se pueden representar las señales son las distintas bases que el espacio $\ell^2(Z)$ admite. De las técnicas convencionales para análisis y procesamiento de señales en el dominio de tiempo - frecuencia, aquellas que son transformadas invertibles corresponden a bases de este espacio, y proporcionan distintos mosaicos o embaldosados del plano de tiempo - frecuencia.

La primera base a considerar es la base canónica. Los elementos de esta base son vectores de dimensión infinita con un único elemento distinto de cero, que tiene valor uno. Las señales musicales se expresan normalmente en esta base, por ejemplo en un disco compacto de audio, o en formatos de archivos de audio simples, como el wav y el aiff. La base canónica puede verse como un posible mosaico del plano:



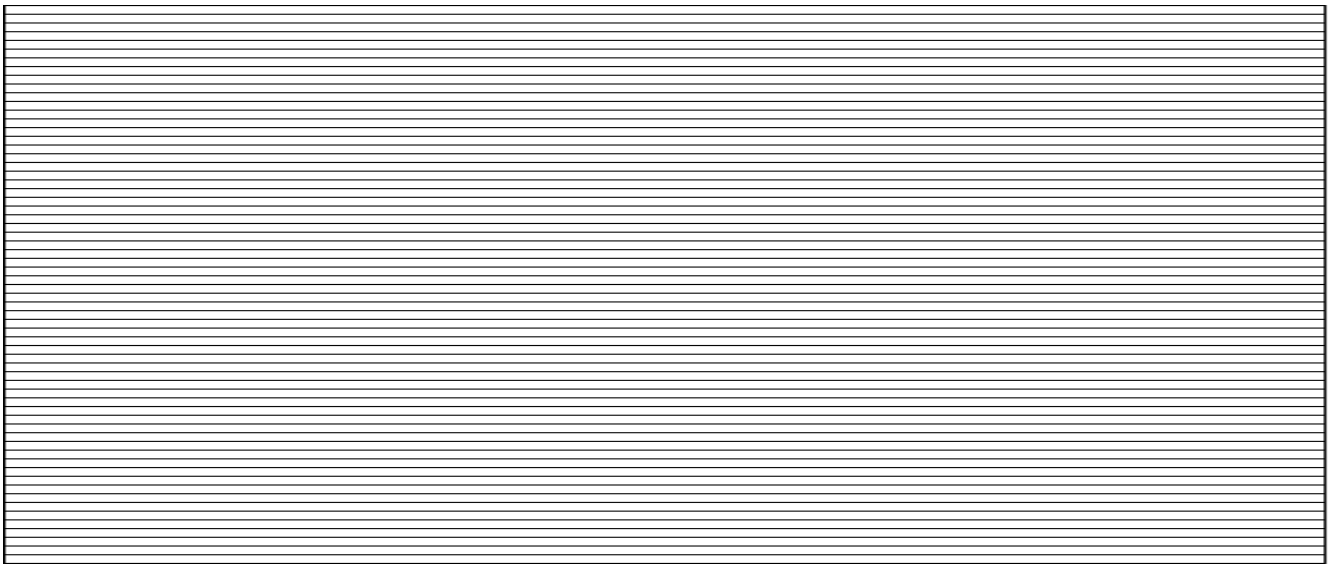
En esta figura (y las 3 siguientes) el eje X corresponde al tiempo y el eje Y a la frecuencia. Consideremos una señal de N = 128 muestras (reales). Se muestra un intervalo temporal de N muestras, y frecuencias desde cero hasta la máxima representable (que es $f_s/2$ donde f_s es la frecuencia de muestreo utilizada). El plano de tiempo - frecuencia se divide en franjas verticales. Cada elemento de la base abarca todas las frecuencias representables, y un intervalo temporal mínimo. La resolución temporal es máxima (y es de 1 / f_s), la resolución frecuencial es nula.

A continuación se describen algunas de las técnicas convencionales, mostrando que las que resultan útiles para el análisis no son buenas para el procesamiento, y las que son aplicables al procesamiento presentan otros inconvenientes. Se evitará una descripción excesivamente detallada de cada una, por haber sido todas ellas descriptas en la literatura en forma exhaustiva.

Transformada de Fourier y transformada discreta de Fourier

La transformada de Fourier (FT) es la técnica de representación frecuencial más antigua, y tiene algunos problemas importantes, pero también tiene virtudes que algunas de las que aparecieron posteriormente perdieron.

Es una transformada invertible. Su versión discreta (DFT) considera una señal de duración finita muestreada. Esto implica que las frecuencias están discretizadas y que el intervalo de frecuencias posibles es acotado. La DFT genera una nueva representación de la señal que utiliza la misma cantidad de coeficientes que la original, y es no redundante. Por lo tanto genera una base del espacio de las señales representadas. Como separa las frecuencias, genera un mosaico del plano de tiempo - frecuencia. Está descripta, por ejemplo, en [ESP/02].



Ahora tenemos N / 2 = 64 franjas horizontales, todas de igual ancho, y tan largas como la señal. Esto significa que tenemos resolución frecuencial máxima (2 f_s / N) igual en todas las bandas, y resolución temporal nula. Representamos entonces la señal con la mitad de los coeficientes (N / 2), pero son coeficientes complejos, y la cantidad de información es la misma.

Es bien conocido que el principal inconveniente de la transformada de Fourier es la completa falta de localización temporal de las componentes obtenidas. La resolución frecuencial es la máxima posible, pero al suponerse que la señal temporal es periódica, se considera también que todas las componentes están presentes en todo momento. Sobre este problema, J. Ville (quien propuso el uso de la Wigner – Ville Distribution como una "densidad temporal - frecuencial") dijo (ver [TOR/99], pág. 1):

"...the representation is mathematically correct because the phases of the tones close to A have managed to suppress it by interference phenomena before it is heard, and to enforce it, again by interference, when it is heard... However this is a deformation of reality: when the A is not heard, it is simply because it has not been played yet..."

("La representación es matemáticamente correcta porque las fases de los tonos cercanos a A (una nota que suena en un determinado momento) tienen éxito en suprimirlo mediante fenómenos de interferencia antes de que se lo escuche, y reforzarlo, de nuevo por interferencias, cuando se lo escucha... Si embargo, esto es una deformación de la realidad: Cuando la nota A no se escucha, es simplemente porque todavía no fue tocada...".)

Esto significa que aparecen componentes espúreas que cancelan parcialmente a otras componentes, pero en realidad ni ellas ni la componente a cancelar deberían aparecer. Este tipo de fenómenos entorpece el análisis, porque sugiere la existencia de características en la señal que en realidad no existen.

La aplicación de la DFT presenta en la práctica un inconveniente adicional, consecuencia de suponer que la señal es estacionaria (periódica), y que el segmento transformado corresponde a una cantidad exacta (entera) de períodos de la misma. Si esta suposición es falsa (en la práctica casi siempre lo es), el resultado obtenido es la transformada de una señal diferente: una señal periódica obtenida concatenando una cantidad infinita de

repeticiones del segmento transformado. Este problema aparece porque muchas implementaciones de la FFT (transformada rápida de Fourier, un algoritmo que calcula la DFT) requieren que la señal tenga tamaño 2^n para algún n entero, y lo que se hace es recortar la señal (tomando una cantidad no entera de períodos) o completarla con ceros. También ocurre porque muchas aplicaciones parten la señal de entrada en segmentos de un tamaño arbitrario (para procesar los segmentos), sin hacer ninguna consideración de la señal en particular que se está procesando en cada momento. De cualquier manera, con elegir mejor el tamaño del segmento a transformar no es suficiente: la señal de entrada normalmente no es periódica. Es en estos casos donde una representación puramente frecuencial no resulta apropiada. El contenido frecuencial de la señal va cambiando con el tiempo. Es precisa una representación en tiempo – frecuencia. Estos problemas, muchas veces ignorados en la práctica, pueden llevar a obtener resultados completamente erróneos. Para más detalles, véase [BRI/88], p.98 a 107.

Wigner - Ville distribution

La Wigner - Ville distribution (WVD) data de 1948 y es históricamente es la primera técnica que busca obtener información sobre una señal, consiguiendo simultáneamente localización temporal y frecuencial. Fue propuesta por J. Ville como una "densidad temporal frecuencial". Cuando se la aplica a oscilaciones puras, proporciona una localización óptima. El buen comportamiento se mantiene también a señales que son ciertas transformaciones simples de una única oscilación pura (p. ej. chirps lineales). Pero aparecen problemas al analizar señales más complejas, por ejemplo la suma de señales simples. En éstos casos el resultado no es la suma de las WVD de aquellas, sino que aparecen "términos de interferencia". En definitiva, la WVD no es lineal. En [TOR/99] hay una descripción más detallada, y se muestran ejemplos donde la alinealidad queda en evidencia.

Existen versiones continuas y discretizaciones (necesarias para trabajar en una computadora) que sirven para el análisis (con diferentes particularidades en cada caso), pero no permiten la reconstrucción de la señal.

En [NEW/97] se hace una comparación entre la WVD discretizada, la STFT y la CWT discretizada aplicándolas al análisis; concluyéndose en las ventajas de la CWT sobre los otros métodos.

Transformada de Fourier con ventana o transformada de Gabor

La transformada continua de Gabor o transformada de Fourier con ventana (STFT), y su versión discreta, son técnicas para dotar de localización temporal a la transformada de Fourier. Están descritas en [TOR/99], [DAU/92] y [ESP/02].

La transformada continua es una transformada invertible entre $L^2(R)$ y $L^2(R^2)$. En la práctica se usan siempre versiones discretizadas.

Se usa una única ventana temporal para el análisis en todas las bandas. Esto significa que la resolución temporal y la resolución frecuencial son constantes, y no pueden ajustarse para crear bandas de distinto ancho. El resultado de esto es un mosaico del plano de tiempo - frecuencia con todas las baldosas rectangulares e idénticas, formando una especie de cuadriculado. La STFT es una representación redundante, pero existen representaciones que comparten muchas de sus propiedades, y que son bases. Una de las más usadas es la MDCT o transformada del coseno modificada, utilizada por el método de compresión de audio MP3.

Estas representaciones presentan otros inconvenientes. Dentro de una ventana se dan problemas similares a los de la DFT, porque se considera que (dentro de la ventana) la señal es estacionaria (periódica). Esto significa que si aparecen componentes cuya frecuencia no es múltiplo del tamaño de la ventana, aparecen componentes espúreas en el resultado. Este problema se da muchas veces al usar la DFT y ya fue comentado. Las distintas alternativas para la ventana intentan paliar este problema (con cierto éxito). También aparecen problemas si en la búsqueda de una resolución frecuencial buena, apropiada para identificar frecuencias bajas (sonidos graves), se toman ventanas temporales grandes. Al utilizarse la misma ventana para las frecuencias altas, es posible que sonidos muy cortos no lleguen a identificarse apropiadamente, agravándose los problemas que comparte con la transformada de Fourier.

En [NEW/97] se hace una comparación entre la WVD discretizada, la STFT y la CWT discretizada aplicándolas al análisis; concluyéndose en las ventajas de la CWT sobre los otros métodos.

Transformada wavelet continua y su discretización

La transformada wavelet continua (CWT) fue propuesta por Grossman y Morlet como alternativa al a transformada continua de Gabor, y está descrita en [TOR/99]. Es una transformada continua al dominio de tiempo - frecuencia. Se toma una única función, llamada wavelet, que contiene la mayor parte de su energía localizada en un intervalo temporal y en un intervalo frecuencial. La primera elección es la llamada wavelet Morlet, una gaussiana modulada, por su óptima localización en el tiempo y en la frecuencia. En [DAU/92], p.76 hay una descripción, en el contexto de las discretizaciones redundantes llamadas "frames" (que no forman bases) de la transformada continua. Esta wavelet es trasladada en el tiempo y simultáneamente dilatada (o sea trasladada en la frecuencia). Cada escala o dilatación modifica la frecuencia en la que está centrado el espectro, y entonces cada dilatación está asociada a una frecuencia "central". La transformada mide para cada instante y cada frecuencia (y su dilatación correspondiente) la correlación entre la señal original y la wavelet trasladada y dilatada.

En las implementaciones para computadoras, se discretiza tiempo y escala. Cuán fino debe ser el muestreo en cada eje depende de la aplicación, pero también debe ajustarse cuidadosamente a las características de localización temporal y frecuencial de la wavelet seleccionada. Se construye así un mapa de la distribución de la energía de la señal en el tiempo y en la frecuencia, con muy buena resolución, acercándose al principio de incertidumbre de Heisenberg. En [NEW/97] se describen ciertas wavelets desarrolladas para el análisis usando la CWT, llamadas Wavelets Armónicas (Harmonic Wavelets).

Eligiendo adecuadamente la wavelet, es posible ajustar la representación a la escala musical. De cualquier manera, estas técnicas son útiles para el análisis pero no para el procesamiento, porque las discretizaciones entregan representaciones muy redundantes de la señal. Si bien esto no impide la reconstrucción, dificulta el procesamiento en el dominio de tiempo – frecuencia, porque al haber redundancia, distintos conjuntos de coeficientes generarán la misma señal, y no resultará fácil determinar qué manipulación de los coeficientes es la que hay que hacer para obtener un resultado en particular. Por ejemplo, un muestreo excesivamente fino del plano de tiempo – frecuencia sugiere que sería posible reconstruir una señal afectando un intervalo temporal y frecuencial extremadamente pequeño, menor que el límite impuesto por el principio de incertidumbre. Es claro que los resultados decepcionarán al usuario: la magia no existe. Esta técnica está descripta en [ESP/02] y [NEW/97]. En [NEW/97] se hace una comparación entre la WV discretizada, la STFT y la CWT discretizada aplicándolas al análisis; concluyéndose en las ventajas de la CWT sobre los otros métodos. En [OLM/99] se detalla una aplicación al análisis de música y reconocimiento automático de melodías. En este trabajo se introduce una familia de wavelets continuas especialmente adaptadas al análisis de música llamada wavelet Log-Morlet.

