

INTRODUCCIÓN AL ANÁLISIS ESPECTROGRÁFICO DE LA INTERPRETACIÓN MUSICAL

Luis COLOMER BLASCO

Abstract

The purpose of this article is to show the possibilities that spectral analysis techniques can offer to musicology for the systematic study of the musical performance. Spectrography supplies direct images of a musical performance, which reveal the performer's actions on the instrument. Items such as tempo fluctuations, rhythmic detail, tuning alterations and, generally, the most part of rhetoric features of performance practice can be measured and compared with the help of the spectrogram.

Resumen

El propósito de este artículo es mostrar las posibilidades que las técnicas de análisis espectral pueden ofrecer a la musicología para el estudio sistemático de la interpretación musical. La espectrografía proporciona imágenes directas de una interpretación musical, las cuales revelan las acciones del ejecutante sobre el instrumento. Cuestiones tales como las fluctuaciones del tempo, el detalle rítmico, las alteraciones de la afinación, y, en general, la mayor parte de los rasgos retóricos de la práctica interpretativa pueden ser medidos y comparados con la ayuda del espectrograma.

El estudio de la interpretación va adquiriendo cada vez más importancia en la musicología actual. No nos referimos aquí a la investigación encaminada a establecer unos criterios de autenticidad histórica para la interpretación de las obras clásicas, sino a la que tiene por objetivo el análisis de la ejecución musical misma¹. Hay un conjunto de rasgos que caracterizan una época, una escuela de interpretación o incluso que individualizan a cada músico. Entre ellos, algunos son específicamente acústicos, como el peculiar timbre de voz de un cantante o el "sonido" de un instrumentista, mientras que otros pertenecen a los recursos prosódicos y retóricos de la interpretación, como las inflexiones de la voz en los instrumentos de afinación libre, la articulación de los motivos y de las frases musicales, las alteraciones del *tempo*, etc. Sería de gran interés disponer de una forma de representación del hecho sonoro que reflejara directamente la música tal y como suena, y que permitiera observar con detenimiento los rasgos peculiares de cada ejecución musical.

1. A modo de ejemplo podemos mencionar el libro de Robert PHILIP, *Early Recordings and Musical Style: Changing Taste in Instrumental Performance 1900-1950*, Cambridge, 1992.

Por otra parte, el estudio de las músicas no escritas requiere también una representación adecuada que facilite su análisis, unas veces porque sus elementos más significativos no se corresponden con nuestro lenguaje musical (por ejemplo, las escalas y afinaciones de muchas músicas étnicas no coinciden con las nuestras), otras, porque nuestro sistema de notación no recoge la pluralidad y riqueza sonora de la ejecución musical (pensemos, por ejemplo, en una improvisación de *jazz*).

El presente trabajo pretende mostrar las posibilidades que las técnicas de análisis espectral ofrecen a la musicología para el estudio de la interpretación musical. Estas técnicas permiten obtener imágenes directas de una ejecución, imágenes que, además, se corresponden con nuestra manera de percibir la música, por lo que pueden ser de gran utilidad para objetivar la multiplicidad de recursos sonoros de una interpretación. En primer lugar explicaremos cómo se pueden obtener estas imágenes; luego examinaremos en un caso concreto algunos rasgos estilísticos de una interpretación musical que pueden ser estudiados mediante su espectrograma; finalmente mostraremos cómo afecta el derrame espectral a la representación espectrográfica de la música. Para ello utilizaremos como ejemplo un fragmento del segundo movimiento —*Adagio molto espressivo*— de la Sonata para violín y piano Op. 24, *Primavera*, de Beethoven (compases 18-21) interpretada por Henryk Szering e Ingrid Haebler².

Fig. 1. Beethoven, Sonata para violín y piano Op. 24, II, cc. 18-21.

2. Henryk SZERING y Ingrid HAEBLER, Beethoven. Sonatas para violín y piano núms. 5 y 9. Philips Classics Productions 1980.

I. Representación espectrográfica de la música

El sonido es la sensación producida en nuestro oído por el movimiento vibratorio de los cuerpos, el cual al propagarse origina pequeñas fluctuaciones en la presión del aire. A la descripción de estas variaciones de la presión del aire a lo largo del tiempo en un punto tomado como referencia (por ejemplo, el lugar donde está el micrófono) se denomina señal de *audio*. Cualquier acontecimiento sonoro puede quedar definido por esta señal, bien sea un solo sonido, una sinfonía entera o varias conversaciones simultáneas. Mediante un sistema digital la señal de *audio* puede ser *muestreada* y transformada en una lista de números, lo que facilita su procesamiento.

Ahora bien, aunque la señal de *audio* es una buena representación del hecho físico sonoro, no es adecuada para describir la ejecución musical, pues no hay analogía entre la señal de *audio* y la ejecución y percepción musicales. Veámoslo con un ejemplo. Con la ayuda de un editor de sonido obtenemos en un ordenador a partir de su registro sonoro la señal de *audio* correspondiente al fragmento antes señalado de la Op. 24 de Beethoven. La gráfica de la figura 2 muestra las 20 milésimas de segundo (ms) en las que se produce la transición del compás 19 al 20; el tiempo se expresa en segundos (s) y la amplitud en unidades arbitrarias (u.a.).

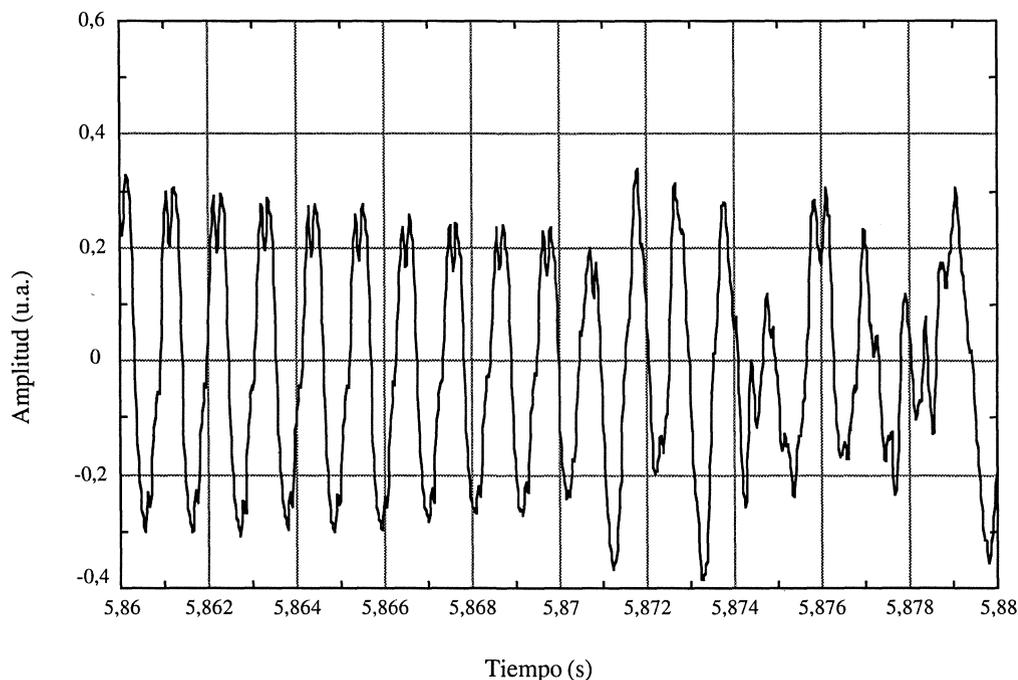


Fig. 2. Señal de *audio* correspondiente a la transición del compás 19 al 20.