Transformada wavelet discreta diádica

Las transformadas wavelet discretas usuales (DWT diádicas, o DDWT) no son una discretización de las CWT.

Se utilizan dos funciones fundamentales continuas: la función de escala Φ_0 y la wavelet Ψ_0. Como primer paso, se toma una cierta función de escala Φ_0. La propiedad principal de esta función es que desplazándola por múltiplos enteros de cierto Δt forman una base ortogonal de ciertos subespacios de $L^2(R)$. En particular, pueden representarse funciones constantes, lineales, y en algunos casos cuadráticas y polinómicas de grados mayores. Llamemos a este espacio V_0. Consideremos ahora la función Φ_{-1}, que es Φ_0 dilatada por un factor 2: $\Phi_{-1}(x) = \Phi_0(x/2)$. Los desplazamientos de Φ_{-1} forman una base de un espacio llamado V_{-1}. Ahora entra en escena Ψ_{-1}. $\Phi_{-1} \cap \Psi_{-1}$, y sus desplazamientos por múltiplos enteros de 2 Δt forman una base de un espacio llamado V_{-1}. Por la forma en que se construyen Φ_{-1} y Ψ_{-1} (para ser utilizadas juntas) resulta que $V_0 = W_{-1} \oplus V_{-1}$.

Esto puede repetirse, obteniendo Φ_{-2} y Ψ_{-2} tales que $V_{-1} = W_{-2} \oplus V_{-2}$. O sea, $V_0 = W_{-1} \oplus (V_{-2} \oplus W_{-2})$. Repitiendo esto k veces, tenemos $V_0 = W_{-1} \oplus (\ldots \oplus (V_{-k} \oplus W_{-k}))$.

El cálculo de la transformada consiste en partir de una secuencia de coeficientes que corresponden a la función contínua original, pero expresada en la base formada por desplazamientos de Φ_0. Entonces, por convolución con dos filtros (asociados a Φ_0 y Ψ_0) y submuestreo, se expresa la señal en la base formada por Φ_{-1} y Ψ_{-1}. Esto se repite para los coeficientes de Φ_{-1}, para expresar la señal en la base formada por Ψ_{-1}, Ψ_{-2} y Φ_{-2}. Esto se repite k veces, y la señal queda expresada en la base formada por Ψ_{-1}, ... Ψ_{-k} y Φ_{-k}. La antitransformada es revertir todos los pasos usando dos filtros apropiados, y sobremuestreo intercalando ceros.

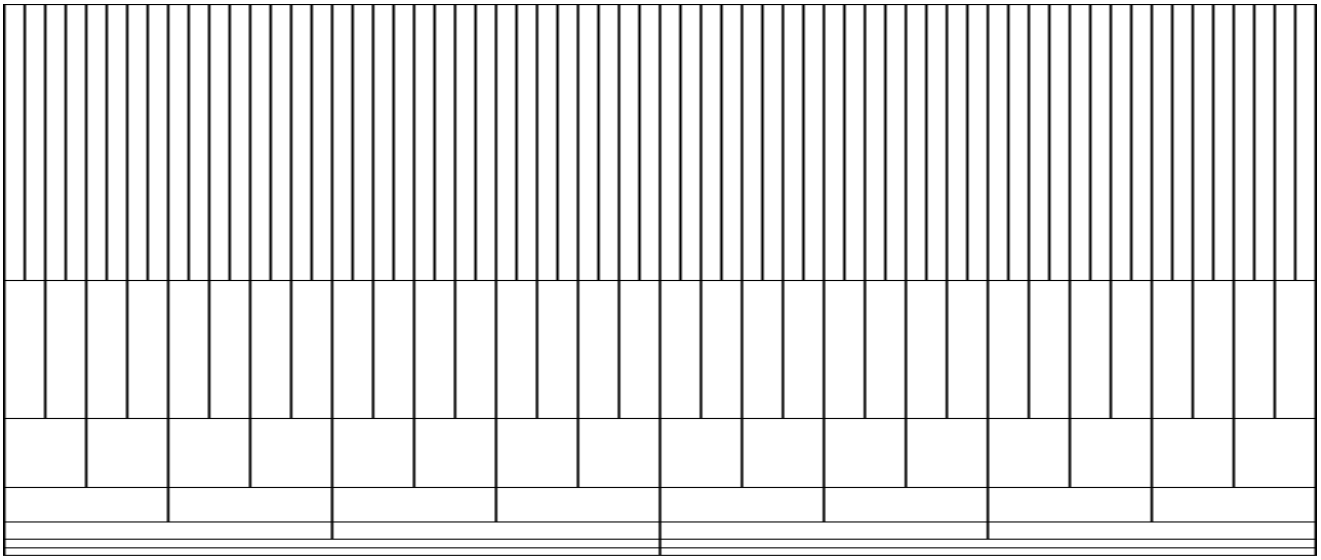
Para aplicar esta transformada a señales discretas en vez de funciones continuas, la secuencia de coeficientes inicial es directamente la señal a transformar. Esto debe tenerse en cuenta el elegir la wavelet a utilizar.

Estas técnicas evitan el costo de almacenar las bases de los distintos espacios; y son muy eficientes en el cálculo: si la cantidad de bandas k es constante, tenemos O(n) con n el tamaño de la señal.

Si las propiedades de localización temporal y frecuencial de Φ_0 y Ψ_0 son apropiadas, el resultado es un mosaico del plano como el que se muestra en la figura, donde las bandas de frecuencias más altas tienen mejor resolución temporal y aquellas correspondientes a las frecuencias más bajas tienen mejor resolución frecuencial.

Los elementos de la base y los coeficientes son reales. Para representar una señal de N muestras temporales se usan N coeficientes.

Como ya dijimos, las representaciones son no redundantes y generan bases. Son entonces transformadas invertibles. La localización temporal y frecuencial depende de la wavelet utilizada, pero en ningún caso será mejor que una octava. Esto significa que cada banda abarca al menos 12 notas de la escala.



En esta figura se muestra el mosaico correspondiente a una base construida con 6 bandas o dilataciones de la wavelet, y la función de escala. Tenemos de nuevo 64 baldosas de igual área. La resolución frecuencial es de una octava: cada banda l de frecuencias va desde cierta $f_{\inf}(l)$ hasta $f_{\sup}(l) = 2 f_{\inf}(l)$ y abarca un ancho ("bandwidth") $bw(l) = f_{\inf}(l)$, la resolución temporal (y la longitud de cada baldosa) es $long(l) = 1/2bw(l)$, y

mantiene la cantidad de baldosas y las áreas iguales a las de la base canónica. La resolución temporal mejora para frecuencias altas, y empeora para frecuencias bajas. Esto resulta deseable para procesar sonido, pero el inconveniente es que nuestro oído tiene una resolución frecuencial mucho mejor que una octava (y una resolución temporal más pobre, porque nuestro oído también debe respetar el principio de incertidumbre de Heisenberg).

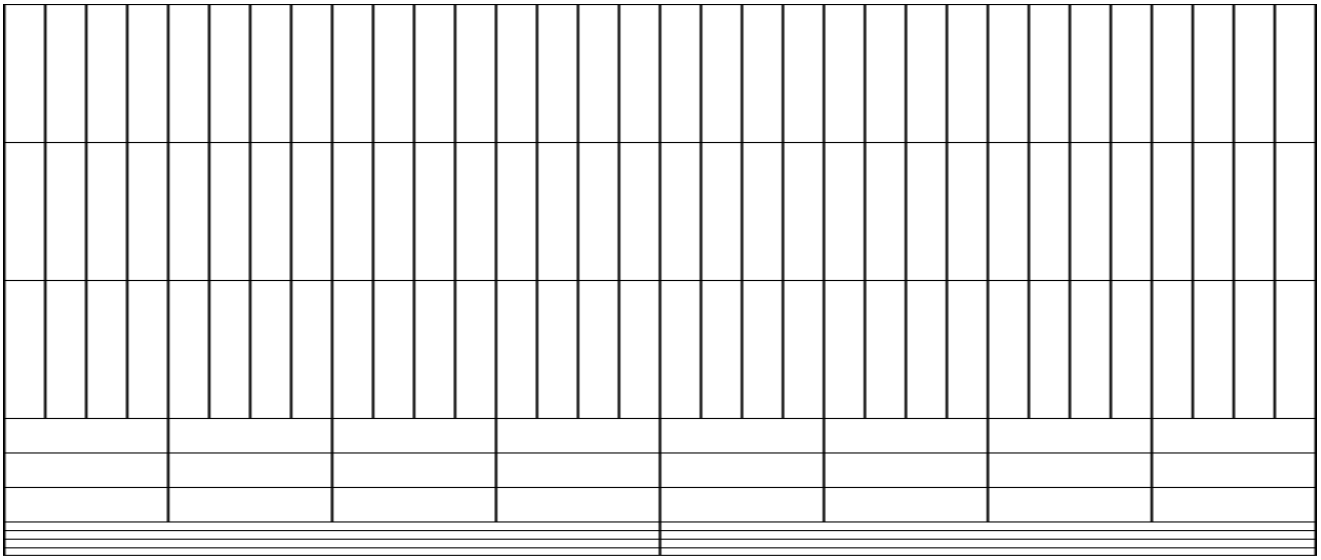
Adicionalmente debe tenerse en cuenta que cambiar la elección de la wavelet afecta la localización temporal y frecuencial de los elementos, modificando la forma en que se ajustan al mosaico del plano; pero el mosaico es siempre el mismo. Esto ocurre incluso con las wavelets discretas de Daubechies, que tienen soporte frecuencial muy amplio, y soporte temporal compacto pero siempre mayor que una baldosa. En aplicaciones en las que interesa especialmente el teselado del plano es importante utilizar wavelets que se ajusten a cada baldosa lo mejor posible.

Esta técnica está descripta en detalle en [DAU/92], [STR/97a], [STR/97b], [TOR/99] y [ESP/02].

Wavelets M-ádicas

Las Wavelets M-ádicas son una generalización de la DWT diádica, y permiten mayor flexibilidad en el balance entre resolución temporal y frecuencial. Generan mosaicos del plano.

Las wavelets M-ádicas buscan incrementar la resolución frecuencial a costa de la resolución temporal. Esto es deseable, pero la manera en que lo hacen es en cada iteración del algoritmo dividir el espectro disponible en m bandas de igual ancho. Luego, se repite el procedimiento sobre la banda inferior recién generada. Esto se repite sucesivamente tantas veces como se desee (y permita la longitud de la señal). Las traslaciones y submuestreos aplicados son siempre enteros. Por esta razón no es posible obtener el balance entre resolución temporal y frecuencial necesario, por ejemplo, para analizar los tonos de una señal musical.



El ejemplo muestra el mosaico correspondiente a m = 4, iterando el algoritmo 3 veces. Tenemos otra vez 64 baldosas de igual área que en la canónica. Tenemos bandas vecinas de igual ancho, y también vemos bandas vecinas donde una tiene el un ancho m veces mayor que la otra. Nuestro oído no funciona de esta manera.

Al igual que en las DWT diádicas, la elección de la wavelet afecta la localización temporal y frecuencial de los elementos de la base, y su ajuste al mosaico, sin modificar el mosaico en sí.

Bases optimizadas para cada señal

Existen diversos trabajos publicados e investigación en curso sobre el problema de elegir una base especialmente adaptada a la señal a representar. El objetivo buscado es elegir una base que permita minimizar la cantidad de coeficientes necesarios para representar cierta señal. La base se arma eligiendo elementos de un diccionario de elementos.

Entre las técnicas de este tipo podemos citar "Matching Pursuit" (S. Mallat y Z. Zhang), "Best Basis" (R. Coifman y V. Wickerhauser) y "Basis Pursuit" (D. Donoho). Están descriptas someramente en [STR/97a], pág. 85.

Estas técnicas generan bases. Son entonces transformadas invertibles. Sus principales aplicaciones incluyen el análisis de señales, y la compresión; pero no el procesamiento en general. La razón de esto es que las bases de este tipo carecen de "ecuanimidad" y favorecen la reconstrucción de ciertas señales (las que se ajustan mejor a la base). Entonces se condiciona el tipo de operaciones que resulta más fácil aplicar. Por ejemplo, de usarse para construir un ecualizador (que permite ajustar el nivel del sonido en bandas de frecuencia), tendríamos un ecualizador cuyas bandas cambian según la señal y entonces permiten distintos tipos de ecualización para distintos tipos de señales.

"Mosaicos arbitrarios del plano de tiempo - frecuencia usando bases locales" (Bernardini / Kovacevic)

En el artículo con este nombre ("Arbitrary Tilings of the Time-Frequency plane using local bases"), ([BER/99]) Bernardini y Kovacevic desarrollan una interesante técnica para obtener bases ortogonales que aproximan cualquier mosaico del plano de tiempo - frecuencia. El resultado es realmente novedoso y prometedor. El problema atacado es similar al de este trabajo, pero es aún más general.

En vez de restringirse a mosaicos del plano donde cada banda tenga un ancho de banda relativo constante, como la escala musical; permiten prácticamente cualquier mosaico del plano. La manera de especificar el mosaico deseado es la siguiente: si tenemos una señal de N muestras, sabemos que podemos representarla en una resolución máxima de N instantes (representación temporal) o una resolución máxima de N/2 frecuencias (utilizando coeficientes complejos en el dominio de Fourier, sabiendo que las señales son reales). Dividamos el plano (acotado en tiempo y frecuencia a la señal y al muestreo) en N/2 franjas horizontales y N franjas verticales. Tenemos una grilla con $N^2/2$ elementos, y sabemos que es imposible tener una representación con tanta resolución. Sabemos que cualquier área rectangular formada por N/2 elementos de la grilla se ajustará al principio de incertidumbre (permitiendo utilizar bases y coeficientes reales). Esto es justamente lo que esta técnica nos permite hacer. Podemos especificar cualquier conjunto de N baldosas, cada una formada por N/2 elementos de la grilla, tales que formen una partición del plano (que no haya superposiciones ni huecos), y la forma de cada baldosa resulte rectangular. Vemos entonces que el ancho (intervalo temporal) de cada baldosa es múltiplo del intervalo de muestreo 1 / f_s , y que la altura (intervalo frecuencial) de cada uno es múltiplo de 1/t, donde t es la longitud de la señal en segundos. Con la técnica propuesta se genera automáticamente una base ortogonal que respeta el mosaico pedido en forma aproximada.

El enfoque es completamente distinto al de este trabajo, sin embargo, es válido comparar los resultados obtenidos. Véase los gráficos de respuesta en frecuencia de la página 23 de [BER/99], y compárese con los incluidos más adelante en este trabajo. Posteriormente se incluye también un comentario sobre la comparación.

Resumen de Técnicas Convencionales

El siguiente cuadro presenta en forma resumida las características más relevantes de las técnicas descriptas.

	Proporciona bases	Transformación lineal	Responde bien a manipulación de coeficientes (*)	Proporciona localización temporal	Proporciona localización frecuencial
Transformada de Fourier y transformada del coseno	si	si	no (1)	no	si
Wigner - Ville Distributions	no	no (2)	-	si	si
Transformada de Fourier con ventana y transformada del coseno modificada	si	si	si	mala (3)	mala (4)
Transformada wavelet continua	no	si	-	si (5)	si (5)
Transformada wavelet discreta diádica	si	si	si (5)	regular (6)	regular (6)
Transformadas wavelet M-ádicas	si	si	si (5)	regular	regular
Bases ortonormales optimizadas para cada señal	si	no (7)	no (8)	regular (9)	regular (9)
"Mosaicos arbitrarios del plano de tiempo - frecuencia usando bases locales"	si	si	si	buena	regular(10)

(*) "Responde bien a manipulación de coeficientes" significa si es posible afectar la amplitud de la señal en una baldosa modificando únicamente el coeficiente correspondiente a ella, afectando razonablemente poco a las otras baldosas, y dependiendo razonablemente poco de los otros coeficientes.

(1) No, por falta de localización temporal y por fenómenos de interferencia que aparecen para compensarla.
(2) No, aparecen "Términos de Interferencia"
(3) Mala, una única resolución temporal a todas las bandas de análisis
(4) Mala, una única resolución frecuencial absoluta a todas las bandas de análisis
(5) Si, si la wavelet es está bien localizada en el tiempo y en la frecuencia
(6) Regular, balance entre resolución temporal y frecuencial no ajustable
(7) No, porque al sumar dos señales se obtiene una señal que debe escribirse en una base que no es la base de ninguna de las señales iniciales.
(8) No, porque al reducirse la cantidad de coeficientes utilizada, se hace que cada coeficiente represente a más de una baldosa. Adicionalmente, al no haber una base canónica, no es posible establecer una semántica para los coeficientes que sea independiente de la señal; y esto dificulta elegir el criterio para su manipulación.
(9) Regular, limitado al diccionario utilizado (usualmente diádico)
(10) Regular. Véanse los resultados obtenidos por los autores en [BER/99].

Una nueva wavelet discreta para música

Antes de este trabajo, no se conocían bases con las características deseadas. Por lo tanto, de existir, era necesario encontrarlas. La estrategia elegida no involucra encajes de subespacios de aproximación, como se hace en wavelets diádicas y M-ádicas.

Se supone una longitud de señal N, y se construye un mosaico del plano de tiempo – frecuencia de N baldosas. A continuación se construye una base de R^N, ajustando sus elementos a este mosaico. Los elementos de estas bases son vectores de dimensión N, y se los construye a partir de una discretización de la wavelet Morlet. La wavelet Morlet es una de las más utilizadas en la CWT, por su óptimo balance entre localización temporal y frecuencial. (Ver [DAU/92], p.76. y [STR/97a], p.67.). La wavelet Morlet es una función compleja (R->C). Como este trabajo trata únicamente sobre señales reales, se eligió como primera candidata a una versión real:

$$\Psi(t) = b.\sqrt{\pi}.e^{-(b.\pi.t)^2}.\cos(2\pi.f_q.t)\ .$$

El parámetro b controla el ancho de la gaussiana, y f_q es la frecuencia donde está centrado el espectro. La wavelet está centrada en tiempo t = 0.

Posteriormente se elimina la correlación entre los elementos de la base para hacerla ortogonal. Esto se hace en varias etapas, para mantener las buenas propiedades de localización temporal y frecuencial de la wavelet.

A continuación se describen someramente los contenidos de los títulos que componen esta sección del trabajo.

Mosaico del plano de tiempo – frecuencia

Aquí se describen las características del mosaico construido para reflejar las características de la escala musical. También se detallan los parámetros que determinan un mosaico en particular, y cómo es la construcción del mosaico a partir de estos parámetros.

Funciones elementales

En una transformada wavelet, todos los elementos de utilizados para representar señales se construyen mediante dilatación y traslación de dos funciones elementales, llamadas wavelet y función de escala. En las bases propuestas en este trabajo, existen tres funciones elementales. En este punto se las describe.

Propiedades algebraicas de las bases

Se desea que las bases construidas sean ortogonales, o sea que la correlación entre cualquier par de elementos sea cero. Conseguir esto manteniendo las otras propiedades de la base es un desafío. En esta sección se describe la correlación entre distintos pares de elemtnos, y cómo se la ataca en cada caso.

Construcción y ortogonalización final de la base

En este punto se describen los pasos finales en la construcción de una base en particular.

Mosaico del plano de tiempo - frecuencia

Descripción

Como ya fue expuesto en la Introducción, el objetivo de este trabajo es encontrar bases wavelet (o al menos pseudo wavelet) cuyo mosaico del plano de tiempo - frecuencia se ajuste a la escala musical. En este mosaico cada banda corresponde a un semitono de la escala musical, su ancho de banda ("bandwidth") es $bw(l) = f_{\sup}(l) - f_{\inf}(l)$, y como $f_{\inf}(l) \cdot \sqrt[12]{2} = f_{\sup}(l) = f_{\inf}(l+1)$, su ancho de banda relativo $bw_r = bw(l) / f_{\inf}(l) = \sqrt[12]{2}$ -1. El ancho de banda relativo es el mismo en todas las bandas (los ingenieros suelen hablar de "factor Q constante"), porque la escala musical está adaptada a nuestro oído, que tiene una respuesta logarítmica a la frecuencia. (Ver Introducción.)

(Nota: Todas las frecuencias se expresan tomando como unidad a la frecuencia de muestreo f_s, y todos los tiempos se expresan tomando como unidad al intervalo de muestreo $1/f_s$. De esta manera, no es necesario conocer el valor de f_s para operar, ni utilizar unidades explícitamente.)

Para cada banda l (determinada por su frecuencia central $f_c(l)$), determinemos la longitud temporal $long(l)$ de sus baldosas. En la representación temporal inicial, el ancho frecuencial de las baldosas es ½, y la longitud temporal es 1. El área de cada baldosa es ½. En el mosaico a construir, la cantidad de baldosas por unidad de tiempo debe ser la misma, y entonces el área de las baldosas también será de ½. Entonces, $bw(l) \cdot long(l)$ = ½, y $bw(l) = \frac{1}{2 \cdot bw(l)} = \frac{1}{2 \cdot bw_r \cdot f_{\inf}(l)}$.

En procesamiento de audio normalmente no es de interés trabajar con las frecuencias que resultan demasiado bajas para resultar audibles por sí mismas, y se las suele filtrar (eliminar). Debido a esto (y a semejanza de la DWT), en las bases presentadas en este trabajo existe un punto arbitrario por debajo del cual se deja de analizar, y todas las frecuencias inferiores se incluyen en una banda especial cuya única función es completar el mosaico, a la que se asocia una función discretizada especial, similar a una función de escala.

Como el mosaico debe ajustarse a la escala musical, es necesario que cada banda esté centrada sobre la frecuencia verdadera de la nota que le corresponde. La afinación de los instrumentos musicales se suele hacer fijando la nota La de la octava central en 440Hz, y construyendo toda la escala a partir de allí. Pero si utilizamos el mosaico descrito anteriormente, nuestra banda de menor dilatación (y mayor frecuencia) capturará las frecuencias entre $\frac{1}{2 \cdot \sqrt[12]{2}}$ y $1/2$. Si la tasa de muestro f_s es de 44100 Hz (como en el compact disc de audio), la frecuencia superior de la banda más alta es de 22050 Hz. Al construir las bandas, dividiendo esta frecuencia por bw_r repetidas veces, la banda que resulta contener a 440Hz va desde aprox. 434Hz hasta aprox. 460Hz, quedando centrada aproximadamente en 447Hz. Esto significa que nuestro mosaico no se ajusta a las verdaderas frecuencias de las notas utilizadas, y podría decirse que está desafinado.

Para corregir esto se agrega una banda superior especial, que captura todas las frecuencias superiores a la máxima banda de análisis que nos interesa. Es de alguna manera análoga y complementaria de la función de escala; porque sólo es incluida para completar el mosaico y la base, y es construida especialmente para ésta función. De la misma manera que la función de escala abarca hasta la frecuencia cero, ésta otra debe abarcar hasta la frecuencia máxima representable $f_s/2$.

Parámetros del mosaico

- N: Longitud de la señal a representar (cantidad de muestras)
- f_0 : Frecuencia mínima de análisis. Debe ser mayor que cero y menor que 1/2.
- L: Cantidad de bandas de análisis - 1. Debe ser tal que $f_{\sup}(L) \leq$ ½. (Ver más abajo cómo se calcula $f_{\sup}(L)$.)
- bw_r : Ancho de banda relativo de todas las bandas de análisis. Su valor es fijo, $bw_r = \frac{f_{\sup}(l) - f_{\inf}(l)}{f_{\inf}(l)}$ $= \sqrt[12]{2}$ -1, donde l denota a cualquiera de las bandas de análisis.

Construcción del mosaico

El mosaico está especificado por los parámetros ya mencionados. Dentro de un mosaico posible, identificamos a una baldosa (y al elemento de la base asociado a ella) mediante:

- l : 0 .. L Es el número de banda. 0 es la banda de frecuencias más bajas (que comienza en f_0) y L es la banda de frecuencias más alta.
- k : 0 .. N_l Es el número de baldosa dentro de la banda l. N_l +1 es la cantidad de baldosas de la banda l.

Adicionalmente definimos la constante a_0 : Factor de dilatación de la wavelet. $a_0 = bw_r + 1 = \frac{f_{\sup}(l)}{f_{\inf}(l)} = \sqrt[12]{2}$.

Para cualquier banda l tenemos:

- $f_{\inf}(l) = f_0\ a_0^{\ l}$ Es la frecuencia inferior de la banda l.
- $f_{\sup}(l) = f_0\ a_0^{\ l+1}$ Es la frecuencia superior de la banda l.
- $f_{cent}(l) = \frac{f_{\inf}(l) + f_{\sup}(l)}{2}$ Es la frecuencia central de la banda l. Se utiliza en la construcción de los elementos de la base.
- $bw(l) = f_{\sup}(l) - f_{\inf}(l) = bw_r \cdot f_{\inf}(l) = bw_r\ f_0\ a_0^{\ l}$ Es el ancho de banda de la banda l. Normalmente será un número irracional.
- $long(l) = \frac{1}{2bw(l)}$ Es la longitud temporal de las baldosas de la banda l. Determina que el área de todas las baldosas del mosaico será igual al área de las baldosas del mosaico de una representación temporal. Normalmente será un número irracional.

- $N(l)$ = $redondeo\left(\frac{N}{long(l)}\right)-1$ La cantidad de mosaicos de la banda l, menos 1. Serán tantos como quepan en la longitud de la señal. El último mosaico puede estar incompleto, pero debe caber al menos la mitad de él.