Esta gráfica corresponde justo a la articulación entre el sib_5 y el re_6 del violín; representa exactamente los momentos finales de la semicorchea sib_5 y los iniciales de la negra re_6 , esta última simultaneada con el ataque del acorde del piano (mi_3 , sol_3 , sib_4 y re_5). Describe el movimiento físico que experimentó la membrana del micrófono (es decir, las rápidas variaciones de la presión del aire en el “punto” donde estaba colocado) durante el brevísimo intervalo de tiempo comprendido entre 5,86 s y 5,88 s. Sin embargo, su falta de adecuación a nuestra percepción musical resulta evidente. Lo primero que observamos es que la escala de tiempos en la que se mueve excede por completo los límites de la percepción temporal humana. Para obtener una gráfica más próxima a nuestra temporalidad con el editor de sonido comprimimos cien veces la representación (ver figura 3).

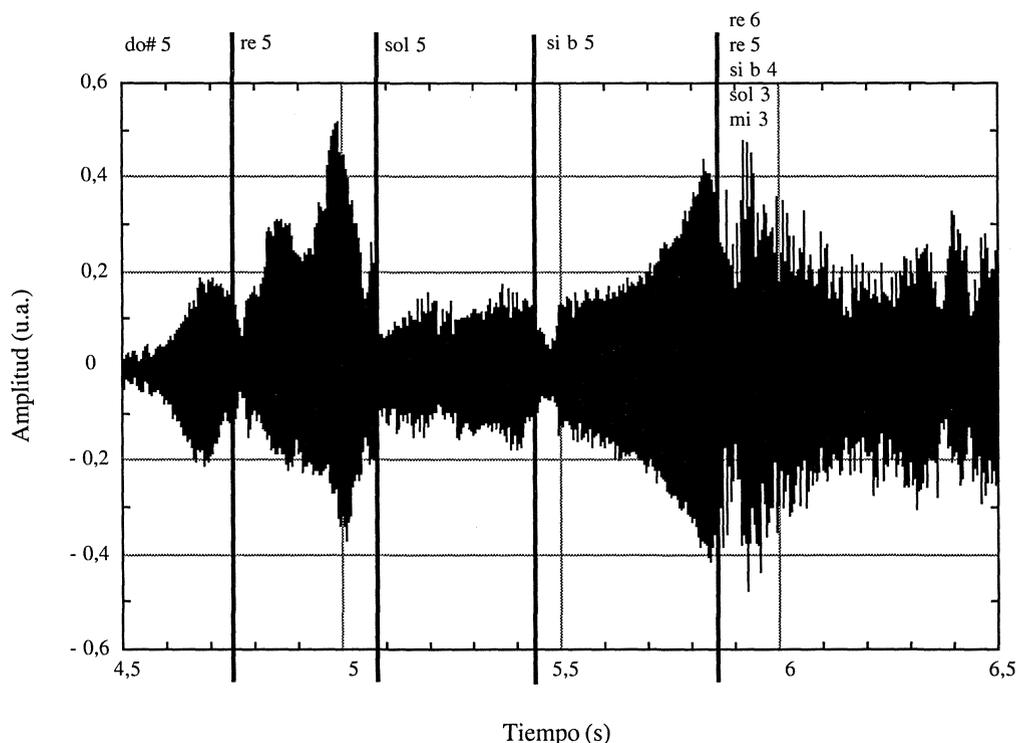


Fig. 3. Señal de *audio* comprimida correspondiente a la tercera parte del compás 19 y al inicio del 20.

Esta nueva gráfica abarca una extensión temporal de dos segundos, desde las cuatro semicorcheas de la tercera parte del compás 19 dadas por el violín ($do\#_5$, re_5 , sol_5 y sib_5), hasta los primeros instantes del comienzo del compás siguiente, con el re_6 del violín y el acorde del piano. Describe, así pues, acontecimientos que evolucionan ya dentro de nuestra escala temporal de per-

cepción, por lo que presenta una cierta analogía con lo que nosotros realmente oímos: distinguimos la articulación de cada nota, el aumento de la presión del arco sobre el *do#₅*, la menor amplitud del *sol₅*, las diferentes duraciones de las semicorcheas, etc. Pero tampoco ahora podemos reconocer las notas que escuchamos.

Únicamente cuando la señal musical está constituida por un sonido aislado es posible hacer conjeturas sobre la nota de la que se trata (por ejemplo, en la gráfica anterior podríamos medir el periodo de los ciclos iniciales, que aproximadamente es de 1,1 ms, y concluir que corresponden a la nota *sib₅*). Pero si pretendemos determinar las notas que están en el inicio del compás 20, sus respectivas intensidades, la forma en la que evolucionan, etc., no tendremos medio de individualizarlas. Ello se debe a que la señal de *audio* está representando el movimiento físico resultante de la suma de perturbaciones producidas por cada instrumento y por cada cuerda del piano vibrando simultáneamente, mientras que el objeto de nuestra percepción auditiva no es el paso a paso detallado del movimiento vibratorio, sino la evolución de sus parámetros *frecuenciales*. Nosotros percibimos frecuencias, notas, no el movimiento físico del acontecimiento sonoro. Por eso, aunque la señal de *audio* describe fielmente la perturbación física que constituye el sonido, no sirve como representación de lo que nosotros verdaderamente oímos.

Sin duda, un sistema de representación de interés musical tiene que ser coherente con nuestra “imagen” acústica; además, debe servirnos para rastrear las acciones del instrumentista, incluso más allá de lo que la mera audición es capaz de percibir. Por ello ha de describir con la mayor precisión posible la evolución a lo largo del tiempo de la frecuencia y amplitud instantáneas de los componentes de los sonidos que constituyen la señal. El procedimiento usualmente empleado para aproximar los valores instantáneos de los parámetros *frecuenciales* en un tiempo *t* consiste en aplicar la transformada de Fourier a un breve fragmento de la señal en torno a *t* y considerar que los picos más relevantes de la función obtenida se corresponden con los componentes sinusoidales presentes en la señal en ese instante. Este procedimiento guarda semejanza con lo que realiza en nuestro oído la membrana basilar. Ello implica presuponer que durante la duración de ese intervalo de tiempo en torno a *t* (al que se suele denominar “ventana de observación”) no se producen cambios significativos en la amplitud y en la frecuencia de los componentes de la señal; más adelante explicaremos la problemática que esta suposición lleva consigo.

Dado que en la práctica, al operar con un ordenador, nos movemos en un mundo discreto (es decir, tratamos con señales digitales, discontinuas, aunque suficientemente *muestreadas*, y realizamos cálculos digitales), las variables con las que trabajemos (tiempo y frecuencia) serán necesariamente también discretas. Puesto que utilizaremos un algoritmo digital para calcular la transformada de Fourier, la *Fast Fourier Transform* (FFT), tendremos que fijar el número de componentes sinusoidales que queramos obtener para conseguir la resolución *frecuencial* que nos interese. Igualmente, no podremos hacer un desplazamiento continuo de la ventana de observación, sino que ésta deberá moverse a saltos. Así pues, el procedimiento consistirá en la transformación desde el dominio del tiempo al de la frecuencia de breves fragmentos de la señal de *audio* sucesivamente desplazados y solapados.