Para cualquier baldosa k de una banda l tenemos:

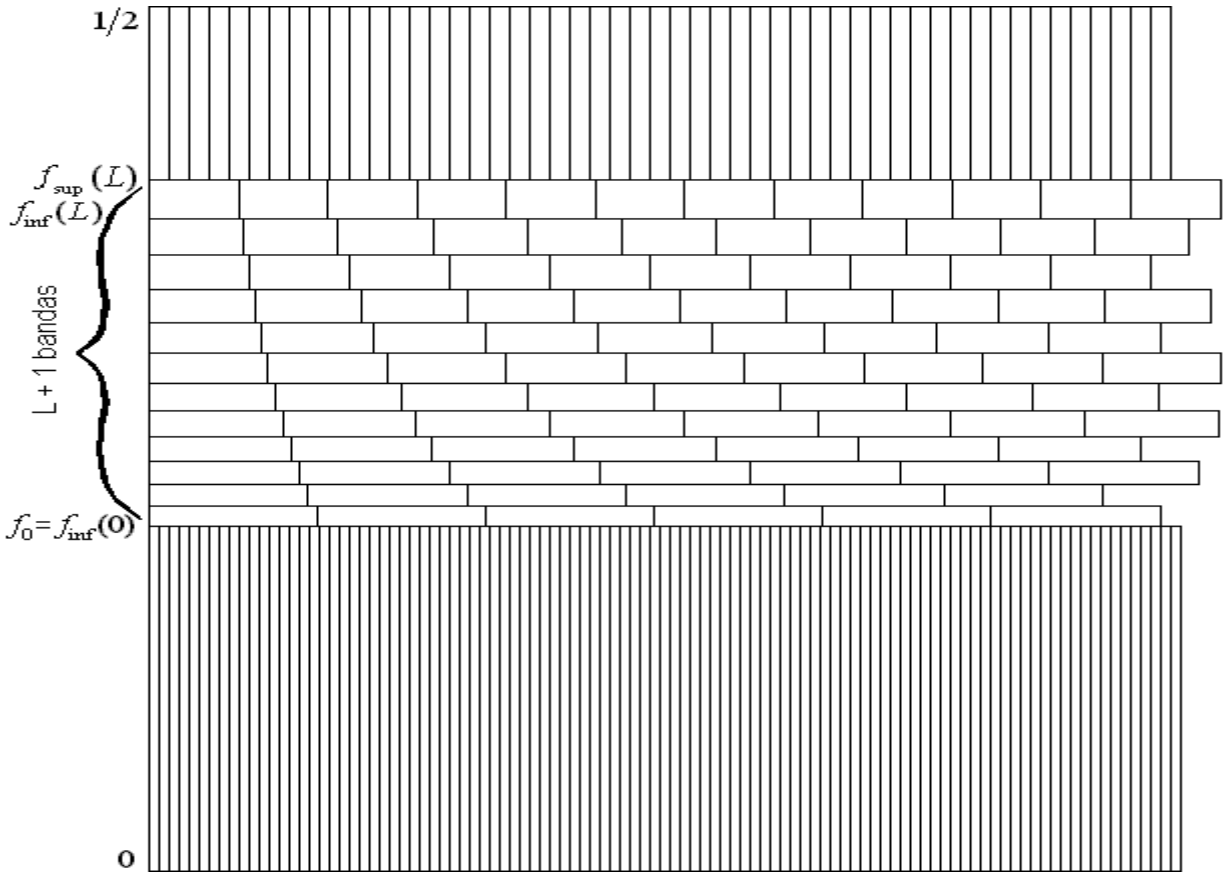
- $t_{inic}(l,k)$ = $k\ long(l)$ El instante donde comienza la baldosa.

- $t_{fin}(l,k)$ = (k +1) $long(l)$ El instante donde termina la baldosa.

- $t_{cent}(l,k)$ = (k +1/2) $long(l)$ El instante central de la baldosa. Se utiliza en la construcción de los elementos de la base.

Adicionalmente, resulta conveniente definir un orden para los elementos de la base y las muestras. Este orden se respetará al construir la base como una matriz con los elementos como columnas, y se respetará también al almacenar secuencialmente los coeficientes de las señales transformadas. Se elije la convención de ordenar por $t_{cent}(l,k)$ y si se repiteran valores, ordenar entre ellos por l.

A continuación se muestra es un ejemplo de mosaico con banda superior y función de escala. Los parámetros son: N = 256, f_0 = 1/5, L = 12-1 = 11. La banda superior resultante captura las frecuencias superiores a 2/5.



Funciones elementales

Ajuste de los elementos de la base al mosaico del plano

Una vez que se establece el mosaico del plano que se desea obtener, es necesario hallar una base con localización temporal y frecuencial que se ajuste lo mejor posible a él. Se ha estudiado extensamente la construcción de bases ortogonales y biortogonales para el caso diádico y M-ádico.

Como ya se dijo, es imposible obtener una señal con soporte compacto a la vez en el tiempo y en la frecuencia. Esto significa que en realidad cada elemento de la base ocupa muchas (de hecho, infinitas) baldosas, y que únicamente está centrado en una baldosa en la cual concentra la mayor parte de su energía. Además, al mejorar la localización en el tiempo se la empeora en la frecuencia, y viceversa. Esto significa que podemos elegir respetar mejor la localización en un dominio que en el otro, sin modificar el mosaico, pero afectando la localización de los elementos en cada dominio. Esta consideración resulta pertinente porque el oído humano es más sensible a la falta de localización frecuencial que temporal (contrariamente a las wavelets discretas convencionales; que privilegian el soporte temporal compacto por sobre la localización frecuencial). Por lo tanto al construir bases se privilegiará el ajuste a la banda de frecuencias de cada baldosa, descuidando (en la medida de lo necesario) la localización temporal, y permitiendo que las wavelets de baldosas vecinas dentro de la misma banda se superpongan en el tiempo de manera considerable.

Como consecuencia de la construcción del mosaico, el intervalo de muestreo de los coeficientes de las distintas bandas en general no es múltiplo del intervalo de muestreo de la señal temporal $1/f_s$. Esto significa que los distintos elementos de la base correspondientes a una misma banda no pueden construirse desplazando en una cantidad entera de muestras un primer elemento ya muestreado, sino que deben ser muestreados independientemente de los otros. Esto sugiere la conveniencia de trabajar almacenando las bases en matrices, donde cada columna contiene un elemento de la base.

Wavelet

Para el análisis de señales mediante la CWT discretizada, suele utilizarse como wavelet una gaussiana modulada, o wavelet Morlet. (Ver [DAU/92], p.76). Esta wavelet alcanza el límite teórico a la localización temporal y frecuencial determinado por el principio de incertidumbre de Heisenberg. (Véase [STR/97a], p.67.)

En una transformada wavelet, todos los elementos de la base que corresponden a bandas de análisis son dilataciones y desplazamientos de un única wavelet fundamental. Esta wavelet es dilatada y desplazada para ubicarla sobre cada baldosa del mosaico. A la baldosa que corresponde a la wavelet fundamental la llamamos baldosa canónica. Como esta wavelet no corresponde a ningún mosaico en particular, la baldosa canónica no forma parte del mosaico.

La wavelet Morlet es una función compleja (R->C). Como este trabajo trata únicamente sobre señales reales, se eligió como primera candidata a una versión real:

$$\Psi_q(t) = b.\sqrt{\pi}.e^{-(b.\pi.t)^2}.\cos(2\pi.f_q.t).$$

El parámetro b controla el ancho de la gaussiana, y f_q es la frecuencia donde está centrado el espectro. La wavelet y su baldosa canónica (como es usual) están centradas en tiempo t = 0. Por lo tanto el único parámetro relevante de la baldosa canónica es f_q.

En la construcción de las bases, es necesario ubicar esta wavelet centrándola en tiempo y en frecuencia sobre cada baldosa. Para ubicarla sobre una baldosa cualquiera (l,k) del mosaico es necesario centrarla temporalmente sobre el t_c = $t_{cent}(l,k)$ y frecuencialmente sobre la f_c = $f_{cent}(l)$ de la baldosa en cuestión. Para esto reemplazamos t por $(t-t_c)\cdot f_c / f_q$. De esta manera:

$$\Psi_{f_c t_c}(t) = b.\sqrt{\pi}.e^{-(b.\pi(t-t_c).f_c/f_q)^2}.\cos(2\pi.f_c(t-t_c)) .$$

Sin embargo, antes de construir la base de esta manera, resulta muy conveniente operar sobre la wavelet para hacerla ortogonal a sus desplazamientos sobre distintos t_c . La forma de hacer esto es construyéndola sobre su baldosa canónica y muestrearla finamente, para luego ortogonalizar esta versión muestreada contra sus desplazamientos. Esta será utilizada posteriormente para construir las bases.

Como vimos antes, las frecuencias se expresan tomando como unidad a la frecuencia de muestreo f_s , y los tiempos se expresan tomando como unidad al intervalo de muestreo $1/f_s$. Esto significa que el muestreo de la wavelet consiste simplemente en evaluarla en valores enteros. Se realiza un muestreo fino, tomando 2^{16} muestras. Esto permite posteriormente obtener los elementos de las bases (remuestreos más gruesos) mediante interpolación lineal, con muy poco error.

Los parámetros para determinar la baldosa canónica son los mismos que ya vimos en la construcción de cada baldosa del mosaico, pero para operar cómodamente en la ortogonalización (sin necesitar remuestrear la wavelet a cada desplazamiento), se elije que el desplazamiento temporal $long_q$ sea entero y se determina f_q a partir de él. Para $long_q$ se adopta el valor 1000, porque proporciona un muestreo detallado de la wavelet, y permite capturar en 2^{16} muestras su parte central, donde alcanza valores relevantes.

Como antes, $bw_r = \sqrt[12]{2}$ -1.

Ya vimos que $long_q$ = $\dfrac{1}{2bw_q}$, y en consecuencia bw_q = 1/2000.

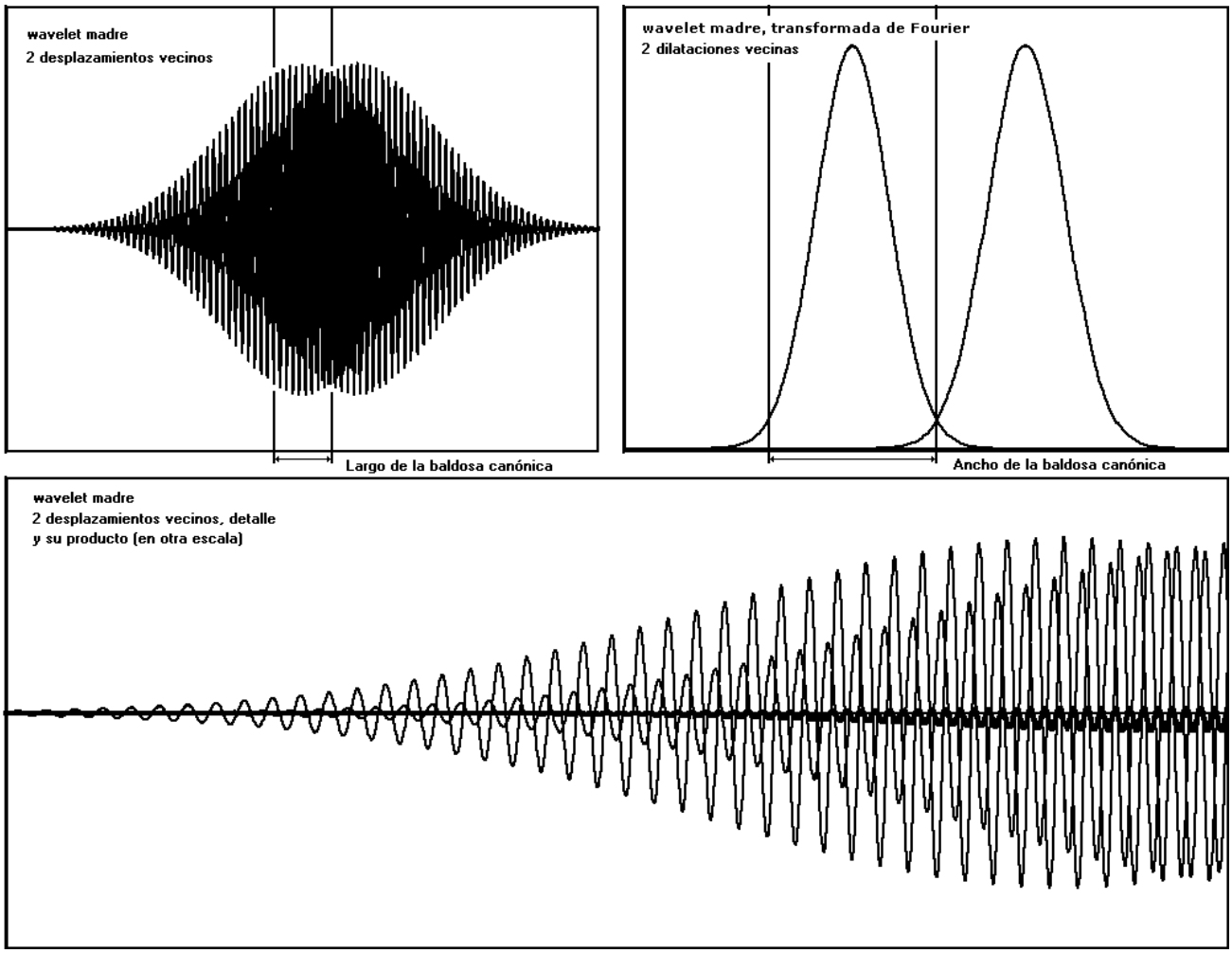
Como $bw_r = \dfrac{bw_q}{f_{\inf q}}$, entonces $f_{\inf q} = \dfrac{bw_q}{bw_r}$.