Veamos este procedimiento utilizando como ejemplo, una vez más, la interpretación de los compases arriba indicados de la Op. 24 de Beethoven. En este caso tomaremos un valor de 50 ms para el tamaño de la ventana de observación (el intervalo a considerar como simultáneo) y de 10 ms para su desplazamiento a lo largo de la señal, lo cual significa que realizaremos 100 transformaciones por segundo, obteniendo una buena aproximación a la aparente continuidad del sistema auditivo humano. Para detallar el funcionamiento del algoritmo elegiremos el fragmento de la señal de *audio* que va desde 1,78 s hasta 1,88 s (ver figura 4). Cada uno de los óvalos de esta gráfica delimita los sucesivos desplazamientos de la ventana de observación.

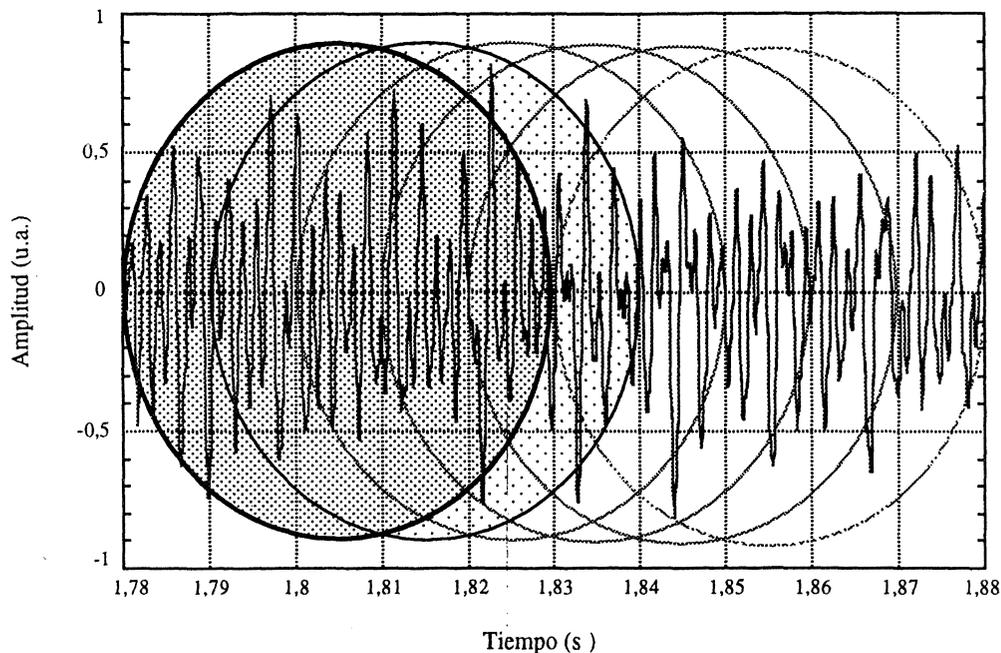


Fig. 4. Desplazamiento de la ventana de observación a lo largo de la señal de *audio* en el inicio del compás 19.

Con la ayuda de un programa de tratamiento de señal³ calculamos la transformada de Fourier del primer fragmento (las 50 milésimas de segundo que van de 1,78 s a 1,83 s y que aparecen en la figura 4 con un óvalo más destacado). La gráfica de la figura 5 presenta la amplitud en función de la frecuencia para los primeros 1.000 hertzios (Hz).

3. El programa MATLAB con el suplemento de las "Signal Processing Toolbox" es el más utilizado en tratamiento de señal.

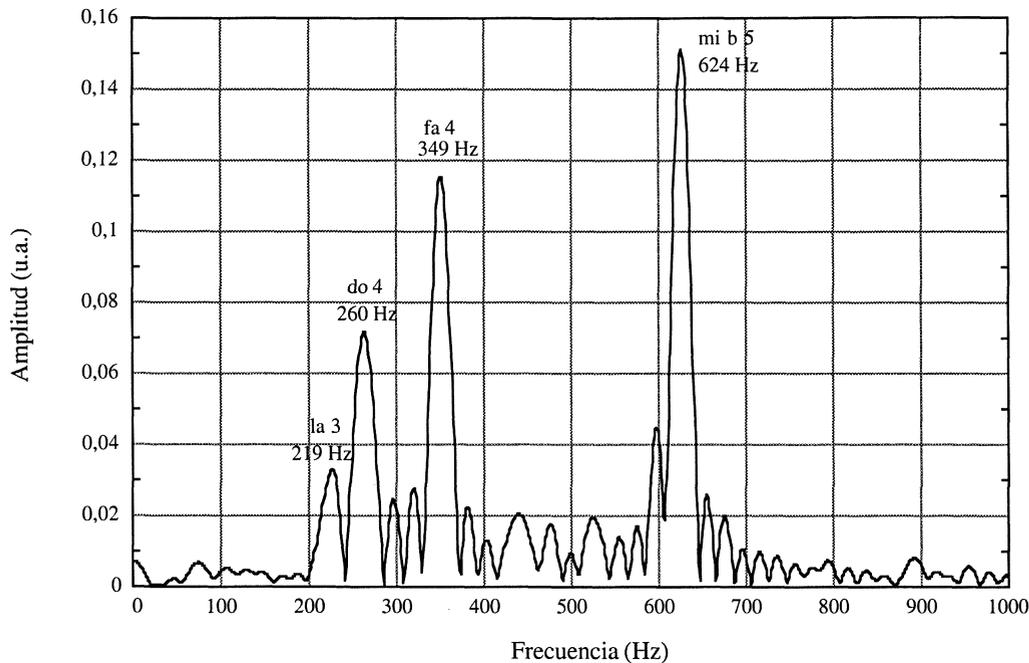


Fig. 5. Transformada de Fourier del primer fragmento de la señal de la figura 4.

Esta gráfica muestra, hablando *grosso modo*, las notas que estaban sonando en ese intervalo de tiempo considerado como estacionario (los primeros 50 ms). Efectivamente, observamos que los máximos más destacados de esta función se corresponden con los componentes sinusoidales de esas notas: los picos en 219, 260, 349 y 624 hertzios coinciden, aproximadamente, con los componentes fundamentales de la_3 , do_4 , fa_4 y mib_5 según la afinación habitual del piano ($la_4 = 440$ Hz).

Para proseguir el análisis desplazaremos ligeramente la ventana de observación (en este caso el desplazamiento que antes hemos elegido, 10 ms) y calcularemos la transformada de Fourier del nuevo intervalo (el segundo óvalo de la figura 4). La sucesiva reiteración de este proceso a lo largo de la señal proporcionará una lista de números capaz de representar de un modo lo suficientemente aproximado la evolución de la amplitud y de la frecuencia instantáneas de los componentes sinusoidales durante todo el fragmento musical. En la gráfica de la figura 6 podemos apreciar la evolución en el tiempo de los parámetros *frecuenciales* de la señal: muestra la amplitud (representada por la altura) en función de la frecuencia y del tiempo, desde 1,6 s a 2,6 s para los primeros 1.000 Hz.

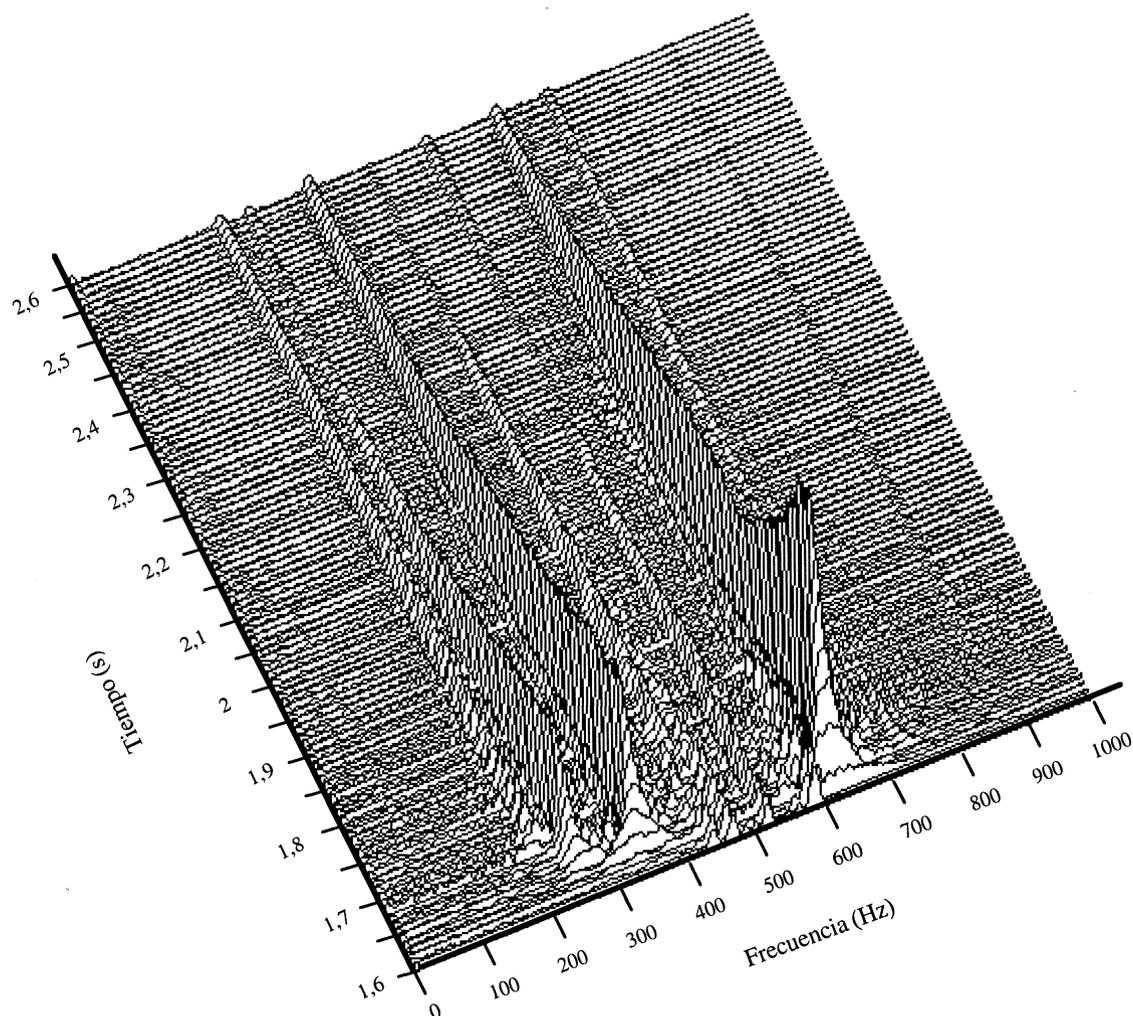


Fig. 6. Representación tridimensional de las sucesivas transformadas de Fourier correspondientes al acorde de piano de la primera parte del compás 19.

Esta gráfica corresponde al acorde del piano de la primera parte del compás 19 (la_3 , do_4 , fa_4 , mib_5). En ella podemos distinguir, entre otras cosas, el carácter destacado del mib_5 (626 Hz), un ligero retraso en el ataque de la nota grave del acorde, la_3 (220 Hz), y la prolongación durante unos instantes del re_5 (587 Hz) del compás anterior. No obstante, esta presentación tridimensional tiene varios inconvenientes: por ejemplo, resulta difícil determinar con precisión el valor

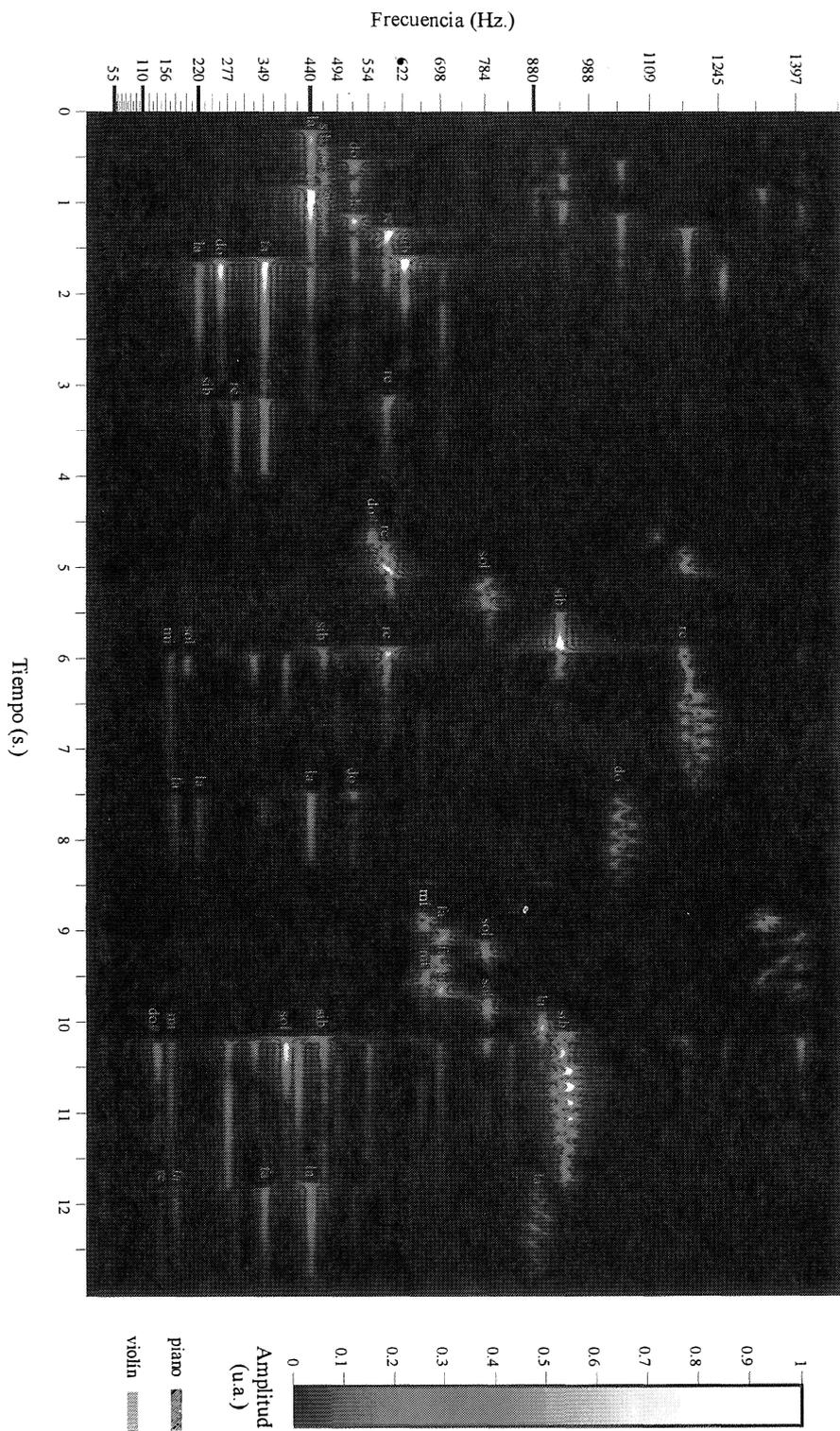
de las frecuencias; los componentes de mayor amplitud pueden ocultar a los más pequeños; y, además, no coincide con nuestros hábitos de notación musical.