Finalmente $f_q = f_{\inf q} + \dfrac{bw_q}{2} = \dfrac{bw_q}{bw_r} + \dfrac{bw_q}{2} = \dfrac{1}{2000(\sqrt[12]{2}-1)} + \dfrac{1}{2000\cdot 2} \cong 0.00865858$.

Se adopta b = $0.31\cdot bw_q$ = 0.000155. Este valor se elije porque graficando la wavelet y su espectro se observa un balance apropiado entre localización temporal y frecuencial. Como veremos más adelate (título “Ortogonalización de los elementos de la misma banda”), tenemos una técnica para controlar la correlación debida a la falta de localización temporal de la wavelet. Pero no tenemos una técnica general para controlar la correlación debida a la falta de localización frecuencial de la wavelet. Por esta razón se decide privilegiar la localización frecuencial por sobre la temporal, y utilizar wavelets con el espectro más concentrado, y con el soporte temporal ensanchado. De esta manera las distintas traslaciones de la misma dilatación tienen mayor solapamiento (y correlación), pero como

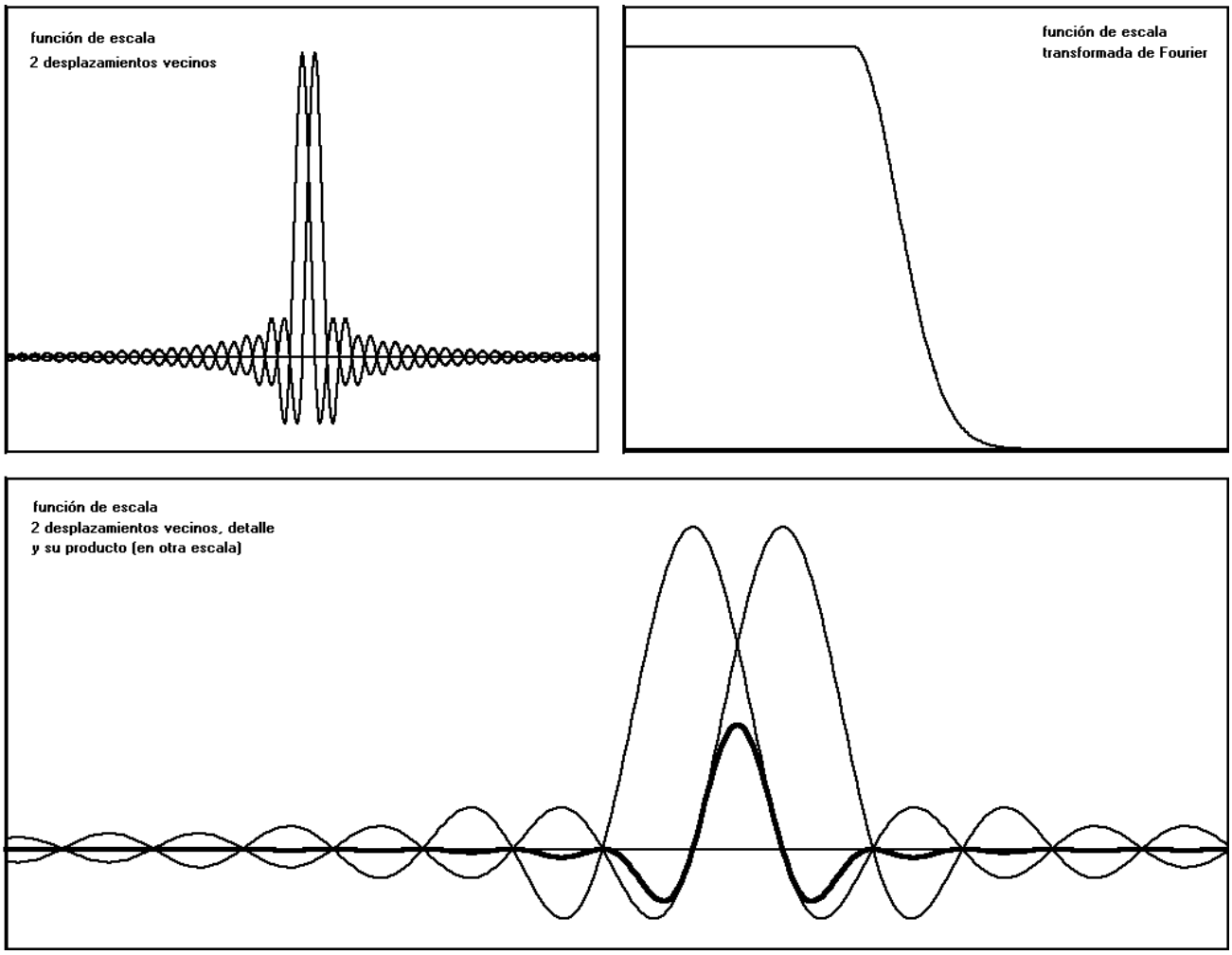
veremos es posible operar entre ellas ortogonalizándolas, propiedad que se mantendrá en todas las dilataciones de la wavelet.

En la siguiente figura se observan dos desplazamientos de la wavelet, y su transformada de Fourier. Debajo se observa un detalle de los 2 desplazamientos y su producto. La suma del producto equivale a la correlación entre los desplazamientos (por estar ya normalizados). El gráfico del producto sirve entonces para tener una primera aproximación a la correlación de los 2 desplazamientos vecinos: Se observa que está mayormente por debajo de cero: habrá correlación, y será negativa.



Función de Escala

Para construir la función de escala, se tomó la mitad superior del espectro de la wavelet (desde el centro de la banda y hacia frecuencias altas); pero se modificó desde el centro de la banda hacia la izquierda, hasta la frecuencia cero, haciéndolo valer 1. Al aplicar la transformada discreta de Fourier inversa se obtiene la función de escala discretizada en el dominio temporal. La razón para construirla de esta manera es que de esta función lo único que realmente interesa es que su espectro abarque las frecuencias inferiores a las bandas de análisis, para no afectar al procesamiento que se realice en el dominio transformado.



"Función de Escala Espejada"

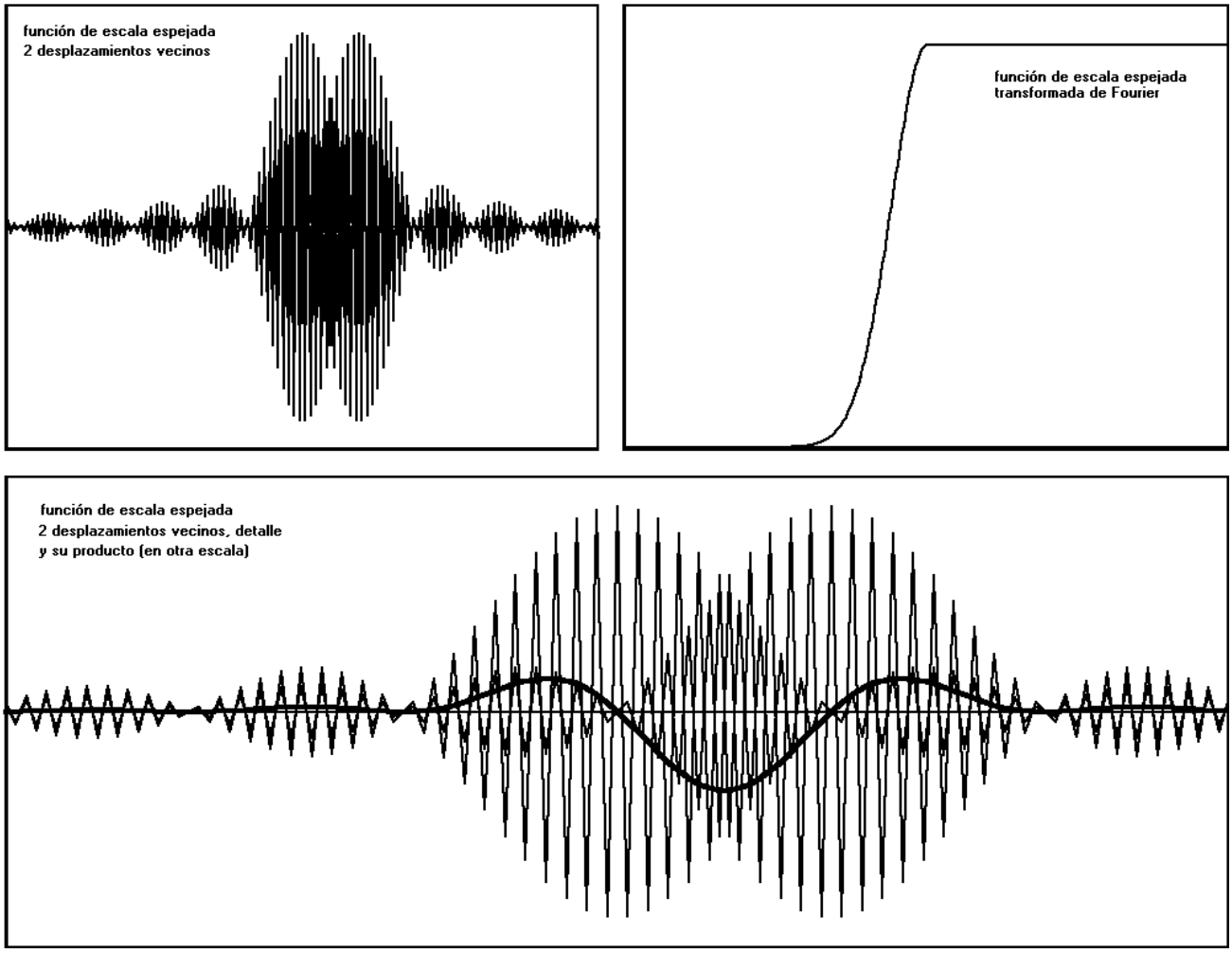
Llamamos de esta manera a la función elemental utilizada para construir los elementos de la base correspondientes a las baldosas de la banda especial superior. Para construir estos elementos se aprovecha que ya fue construida la función de escala, como se describe a continuación.

Si una señal es real y par (y entonces su transformada de Fourier es par y real), la manera de desplazar su espectro operando en el dominio temporal es multiplicándola por un coseno de frecuencia igual al desplazamiento buscado. O sea, si H(f) es la transformada de Fourier (real) de h(t), entonces $H(f - f_0)$ es la transformada de Fourier de $h(t)\cos(2\pi f_0 t)$

Esta técnica de modulación es equivalente al desplazamiento en el dominio de Fourier y la posterior antitransformación. Si se hace esto con un coseno de frecuencia $f_s/2$, entonces el espectro se desplaza en esa misma cantidad. El efecto es reemplazar el espectro entre 0 y $f_s/2$ por el espectro entre - $f_s/2$ y 0 (frecuencias negativas). Como en este el espectro es par, lo que se consigue es invertir el espectro horizontalmente, y la función de escala se transforma en una señal que captura todas las frecuencias superiores a una dada. Esto es justamente lo requerido para la banda superior. Como la unidad de frecuencia es f_s, entonces la frecuencia $f_s/2$ la escribimos como ½. El coseno de frecuencia ½ es $\cos(2\pi \frac{1}{2} t) = \cos(\pi t)$, y es la secuencia (1, -1, 1, -1, ...), y el

espejado del espectro se consigue simplemente cambiando el signo de la señal muestra por medio. (Este efecto es muy popular, por su facilidad de aplicación.)

Por ejemplo, supongamos que se analizará una señal muestreada con calidad de CD (compact disc). La frecuencia de muestreo f_s es de 44,1 KHz, y la máxima frecuencia representable $f_s/2$ es de 22,05 KHz. Supongamos que sólo es necesario analizar hasta 18KHz. Entonces se construye una función de escala remuestreada para tener su frecuencia de corte a 22050-18Hz, o sea 4050Hz. El ancho de las baldosas será el correspondiente a 4050Hz. Y para acomodar el espectro a las frecuencias altas, se multiplica por –1 muestra por medio.



Propiedades algebraicas de las bases

Los elementos de las bases construidas al centrar una wavelet o función de escala en cada baldosa resultan linealmente independientes (y por eso podemos construir bases). Pero presentan una correlación excesiva para poder calcular la transformada en forma numéricamente estable.

Idealmente quisiéramos una bases ortogonales, pero obtener bases de Riesz será suficiente. (Una base A es de Riesz si los autovalores de A^tA están acotados por encima y por debajo por dos números positivos.) Utilizar una base de Riesz es necesario para poder calcular la transformada y la antitransformada en forma estable. (Para más detalles ver [STR/97a], p.69) Si además la base resulta ser ortogonal se simplifican enormemente los cálculos, ya que calcular la transformada es simplemente multiplicar la señal por la traspuesta de la base. Por lo tanto buscaremos siempre conjuntos de vectores con localización ajustada al mosaico, y con la menor correlación posible (correlación cero entre todo par de vectores significa que la base es ortogonal).

Correlación entre los elementos

Para disminuir la correlación entre los elementos de las bases (ya sea para obtener bases ortogonales, o para obtener bases de Riesz), es preciso restar a cada elemento cierta combinación lineal de elementos con los que tiene una correlación significativa.

Un elemento de una base sólo puede tener correlación significativa con otros elementos que sean cercanos temporalmente. Veamos por qué. En álgebra lineal se define a la correlación entre dos vectores f y g como $\frac{< f, g >}{\|f\| \cdot \|g\|}$. Si los vectores (o señales) están normalizados, la correlación es <f, g> = $\sum_t f(t) \cdot g(t)$. Si no hay ningún intervalo temporal en donde tanto f como g tengan valores considerables, la sumatoria será muy pequeña y tendremos muy poca correlación.

Adicionalmente, un elemento de una base sólo puede tener correlación significativa con otros elementos que tengan espectro cercano frecuencialmente. Veamos por qué. Llamemos h(t) = f(t)·g(t), y llamemos F, G y H a las transformadas discretas de Fourier de f, g y h. Por propiedades de la transformada de Fourier, H = F * G (donde * denota la operación de convolución). $H(f) = \sum_w F(w) \cdot G(f - w)$. En particular, $H(0) = \sum_w F(w) \cdot G(-w)$, y como el espectro es simétrico, $H(0) = \sum_w F(w) \cdot G(w)$. Si no hay ningún intervalo de frecuencias donde tanto F como G tengan valores considerables, H(0) resulta muy pequeño. Pero $H(f) = \sum_t h(t) e^{-i 2\pi f / N}$ (Definición de la transformada de Fourier discreta), y entonces $H(0) = \sum_t h(t) = \sum_t f(t) \cdot g(t) = < f, g >$. Por lo tanto la correlación será muy pequeña también en este caso.

Esto significa que no es tan importante la correlación entre elementos centrados en tiempos o frecuencias lejanas, pero sí será necesario operar entre elementos de la misma banda y de bandas cercanas que además estén cercanos en el tiempo. Esto provocará una dilución tanto de la localización temporal como de la localización frecuencial a estas bandas y tiempos cercanos. Es importante en estos casos atacar la correlación con técnicas específicas, que preserven en la medida de lo posible la localización temporal y frecuencial de los elementos de la base.

Ortogonalización de los elementos de distintas bandas

La distancia entre el t_{inic} (l_1) de cada baldosa de una banda l_1 y t_{inic} (l_2) de la baldosa más cercana de una banda l_2 va variando a lo largo del mosaico, pero si existe un k entero tal que k $long(l_1)$ = (k+1) $long(l_2)$, entonces el patrón se repite. Sin embargo, para dos bandas l_1 y l_2, la razón entre sus longitudes será $a_0^{l_2-l_1}$, generalmente un número irracional (excepto para las bandas separadas a n octavas: la razón entre sus longitudes será 2^n). Por esto, la distancia temporal entre pares de elementos de 2 bandas distintas nunca se repite. Esto dificulta en gran medida el análisis de la correlación entre elementos de distintas bandas: habrá que estudiar y eliminar la correlación entre todos los posibles pares de elementos uno por uno. Para esto, es preciso construir la base primero y ortogonalizarla después. Ver sección siguiente: "Construcción y ortogonalización final de la base".

Ortogonalización de los elementos de una misma banda

Al considerar pares de elementos que correspondan a la misma dilatación, resulta posible realizar la ortogonalización sólo una vez. El resultado es una wavelet que es ortogonal a sus desplazamientos sobre todas las baldosas de su misma banda. Posteriormente esta nueva wavelet será utilizada para construir todos los elementos de todas las bandas de análisis de la base.

Dentro de cualquier banda, y en particular de la banda de la baldosa canónica, la distancia entre dos elementos es múltiplo de la distancia entre elementos consecutivos ($long_q$ en este caso), por la construcción del mosaico.

Llamemos F(t) = Ψ_q(t), y G(t) = Ψ_q(t – k $long_q$), con k $\in$ N. Vamos a mostrar que F $\perp$ G => f $\perp$ g, donde son muestreos con una tasa suficientemente alta de F y G.

Llamemos H(t) = F(t) G(t). Sean f y g muestreos de F y G con una frecuencia de muestreo F_s igual o mayor a 4 veces la componente de mayor frecuencia de F y G. Llamemos h(t) = f(t) g(t). Entonces h(t) será un muestreo válido de H(t), ya que su componente de mayor frecuencia no podrá superar a la suma de las frecuencias de las componentes de mayor frecuencia de f y g, o sea ½ de F_s.

Sean $\hat{H}$ la transformadas de Fourier de H. $\hat{H}(f) = \int_t H(t)e^{-i2\pi ft}\,dt$.

Sea $\hat{h}$ la transformada discreta de Fourier de h. $\hat{h}(f) = \sum_t h(t)e^{-i2\pi f/N}$.

Supongamos F $\perp$ G. Entonces 0 = <F, G> = $\int_t F(t)\cdot G(t)dt$ = $\int_t H(t)dt$ = $\hat{H}(0)$.

Como h es un muestreo válido de H, al aplicarle un filtro pasa bajos ideal con frecuencia de muestreo ½ F_s se obtiene H. Esto significa que $\hat{h}_{(f)} = \hat{H}_{(f)} \forall f \le F_s/2$. En particular, $\hat{h}_{(0)} = \hat{H}_{(0)} = 0.$

Pero $\hat{h}(0) = \sum_t h(t) = \sum_t f(t)\cdot g(t)$ = <f, g>.

Por lo tanto f y g son ortogonales.

Esto significa que es posible realizar la ortogonalización de la wavelet contra los demás desplazamientos de su misma dilatación sólo una vez. La wavelet resultante será ortogonal a sus desplazamientos múltiplos de $long_q$ en cualquier muestreo que cumpla con las hipótesis.

Ortogonalización contra desplazamientos a distancias impares

En el curso de los experimentos que se hicieron para desarrollar todas estas ideas, al ortogonalizar la wavelet construida contra los desplazamientos inmediatamente vecinos de la misma dilatación, (+/- 1 $long_q$) se observó que el espectro se desplaza aproximadamente un 1% hacia las frecuencias altas. Lo que se consigue al hacer esta ortogonalización es ajustar la frecuencia f_q de la modulación de la wavelet para que los cosenos de los desplazamientos vecinos tengan un desfasaje muy cercano a π/2, consiguiéndose de esta manera que sean ortogonales. Pero la relación entre f_q y $long_q$ depende del valor de a_0, y no puede cambiarse sin modificar el mosaico, porque para la banda canónica q (y análogamente para cualquier otra banda) tenemos:

$f_{cent\ q} = \frac{f_{\inf q} + f_{\sup q}}{2}$, o sea $f_{\inf q} = \frac{2 \cdot f_{cent\ q}}{bw_r + 2}$;

$bw_q = f_{\sup q} - f_{\inf q} = bw_r \cdot f_{\inf q} = \frac{bw_r \cdot 2}{bw_r + 2} f_{cent\ q}$;

y finalmente $long_q = \frac{1}{2 \cdot bw_q} = \frac{bw_r + 2}{4 \cdot bw_r} \cdot \frac{1}{f_{cent\ q}}$.

Por otra parte, si modificamos el mosaico, tomando bw_r = 1/k, (con k entero) entonces $long_q = \frac{2k+1}{4 f_{cent\ q}}$.

Recordemos la expresión de la wavelet centrada sobre una baldosa:

$$\Psi_{f_c t_c}(t) = b.\sqrt{\pi}.e^{-(b.\pi(t-t_c).f_c/f_q)^2}.\cos(2\pi.f_c(t-t_c)).$$

Consideremos ahora dos desplazamientos distintos, para alguna dilatación. Uno de ellos será t_c = 0, y el otro t_c = $l \cdot long_q$ con l entero. Los cosenos correspondientes a ellos serán

$\cos(2\pi.f_q t)$ y

$\cos(2\pi.f_q(t - l \cdot long_q)) = \cos(2\pi \cdot f_q \cdot t - 2\pi \cdot \frac{2lk+1}{4}) = \cos(2\pi \cdot f_q \cdot t - \pi \cdot (lk + 1/2))$.

El desfasaje entre ambos cosenos será pues de ±π/2, para cualquier l y k enteros.

El resultado es una correlación extremadamente baja, cercana a los límites de precisión del tipo numérico empleado (Float de 32 bits), casi ortogonalidad. La siguiente tabla muestra la correlación obtenida para los primeros vecinos a distancias impares. A distancias mayores, la correlación se reduce aún más.

Vecino a distancia	Correlación
± 1	3.99 e-12
± 3	-2.77 e-11
± 5	3.19 e-12
± 7	9.28 e-13
± 9	-5.42 e-15

Aproximación del mosaico por racionales

Como consecuencia de lo recién expuesto, vemos que resulta necesario modificar el mosaico. El paso siguiente es entonces construir un nuevo mosaico que aproxime lo mejor posible al ideal, pero que cumpla las nuevas restricciones, que consisten en utilizar únicamente bandas l tales que $a_0 = f_{\sup}(l) / f_{\inf}(l) = 1 + 1/k$ con k entero.

Una posibilidad consiste en usar dos dilataciones distintas, cercanas a la ideal, e ir intercalando bandas construidas con ellas para evitar acumular errores que a los pocos semitonos terminen haciéndose inaceptables

Las fracciones de la forma $1 + 1/k$ más cercanas a $\sqrt[12]{2} \cong 1.05946$ son $1 + 1/17 \cong 1.05882$ y $1 + 1/16 = 1.0625$. Entonces, podemos aproximar una octava (12 semitonos de $\sqrt[12]{2} \cong 1.05946$) por 10 bandas ligeramente más angostas y 2 bandas ligeramente más anchas: $(1 + 1/17)^{10} \cdot (1 + 1/16)^2 \cong 1.9993725$. El error cometido es muy pequeño. En cada octava (de 12 bandas) se construyeron con $1 + 1/16$ las bandas tercera y novena, y con $1 + 1/17$ las demás.

Ortogonalización contra desplazamientos a distancias pares

Ya tenemos prácticamente ortogonalidad de un elemento de la base contra los desplazamientos inmediatamente anterior y posterior dentro de su misma banda (dilatación). También tenemos prácticamente ortogonalidad contra vecinos a distancia n impar. Pero tenemos una correlación significativa contra algunos elementos desplazados a distancias n par (para distancias mayores a las mostradas en la tabla, la correlación se reduce aún más):

Vecino a distancia	Correlación
± 2	-0.62
± 4	0.15
± 6	-0.014
± 8	0.0005
± 10	-0.0000071

El primer problema que aparece al intentar ortogonalizar restando la proyección contra el otro elemento (como lo hace Gram Schmidt) es que como todos los elementos están construidos con la misma wavelet, al modificar uno de ellos, en realidad los estamos modificando a todos. En particular aquel contra el que queremos ortogonalizar. Por lo tanto, el nuevo elemento obtenido al restar la proyección es ortogonal al desplazamiento de la wavelet antigua, pero no al de la nueva.

Por otra parte, para mantener la simetría de la wavelet, al restar su proyección sobre su vecino ubicado a la derecha, también lo debemos hacer con su vecino ubicado a la izquierda.

Para complicar aún más las cosas, al operar contra los vecinos ubicados a +/- 2 baldosas afecta la ortogonalidad contra vecinos más lejanos, ubicados a ± 4, ± 6, ± 8 y ± 10 baldosas.

Sin embargo, es suficiente con reducir la correlación por debajo de cierto umbral, sin ser necesario ortogonalizar realmente. Esto es así porque se observó una correlación significativa (cercana en algunos casos a 0.07) entre elementos de bandas vecinas correspondientes a una aproximación racional distinta (una banda con k= 17 y otra con k = 16). Por lo tanto será necesaria una ortogonalización final de la base completa. Sin embargo sí es conveniente reducir la correlación todo lo posible para evitar la gran pérdida de localización temporal y frecuencial que se produciría al hacer la ortogonalización final partiendo de elementos con correlación tan alta. Se establece como umbral aceptable un valor de correlación de 0.01.

Teniendo en cuenta lo anterior, se procedió así:

- Recorrer los vecinos ubicados a n = 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22 y 24.

- Si la correlación contra el vecino a distancia n resulta relevante (> 0.01), restar la proyección sobre +n y sobre –n, multiplicándola por un factores c entre 0 y 1, obteniendo una nueva wavelet para cada valor de c. Elegir el c que minimiza la correlación entre la wavelet resultante y sus desplazamientos +n y –n. La nueva wavelet es la correspondiente a este c.

- Pasar al siguiente vecino par.

Esta técnica dio buen resultado. El siguiente gráfico muestra la wavelet resultado de aplicar esta idea. Sólo fue necesario restar las proyecciones contra n = 2, 4 y 8, y los coeficientes c usados fueron 0.647453, 0.5556 y 0.5045. Se muestra sólo una parte porque es simétrica (par). Las líneas verticales muestran dónde esta centrado el elemento, y dónde estarán centrados sus desplazamientos vecinos (para la misma dilatación o banda).