Dos pequeñas modificaciones permiten obtener una visualización más eficaz e intuitiva: representamos el tiempo en el eje horizontal, discurriendo de izquierda a derecha, y las frecuencias en el vertical, del grave al agudo; sustituimos la tercera coordenada, la altura, por la mayor o menor intensidad lumínica del punto correspondiente. De este modo, a partir de la lista numérica resultante del algoritmo anterior, podemos generar una imagen en la que la coordenada horizontal y vertical de cada *pixel* de la pantalla del monitor se corresponda, respectivamente, con el tiempo y la frecuencia, y la amplitud quede reflejada por una escala de niveles de gris, o, mejor, por una escala de color asociada a un fenómeno físico familiar (por ejemplo, los colores por los que pasa el hierro al calentarse). Cada una de las columnas de *pixeles* de la imagen representará los resultados de las sucesivas transformadas de Fourier. Sobre el espectrograma así obtenido podremos establecer mediciones y efectuar manipulaciones (particularmente cambios en la intensidad y en el contraste) que nos ayuden a apreciar con claridad cosas que de otro modo pudieran haber pasado desapercibidas. La figura 7 muestra en una escala en color la imagen correspondiente a todo el fragmento de la Op. 24 de Beethoven que estamos utilizando en nuestros ejemplos. Si bien los diferentes niveles lumínicos que se aprecian en la pantalla del ordenador se pierden en el papel impreso, sirve para aproximarnos a las múltiples posibilidades que este tipo de visualización puede ofrecer al estudioso de la interpretación musical.

II. Análisis espectrográfico de un fragmento de la Primavera de Beethoven interpretada por Szering y Haebler

Veamos mediante el estudio de este espectrograma qué tipo de información de interés musical es posible obtener de estas imágenes y algunas de sus aplicaciones al análisis de la interpretación. Por razones de espacio hemos limitado la representación de las frecuencias a 1.500 Hz. La intensidad sonora está representada mediante una escala de color que va del negro (valor mínimo) al blanco (valor máximo).

En la imagen se observan dos tipos de líneas, unas nítidas y de cierto grosor, a las que llamaremos líneas principales, y otras más finas, apagadas y difuminadas que repiten por arriba y por abajo el trazado de la principal. Estas últimas aparecen como ecos o sombras de las principales, cada vez más desdibujadas conforme se alejan de ellas, y contribuyen a dar al conjunto un aspecto enmarañado; como luego veremos, se deben al derrame espectral que lleva consigo el propio análisis de Fourier y han de ser consideradas como información espuria, como el “ruido” que todo análisis espectral genera. Las líneas principales son producidas por los componentes de cada sonido: unas por los componentes fundamentales (las que hemos señalado con los nombres de las notas), y otras por sus correspondientes armónicos (aquellas que representan frecuencias múltiplos de las fundamentales, aunque la mayor parte de los armónicos no aparecen en esta imagen porque sobrepasan por el agudo los límites que hemos



elegido). El grosor de las líneas depende, como luego explicaremos, del tamaño de la ventana de observación.

Dentro de las líneas principales fácilmente se distinguen las líneas que corresponden al piano y las que son del violín. Todas las líneas de los componentes del piano siguen el mismo patrón formal: franjas horizontales (debido a que las notas del piano mantienen la frecuencia constante) de mayor o menor longitud según su duración temporal, las cuales alcanzan su máxima intensidad inmediatamente después del ataque para ir apagándose progresivamente. Las líneas de los componentes del violín presentan un trazo y una luminosidad variables: su dibujo sigue los movimientos del dedo del violinista, mientras que las variaciones en la intensidad lumínica describen las modificaciones de la presión del arco sobre la cuerda. Los ensanchamientos que se producen en algunas líneas (en los momentos iniciales de las notas del piano o en los finales de alguna nota del violín como en el *si_b* del compás 19) reflejan lo abrupto de las transiciones, bien por la rapidez del ataque en el caso del piano o por el corte brusco del sonido en el del violín. Los ligeros desvanecimientos y reapariciones de la luz de algún componente (como en el fundamental del *si_b* del acorde del piano del compás 21) son producidos por los batidos que se originan al juntarse componentes de frecuencias próximas; se trata de elementos aleatorios que, cuando no son el resultado de claras desafinaciones, contribuyen a crear una sonoridad más cálida en el instrumento.

Podemos comprobar que las ocho primeras fusas del piano han tenido una ejecución ultraligada: se observa que los sonidos se prolongan mucho más allá del ataque de las notas siguientes, solapándose en el tiempo unas a otras. Ello sugiere un efecto de “pedal de dedo”, ya que constatamos que el pedal de pie no está metido, pues se produce un claro decaimiento del sonido antes de repetirse el ataque de la misma nota. Es posible apreciar, además, que en todas las ocasiones existe un ligero desfase entre los sonidos de los acordes del piano, un ligerísimo arpegiado que siempre va del agudo al grave y que es más destacado en los dos acordes del compás 19, donde el piano está solo. Este desfase no es perceptible por la audición, pero contribuye a destacar la nota aguda de la melodía.

A simple vista es posible distinguir en el espectrograma las tres manifestaciones del diseño melódico y rítmico que constituyen este fragmento () y apreciar los recursos expresivos más destacados. Pero, además, podemos cuantificar los valores temporales y *frecuenciales* de cada nota, con lo que es posible objetivar la alteración de los valores métricos y tonales que se ha producido en la articulación de la frase en esta interpretación.

Para establecer los parámetros temporales hemos medido el tiempo de inicio de cada nota, tomando como valor métrico el intervalo temporal transcurrido entre el ataque de una nota y el de la siguiente (al margen de que las notas se puedan prolongar más o menos, lo cual afecta a la sonoridad, pero no es significativo desde el punto de vista métrico). La tabla de la figura 8 presenta el valor métrico de cada nota, expresado en centésimas de segundo (cs), en las tres realizaciones del diseño musical. Para facilitar la comparación, incluye el tiempo que hubiera correspondido a cada nota si hubiera sido ejecutada con una exacta regularidad métrica. En aquellas notas que van seguidas de un silencio, o en los silencios mismos, no se puede precisar su exacto

Nota	Tiempo inicio	Valor regular	Valor medido
la 4 (p)	26	17,5	17
sib 4 (p)	43	17,5	14
do 5 (p)	57	17,5	14
sib 4 (p)	71	17,5	15
la 4 (p)	86	17,5	15
sib 4 (p)	101	17,5	15
do 5 (p)	116	17,5	15
re 5 (p)	131	17,5	31
mb 5 (p)	162	140	149
fa 4 (p)	164	140	-
do 4 (p)	165	140	-
la 3 (p)	168	140	-
re 5 (p)	311	70	≈89
fa 4 (p)	314	70	-
re 4 (p)	317	70	-
sib 3 (p)	318	70	-
silencio	≈400	70	≈50

Nota	Tiempo inicio	Valor regular	Valor medido
do# 5 (vl)	450	35	24
re 5 (vl)	474	35	33
sol 5 (vl)	507	35	36
sib 5 (vl)	543	35	44
re 6 (vl)	587	140	153
re 5 (p)	583	140	157
sib 4 (p)	584	140	-
sol 3 (p)	588	140	-
mi 3 (p)	589	140	-
do 6 (vl)	740	70	≈80
do 5 (p)	740	70	≈80
la 4 (p)	740	70	-
la3 (p)	742	70	-
fa 3 (p)	743	70	-
silencio	≈820	70	≈50

Nota	Tiempo inicio	Valor regular	Valor medido
mi 5 (vl)	870	17,5	21
fa 5 (vl)	891	17,5	13
sol 5 (vl)	904	17,5	14
fa 5 (vl)	918	17,5	14
mi 5 (vl)	932	17,5	15
fa 5 (vl)	947	17,5	17
sol 5 (vl)	964	17,5	15
la 5 (vl)	979	17,5	22
sib 5 (vl)	1001	140	166
sib 4 (p)	1004	140	159
sol 4 (p)	1004	140	-
mi 3 (p)	1010	140	-
do# 3 (p)	1010	140	-
la 5 (vl)	1167	70	≈101
la 4 (p)	1163	70	≈105
fa 4 (p)	1167	70	-
fa 3 (p)	1168	70	-
re 3 (p)	1168	70	-
silencio	≈1268	70	-

Fig. 8. Tabla de valores métricos.

valor métrico, por lo que en la tabla éste va precedido por el signo \approx . Las líneas de separación más destacadas señalan las barras de compás.

En primer lugar comprobamos que se mantiene la regularidad métrica en los dos compases que están enteros. Si tomamos como referencia para el inicio del compás las notas agudas del piano (*mib₅*, *re₅* y *sib₄*) observamos que los dos compases duran exactamente lo mismo, 421 cs, lo que equivale a 43 negras por minuto, medida que se ajusta al tempo propuesto por Czerny, 84 corcheas por minuto⁴. Así pues, la duración regular, la esperada según el tiempo del metrónomo, sería de 140 cs para la negra, 70 cs para la corchea, 35 cs para la semicorchea y 17,5 cs para la fusa.

Vemos, por otra parte, que las tres repeticiones del diseño rítmico siguen el mismo patrón interpretativo:

1) Alargamiento de las negras: en 9 cs en el *mib₅* del piano; en 13 cs en el *re₆* del violín; y en 26 cs en el *sib₅* del violín.

2) Prolongación de las corcheas finales, que invaden parte del tiempo del silencio: aproximadamente, 19 cs en el primer caso, 10 cs en el segundo y 31 cs en el tercero, en el *la₅* del violín.

3) Alargamiento de la nota anterior a la negra para realzar su articulación: el *re₅* fusa del piano del compás 18 casi duplica su valor (31 cs frente a los 17,5 que le corresponderían); el *sib₅* semicorchea del violín en el final del compás 20 extiende su valor métrico en 9 cs (esta nota también queda resaltada por ser la única del violín realizada sin vibrato y con la afinación mantenida, así como por su ligero crescendo con final abrupto); el *la₅* fusa del violín del compás 20 se prolonga en 5 cs (esta nota también adquiere más relevancia por su afinación sobreaguda).

4) Acortamiento de las restantes semifusas, tanto en la primera realización del diseño como en la tercera, a excepción de la primera nota de cada serie, que está en su valor justo en el primer caso y ligeramente alargada en el tercero.

5) El valor total de los grupos de 8 fusas o 4 semicorcheas no altera de modo significativo el de la negra equivalente: debería de ser 140 cs y es, respectivamente, 136, 137 y 131 cs (ó 135 cs si tomamos como referencia la nota superior del piano).

En lo que respecta a la determinación de las frecuencias hemos procedido del siguiente modo. Para obtener la frecuencia de cada nota hemos promediado el valor de sus máximos (es decir, los puntos que presentan la mayor intensidad lumínica en el espectrograma). En los sonidos del violín realizados con vibrato hemos tomado el valor en torno al cual oscila la nota. Cuando las notas (con vibrato o sin vibrato) muestran una clara tendencia hacia el agudo o hacia el grave hemos calculado dos valores, el inicial y el final. El margen de error de este espectrograma es de aproximadamente ± 2 Hz⁵. Hay que tener en cuenta que el fragmento musical que estamos analizando ha sido elegido para que los resultados no se vean sensiblemente afectados por

4. CZERNY, Carl. *On the Proper Performance of all Beethoven's Works for the Piano*, edited and with a Commentary by Paul Badura-Skoda, Universal Edition, Viena 1970, pág. 68.

5. Este margen de error resulta suficiente para estudiar los rasgos interpretativos de este fragmento musical, las desviaciones intencionadas del violinista, pero no hubiera sido tolerable para precisar la afinación del piano. Para incluir dentro del estudio la afinación del instrumento (lo cual pudiera ser relevante en la interpretación de música no temperada en el clavicén o en el clavicordio) habría que efectuar primero un *pretratamiento* de la señal adecuado a este objetivo y diseñar un algoritmo alternativo.

Nota	Frecuencia temperada	Frecuencia medida	Diferencia en cents
la 4 (p)	440	440	-
sib 4 (p)	466	466	-
do 5 (p)	523	524	-
sib 4 (p)	466	466	-
la 4 (p)	440	440	-
sib 4 (p)	466	466	-
do 5 (p)	523	524	-
re 5 (p)	587	588	-
mi 5 (p)	622	624	-
fa 4 (p)	349	349	-
do 4 (p)	262	260	-
la 3 (p)	220	219	-
re 5 (p)	587	588	-
fa 4 (p)	349	349	-
re 4 (p)	294	293	-
sib 3 (p)	233	232	-
silencio			

Nota	Frecuencia temperada	Frecuencia medida	Diferencia en cents
do# 5 (vi)	554	560	19
re 5 (vi)	587	588 / 594	- / 20
sol 5 (vi)	784	≈788	≈9
sib 5 (vi)	932	930	-
re 6 (vi)	1174	1175 / 1196	- / 32
re 5 (p)	587	588	-
sib 4 (p)	466	466	-
sol 3 (p)	196	195	-
mi 3 (p)	165	163	-
do 6 (vi)	1086	1048	-
do 5 (p)	523	524	-
la 4 (p)	440	440	-
la 3 (p)	220	219	-
fa 3 (p)	175	174	-
silencio			

Nota	Frecuencia temperada	Frecuencia medida	Diferencia en cents
mi 5 (vi)	659	≈664	≈13
fa 5 (vi)	698	≈700	-
sol 5 (vi)	784	≈786	-
fa 5 (vi)	698	≈698	-
mi 5 (vi)	659	664	13
fa 5 (vi)	698	698	-
sol 5 (vi)	784	786	-
la 5 (vi)	880	895	29
sib 5 (vi)	932	932 / 943	- / 20
sib 4 (p)	466	466	-
sol 4 (p)	392	391	-
mi 3 (p)	165	163	-
do# 3 (p)	139	137	-
la 5 (vi)	880	882	-
la 4 (p)	440	440	-
fa 4 (p)	349	349	-
fa 3 (p)	175	174	-
re 3 (p)	147	146	-
silencio			

Fig. 9. Tabla de frecuencias.

los problemas derivados del derrame espectral, de los que más adelante hablaremos: los pocos casos en los que hay componentes simultáneos muy próximos corresponden a notas de piano de larga duración, en las que, al ser su frecuencia estable, un promediado de sus valores minimiza las posibles desviaciones debidas a la interacción de los componentes. La tabla de la figura 9 muestra la frecuencia medida en cada nota y la que le correspondería según la afinación temperada, así como su desviación en centésimas de semitono (cents) cuando ésta excede el margen de error aceptado. En las notas del violín de muy breve duración y frecuencia inestable sólo se puede indicar su valor de un modo aproximado, por lo que en la tabla va precedido por el signo \approx .