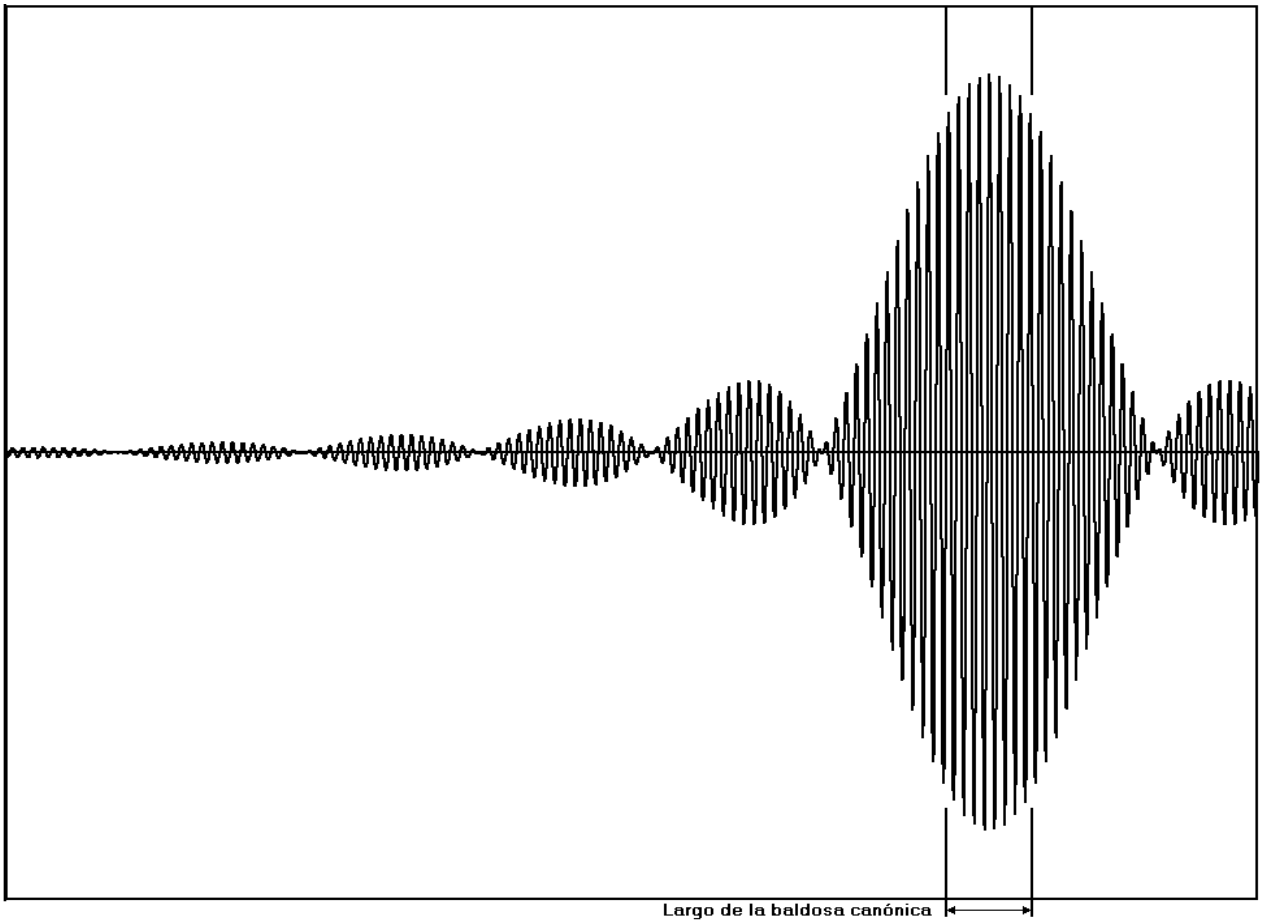
Resulta oportuno comentar por qué esta técnica funciona. Al ortognalizar contra los vecinos a una cierta distancia n, se afecta la correlación contra los vecinos a otras distancias. Los vecinos a distancias mayores no son un problema porque su correlación será atacada posteriormente. Pero además se afecta la correlación contra los vecinos más cercanos, que fue atacada recientemente.

En este caso se ortogonalizó contra n = 2, 4 y 8. Al tomar n = 2, no hay problema, no hay pasos previos que pudieran resultar afectados.

Al ortogonalizar con n = 4, se podría afectar la correlación con los vecinos +2 y –2. Con +2 no hay problema, ya que para el elemento +4, el elemento +2 resulta ser su elmento –2, y por lo tanto son ortogonales. Con –2 sí tenemos un problema. Para el elemento +4, el –2 resulta ser su –6. Pero el problema no es serio, porque la correlación original a ± 6 era –0.014; y como el factor c es 0.5556, se genera una nueva correlación con el vecino –2, pero menor que nuestro umbral de 0.01.

Al ortogonalizar con n = 8, pasa algo parecido, pero la correlación introducida sería con sus elementos –10, -12, y –14. En todos estos casos la correlación es despreciable.

De esta manera se consigue que la correlación contra todos los vecinos esté por debajo del umbral adoptado.



Construcción y ortogonalización final de la base

Para la construcción de la base se eligen los parámetros como fue descrito en "Mosaico del plano de tiempo - frecuencia". A continuación se crea una matriz de N por N. Luego se recorren las baldosas en el orden ya descrito y para cada una se determinan $t_c = t_{cent}(l,k)$ y $f_c = f_{cent}(l)$. Con estos parámetros se muestrea $\Psi_{f_c t_c}(t)$ en valores enteros de t, y se almacena el vector resultante en la columna correspondiente de la matriz. Para los elementos de la base correspondientes a la función de escala y la función de escala espejada se procede de manera análoga. Los muestreos mencionados se obtienen remuestreando las funciones elementales discretizadas ya descritas, utilizando interpolación lineal. Esto se hace así porque la wavelet obtenida en la sección anterior es el resultado de un algoritmo que trabaja con una versión muestreada, y no existe una fórmula para describirla.

La correlación obtenida entre elementos de distintas bandas es bastante reducida. Entre todos los pares de bandas resulta ser menor a 0.01, excepto entre las bandas vecinas obtenidas con aproximaciones racionales distintas. En estos casos la correlación es cercana al 0.07. En consecuencia, resulta imprescindible ortogonalizar la base para poder hacer un cálculo estable de la transformada. Una consecuencia muy deseable de esto es la simplificación de los cálculos, ya que en vez de resolver un sistema de ecuaciones muy grande, alcanza con hacer el producto entre la señal y la traspuesta de la matriz de la base. La ortogonalización implica abandonar parte de la generalidad de

los experimentos realizados: hasta ahora las bases eran de tamaño indefinido (tan grande como se deseara), a partir de aquí es necesario trabajar con bases particulares de cierta longitud.

Para efectuar la ortogonalización se utilizó el algoritmo de Gram Schmidt Modificado. El algoritmo de ortogonalización ataca la correlación entre los elementos de a pares. (Los elementos de la base se mantienen siempre normalizados, pero no se incluye esto en las siguientes fórmulas para facilitar la lectura.) El algoritmo toma cada par de elementos v_1 y v_2 (con correlación $<v_1, v_2>$), y a v_1 le resta su proyección sobre v_2, haciendo $v_1 := v_1 - <v_1, v_2> v_2$. Esta operación degrada la localización temporal y la localización frecuencial de v_1, que ahora se ven ensanchadas abarcando también las de v_2. Pero esto sólo es relevante si la correlación entre ellos era significativa. Por esto es que es de vital importancia que la correlación entre los elementos de la base sea lo más baja posible antes de ortogonalizar.

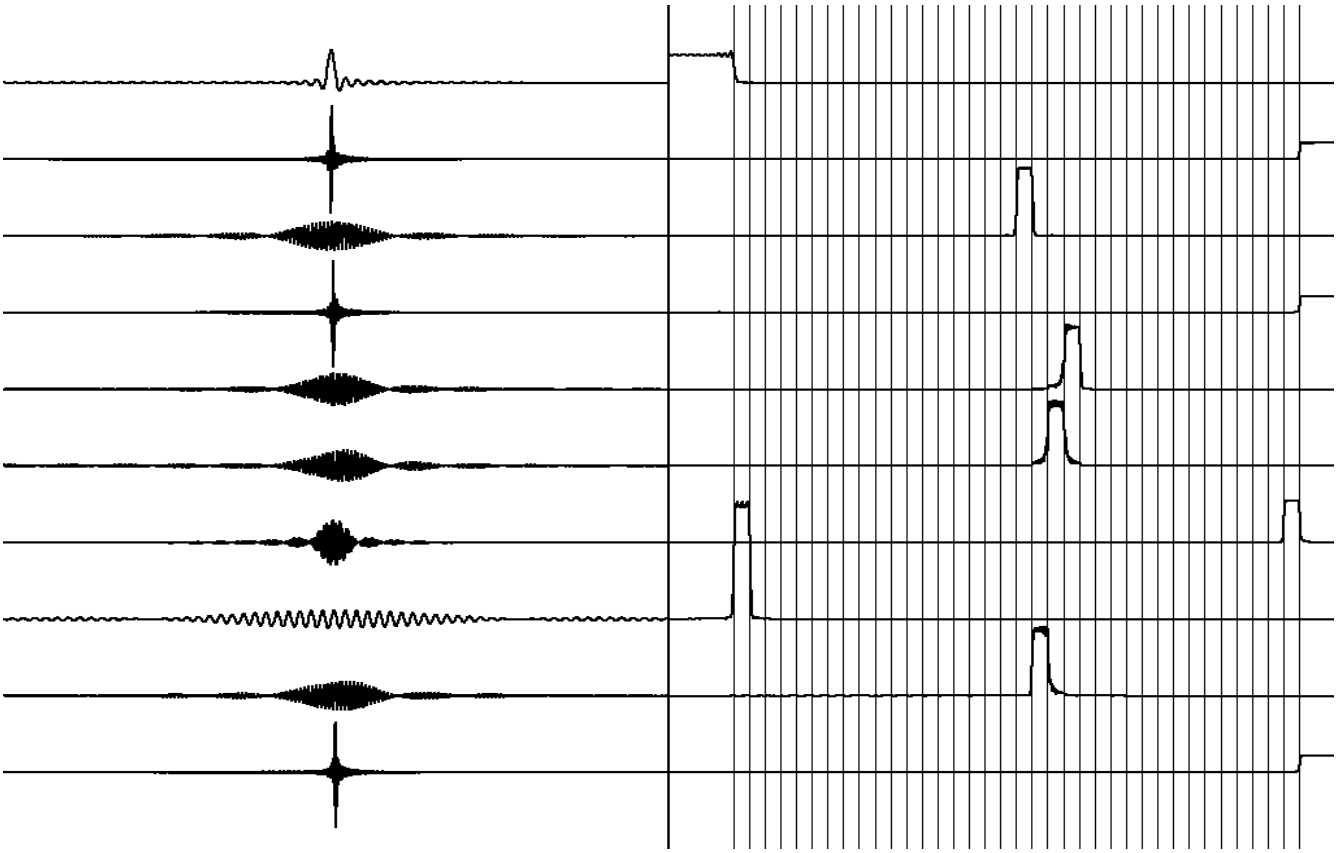
Al aplicar el algoritmo aparece un problema adicional. Como las bases son matrices cuadradas es inevitable que los elementos cercanos al principio y final de la matriz estén recortados. Esto les arruina su localización frecuencial. Esto no es un problema insalvable, porque es suficiente con tomar una base más grande que la señal a procesar, y rellenar ésta con ceros antes y después para mitigar el efecto. Pero si se aplica el Gram Schimdt Modificado, los primeros elementos son restados de todos los demás. Como su localización frecuencial fue seriamente afectada, tendrán correlación muy alta con muchos otros elementos de la base, y al ortogonalizar se arruina la localización frecuencial de ellos también. La solución adoptada es armar una nueva matriz, retirando las primeras columnas y moviéndolas al final. Después se ortogonaliza esta matriz, y las columnas que habían sido movidas se vuelven a colocar en su posición original. (En realidad se modificó la iteración del algoritmo para conseguir este efecto sin mover columnas.) De esta manera se preservan las buenas propiedades de la parte central de la base.

Resultados obtenidos

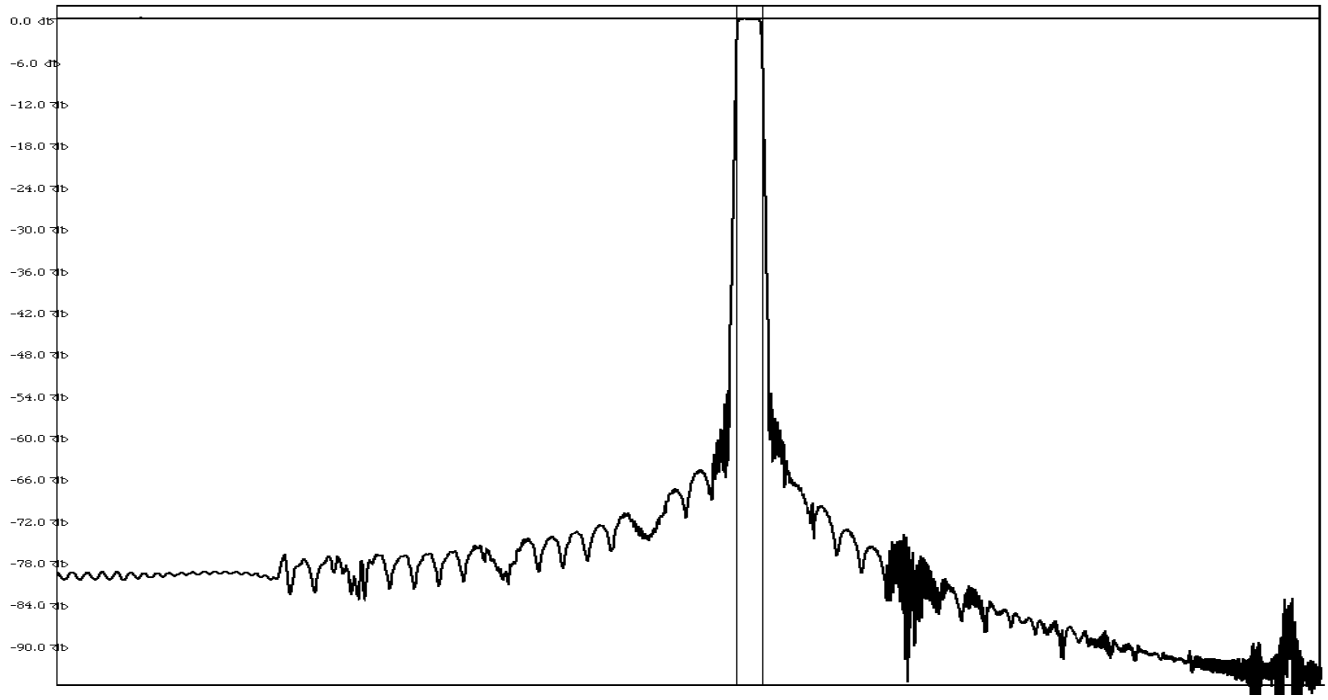
Para realizar pruebas con segmentos de audio se construyó una base con 3 octavas de análisis, desde el do central de 261.5Hz hasta el si 2 octavas más arriba, de 1974.585Hz. La base incluye 36 bandas de análisis, una banda inferior (o función de escala) y una banda superior. La frecuencia de muestreo es baja de 5512.5Hz. Se puede aplicar a señales de 10000 muestras, aproximadamente 1,8 seg. de duración. Esto puede parecer poco, pero es suficiente para obtener resultados audibles, y como la matriz es de elementos de tipo float ocupa 10000 x 10000 x 4 bytes, o sea casi 400Mb, que está cerca del límite de la capacidad de la máquina utilizada para las pruebas.

A continuación se muestra parte de la base construida. En la parte de la derecha se grafica el espectro (módulo de la transformada de Fourier). Se ven las 36 bandas de análisis, con su frecuencia central señalada por líneas verticales. La escala de frecuencias es logarítmica, y en consecuencia todas las bandas se ven de igual ancho. No se grafican todas las frecuencias, se recortaron las frecuencias bajas y altas (correspondientes a la función de escala y la función de escala espejada), para que se vean mejor las bandas de análisis.

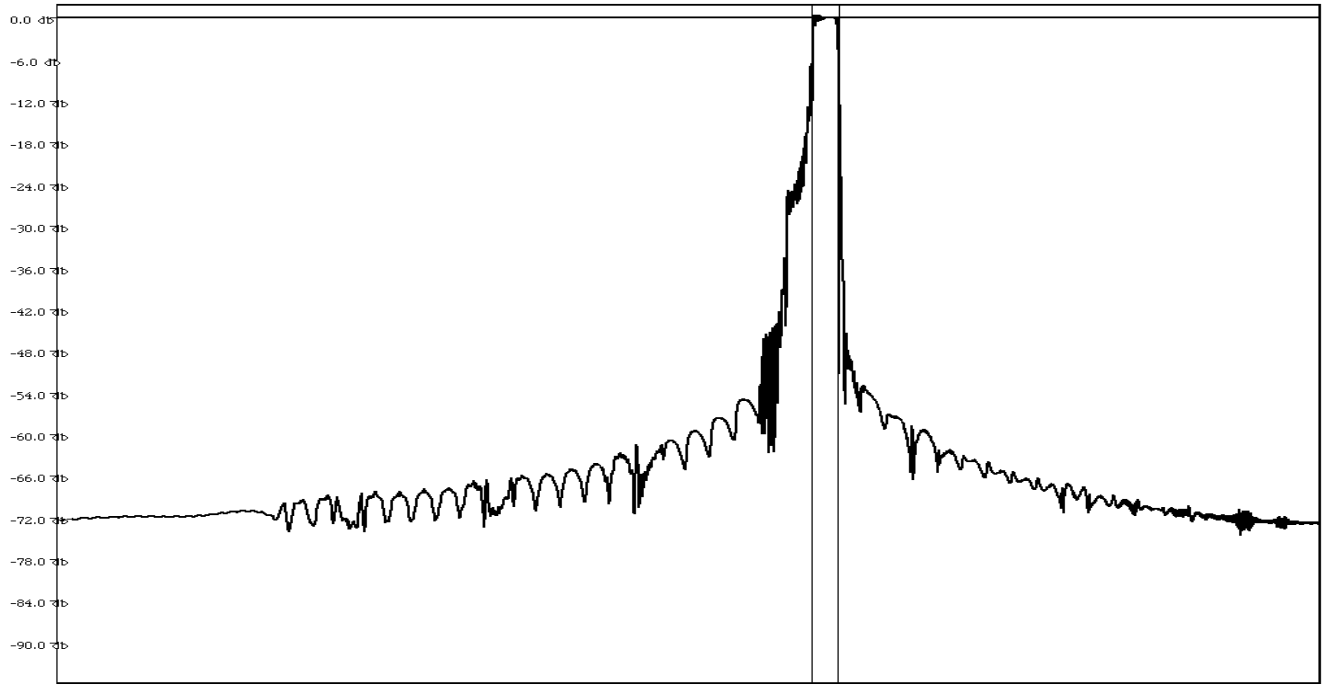
El sexto elemento corresponde a una banda construida con aproximación racional $1+\frac{1}{16}$, y por lo tanto distinta de las vecinas. Antes de la ortogonalización final, este elemento tenía una correlación importante con sus vecinos. Al ortogonalizar se produce una dilución de la localización frecuencial en él y sus vecinos que se observa claramente en el gráfico.



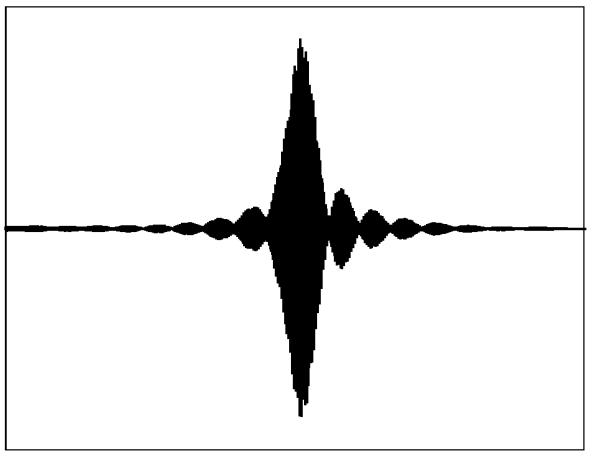
A continuación se muestra la respuesta en frecuencia de un elemento típico de la base. El gráfico muestra la magnitud de la transformada de Fourier , expresada en db (o decibeles, una escala logarítmica usada en audio). El elemento es el tercero mostrado en el gráfico de arriba.

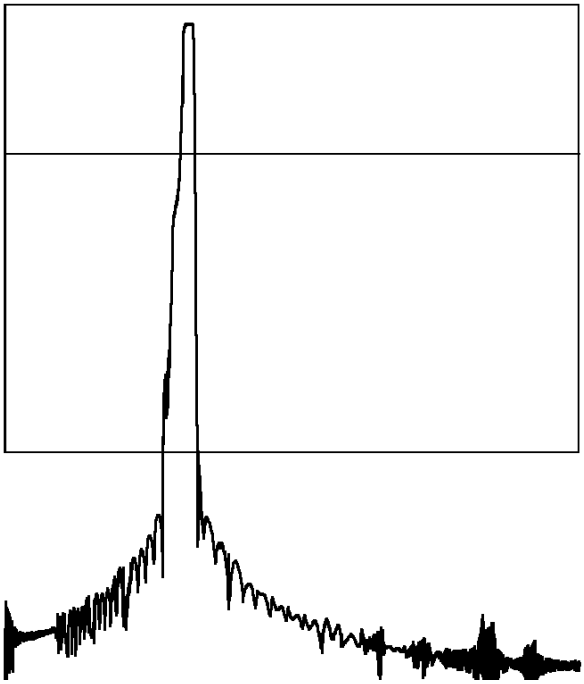


A continuación se muestra la respuesta en frecuencia del sexto elemento en el gráfico de la base, más arriba. La banda vecina inferior (que es el quinto elemento mostrado más arriba) está construida con una aproximación racional diferente. La respuesta en frecuencia no es tan buena como la mostrada recién, pero es muy buena de cualquier manera.



En el subtítulo “Mosaicos arbitrarios del plano de tiempo - frecuencia usando bases locales” se comenta sobre la técnica desarrollada por Bernardini y Kovacevic en [BER/99]. Los siguientes dos gráficos están preparados para ser directamente comparados con los que aparecen en el trabajo de Bernardini y Kovacevic en la página 23. Para ello, el gráfico de respuesta en frecuencia esta graficado en escala frecuencial lineal (eje horizontal), y con el eje vertical entre 20 y –40db.





Se observa que la localización frecuencial es mecho mejor. Esto es de esperar, teniendo en cuenta que la técnica de Bernardini y Kovacevic es más general, y que este trabajo busca una familia más restringida de bases, para aplicaciones más específicas.

Conclusiones

Este trabajo presenta el resultado de una iniciativa de investigación sobre bases ortonormales para representación en tiempo y frecuencia de señales musicales. Las bases construidas son las primeras desarrolladas específicamente para este problema.

Sus virtudes incluyen:

- Son ortonormales
- Tienen excelente localización frecuencial
- Tienen buena localización temporal
- Son relativamente fáciles de construir

Sus defectos incluyen:

- Cada base se construye específicamente para una cierta longitud de señal
- Tienen problemas de localización frecuencial en los extremos, y requieren rellenar con ceros la señal tanto al inicio como al final
- No existe una representación concisa de las bases
- Para una señal de longitud n, se requiere n^2 espacio y $O(n^3)$ tiempo de cómputo

Estos defectos son consecuencia de la aproximación racional y de la necesidad de una ortogonalización global

Trabajos Futuros

El próximo objetivo debería ser hallar bases que corrijan los defectos recién mencionados sin compromenter las virtudes. También queda como trabajo futuro explorar el comportamiento de estas bases en aplicaciones prácticas.

Glosario

CWT Continuos Wavelet Transform. Véase p.11.
DFT Discrete Fourier Transform. Véase p.8.
DWT Discrete Wavelet Transform. Véase p.11.
DDWT Dyadic Discrete Wavelet Transform.
FT Fourier Transform. Véase p.8.
FFT Fast Fourier Transform. Algoritmo usual para calcular la DFT. Véase p.9.
JPEG Joint Picture Expert Group. Formato comprimido para imágenes (fotos) muy popular.
MDCT Modified Cosine Transform. Véase p.10.
MP3 Formato de audio comprimido muy popular.
STFT Short Time Fourier Transform. Véase p.10.
WVD Wigner-Ville Distribution. Vease p.10.

Bibliografía

[BER/99] Bernardini R., Kovacevic J. “Arbitrary Tilings of the Time-Frequency Plane Using Local Bases”. IEEE Transactions on Signal Processing, vol. 47, nro. 8, pags. 2293-2304. Agosto 1999. Disponible en http://cm.bell-labs.com/who/jelena/Papers/journals_abstracts.html

[BRI/88] Brigham E. “The fast Fourier transform and its applications”. Prentice Hall Signal Processing Series, Englewood Cliffs, New Jersey. 1988.

[DAU/92] Daubechies I. ”Ten Lectures on Wavelets”. Vol. 61, CBMS-NSF Regional Series in Applied Mathematics. 1992.

[ESP/02] Espen R. “Drum Analysis”. Tesis del Departamento de Informática de la Universidad de Bergen. 2002. Disponible en http://www.ii.uib.no/~espenr/hovedfag/thesis.pdf

[NEW/97] Newland D. “Practical Signal Analysis: Do Wavelets make any difference?”. Proceedings of DETC’97 1997 ASME Design Engineering Technical Conference. Sacramento, California. 1997. Disponible en http://cwllab.kaist.ac.kr/cwllab/lectures/Data/MAE591/Wavelet2.pdf

[NUÑ/92] Nuñez A. “Informática y electrónica musical”. Editorial Paraninfo, Madrid, 1992.

[OLM/99] Olmo G., Dovis F., Benotto P., Calosso C., Passaro P. “Instrument-Independent Analysis of Music by Means of the Continuous Wavelet Transform”. SPIE Conf. on Wavelet Applications in Signal and Image Processing VII, (Denver, Colorado), SPIE Vol. 3818, pag. 716 – 726, 1999. Disponible en http://www1.tlc.polito.it/SAS/olmo_pdb.shtml

[OLS/67] Olson H. “Music, Physics and Engineering”. Dover Publications, Inc., New York. 1967.

[STR/86] Strang, G. “Algebra Lineal y sus aplicaciones”. Addison-Wesley Iberoamericana. Wilmington, Delaware. 1986

[STR/97a] Strang G., Nguyen T. ”Wavelets and Filter Banks”. Wellesley-Cambridge Press, Wellesley, MA. 1997.

[STR/97b] Strang G. “The search for a Good Basis”. No impreso. Disponible en http://www-math.mit.edu/~gs/papers/search.ps.gz

[TOR/99] Torrésani B. “An Overview of Wavelet Analysis and Time-Frequency Analysis”. in Self-Similar Systems, proceedings of the International Workshop (Dubna, Rusia), 1998. Disponible en http://www.cmi.univ-mrs.fr/~torresan/publi.html

[VIL/48] Ville J. “Théorie et applications, de la notion de signal analytique”. Cables et Transmissions 2, 61-74. 1948. Traducido al inglés por I. Selin, RAND Corp. Report T-92, Santa Monica, CA. 1958.