La mayor parte de las desviaciones respecto a la afinación convencional que presentan muchas de las notas del violín son recursos con los que el intérprete expresa el sentido general del diseño melódico, en el mismo sentido que hemos comentado al hablar de la alteración de los valores métricos:

1) En las dos negras (re_6 y sib_5) que son el punto culminante de las dos realizaciones del diseño melódico en el violín, se produce una desviación significativa hacia el agudo durante la ejecución del vibrato: el re_6 del compás 20 empieza en la altura tonal que le corresponde, pero termina 32 cents más arriba; algo menos destacada es la desviación del sib_5 del compás 21, que comienza también en la afinación que le corresponde, pero asciende progresivamente hasta terminar 20 cents más arriba, a la vez que la profundidad del vibrato disminuye.

2) La última fusa del compás 20, el la_5 , se realiza en 895 Hz en lugar de los 880 Hz que le corresponderían, es decir 29 cents por encima de la afinación convencional, lo que intensifica su tendencia hacia el sib_5 negra. Esta desviación se hace más significativa porque el la_5 corchea en el que resuelve ese sib_5 está ya en la altura tonal que le corresponde. Esta alteración de la frecuencia contribuye a destacar la nota anterior al clímax, igual que en las dos repeticiones anteriores del diseño lo ha hecho la realización métrica.

3) La segunda semicorchea del violín en el compás 19, re_5 , empieza en la frecuencia habitual y termina 20 cents más arriba, potenciando así el sentido ascendente de esta primera parte del diseño.

Hasta aquí hemos podido observar la manera en la que las alteraciones de los valores métricos y *frecuenciales* cooperan en la expresión del diseño musical, marcando el clímax con la prolongación y sobretensión progresiva de la nota negra, y subrayando su articulación al destacar de alguna manera la nota inmediatamente anterior. Podemos señalar, además, dos rasgos del violín que se mantienen constantes, los cuales, más que ligados a la intención expresiva de la frase, podrían ser considerados como propios del estilo. Por un lado, todas las notas que van seguidas por un semitono ascendente están sobretensadas: el $do\#_5$ semicorchea en 19 cents, los dos mi_5 fusas aproximadamente en 13 cents, y el la_5 fusa en 29 cents (éste, además, por razones expresivas como antes hemos explicado). Por otro, todas las notas han sido ejecutadas con vibrato, a excepción de la última semicorchea del compás 19, sib_5 , donde su ausencia adquiere valor expresivo. El periodo y la profundidad media de los vibratos en las cuatro notas de cierta duración, corcheas y negras, son similares (el periodo está en torno a 17 cs, es decir, aproximadamente 6 ciclos por segundo, y la profundidad media en torno a los 35 cents).

Si bien la corta duración del fragmento impide estudiar con rigor las características estilísticas de la interpretación, el fragmento que hemos analizado puede servir como muestra de la utilidad del análisis espectral para un estudio sistemático de los recursos interpretativos. Hemos presentado las posibilidades que ofrece el examen de las frecuencias y de las duraciones, los dos parámetros más fácilmente mensurables, tanto por la inmediatez con la que se pueden obtener los datos, como por estar cuantificados en nuestro sistema musical. La medición de las intensidades para realizar un estudio similar de las dinámicas plantea dos problemas importantes: por un lado, la obtención de datos fiables a partir del análisis espectral, y, por otro, su transformación para adecuarlos a nuestra percepción musical (nuestro lenguaje musical posee una escala de duraciones y otra de alturas tonales, pero no tiene una escala de intensidades).

III. Problemática del derrame espectral

Examinemos ahora los problemas derivados del derrame espectral, en la medida en la que afectan a la utilización de las técnicas espectrográficas para el análisis de la interpretación. El proceso por el cual la energía de un componente parece distribuirse a ambos lados del espectro se denomina derrame espectral. El derrame espectral, que es inherente al análisis de Fourier mismo, condiciona tanto la capacidad de distinguir componentes próximos como la de determinar con precisión el valor de cada uno de ellos, por lo que es responsable de las inexactitudes a las que pueden dar lugar algunos espectrogramas. Tres son los problemas que puede ocasionar: 1) componentes próximos pueden quedar solapados y no ser detectados individualmente; 2) el derrame de un componente fuerte puede afectar a otros más débiles, con el consiguiente desplazamiento en la localización de su frecuencia; 3) componentes espurios pueden ser tomados por componentes reales.

Veamos con un ejemplo de qué manera el derrame espectral puede afectar a la exactitud de los resultados del análisis y, en consecuencia, a la fiabilidad de la representación espectrográfica de la música. Generemos mediante ordenador una señal $s_1(t)$ constituida por un solo componente sinusoidal de 0,1 de amplitud y 220 Hz de frecuencia (lo que corresponde al la_3) y calculemos la transformada de Fourier de un intervalo de 50 ms. La gráfica de la figura 10 muestra la amplitud normalizada para los primeros 400 Hz.

Observamos que esta función tiene un pico en los 220 Hz, pero que no se anula para todos los demás valores, sino que presenta lóbulos a derecha e izquierda progresivamente atenuados.

Generemos a continuación dos señales más débiles: una, $s_2(t)$, constituida por un componente sinusoidal de 233 Hz (sib_3) y 0,025 de amplitud, es decir, muy próxima, a un semitono de la anterior, y de 12 decibelios (dB) menos de intensidad; y otra, $s_3(t)$, por un componente de 139 Hz ($do\#_3$) y 0,01 de amplitud, es decir, más alejada, a ocho semitonos de la anterior, y de 20 dB menos de intensidad. Hemos elegido estos valores a propósito para que la perturbación sea nota-

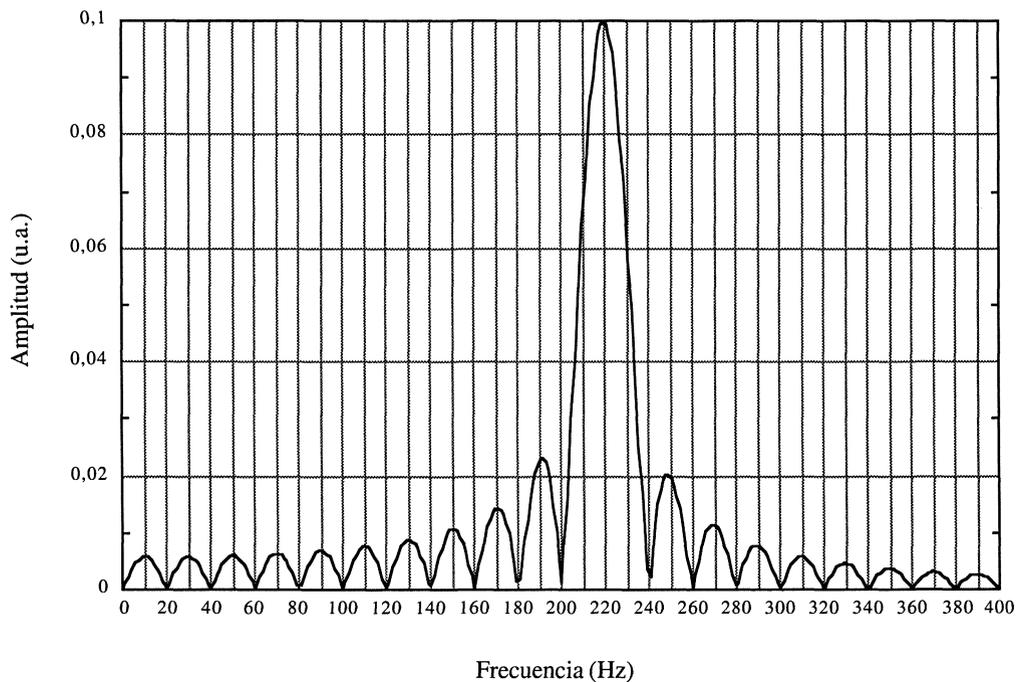


Fig. 10. Transformada de Fourier de un intervalo de 0,05 s de $s_j(t)$.

ble. La transformada de un intervalo de 50 ms de cada una de estas señales tendrá una forma lobulada similar a la de la anterior, pero con el lóbulo principal centrado en su respectiva frecuencia. Sumamos ahora estas tres señales y obtenemos una nueva señal, $s_4(t) = s_1(t) + s_2(t) + s_3(t)$, que estará constituida por los tres componentes sinusoidales anteriores. Calculamos la transformada de Fourier de un intervalo de 50 ms (ver figura 11).

Comprobamos que el componente de 233 Hz ha quedado absorbido por el lóbulo principal del de 220 Hz (sólo una ligera deformación a la derecha delata su presencia). El componente de 139 Hz aparece desplazado a 131 Hz, es decir, 103 cents, un semitono más bajo, mientras que su amplitud ha sido aumentada de 0,01 a 0,017 (4,6 dB más de intensidad). Además, los picos de los lóbulos laterales del componente de 220 Hz (en 190 Hz y en 250 Hz) podrían ser tomados erróneamente por componentes reales. La explicación de lo que ha sucedido es que, al ser la transformada de esta señal la suma de las transformadas de las señales individuales, los lóbulos de cada uno de sus respectivos componentes han interferido entre sí. Por ello, si pretendiéramos identificar los picos más significativos de esta función con los componentes sinusoidales realmente presentes en la señal, obtendríamos un resultado engañoso.

Todos los problemas mencionados de resolución, desplazamiento e información espuria tienen, así pues, su origen en la lobulización que es intrínseca a la transformada de Fourier de un

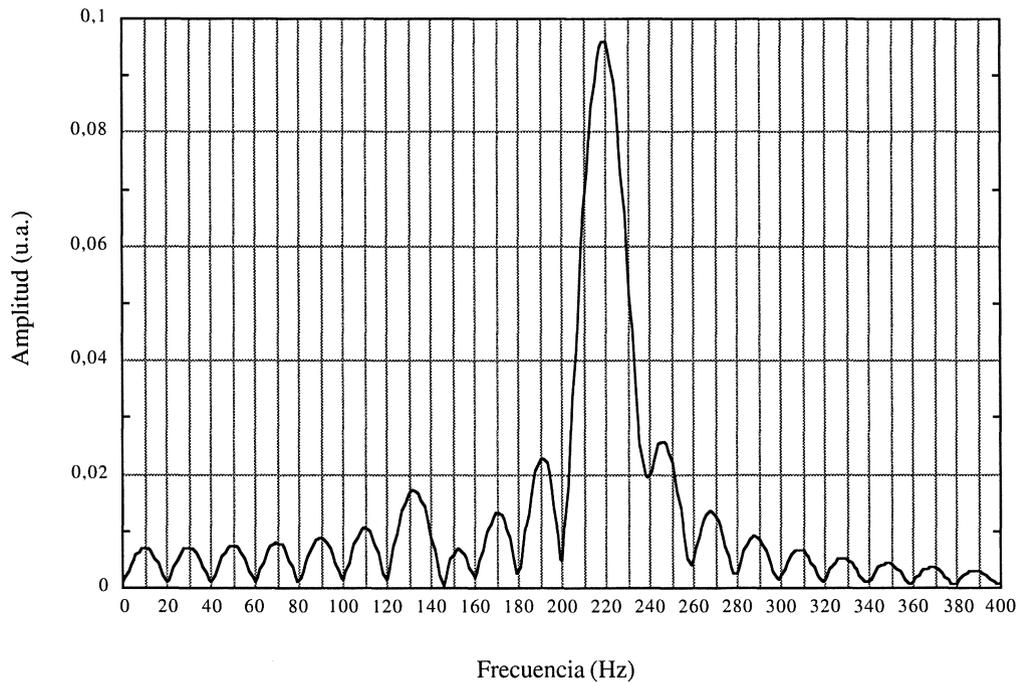


Fig. 11. Transformada de Fourier de un intervalo de 0,05 s de $s_4(t)$.

fragmento finito de señal⁶. Esta lobulización disminuye conforme aumenta el tamaño del intervalo sobre el que se realiza la transformación. Veámoslo tomando sobre la señal $s_4(t)$ un intervalo temporal cuatro veces mayor que el anterior, es decir, de 0,2 s, y calculando su transformada de Fourier (ver figura 12).

6. Calcular la transformada de Fourier de un fragmento finito de una señal sinusoidal equivale a calcular la transformada de Fourier del producto de dos señales infinitas en el tiempo: una, la señal sinusoidal $s(t) = A \sin(2\pi ft)$; y otra, la que define el intervalo de tiempo sobre el que se efectúa el análisis, $v(t)$. Esta última es una función rectangular del tiempo, cuyos valores se hacen 0 para todo tiempo fuera del intervalo de observación y 1 para el tiempo de observación: $v(t) = 1$ para $t_a < t < t_b$, y $v(t) = 0$ para todos los demás valores de t . La transformada de Fourier de la señal sinusoidal infinita, $S(f)$, es una función delta de Dirac desplazada en la frecuencia del componente sinusoidal que la constituye, en la cual el valor de su integral refleja la amplitud. La transformada de Fourier de la señal rectangular es la función

$$V(f) = \int_{t_a}^{t_b} e^{-i2\pi ft} dt,$$

cuyo valor absoluto o magnitud tiene el aspecto lobulado de la figura 10, pero centrada en torno a 0 y simétrica. Teniendo en cuenta que la transformada de Fourier del producto de dos señales es la integral de circunvolución de sus dos transformadas, definida como

$$S(f)*V(f) = \int_{-\infty}^{\infty} S(u)V(f-u)du,$$

resulta evidente que la transformada de Fourier de un fragmento finito de señal constituido por un componente sinusoidal será la misma función $V(f)$ desplazada en la frecuencia del componente sinusoidal y multiplicada por su amplitud.

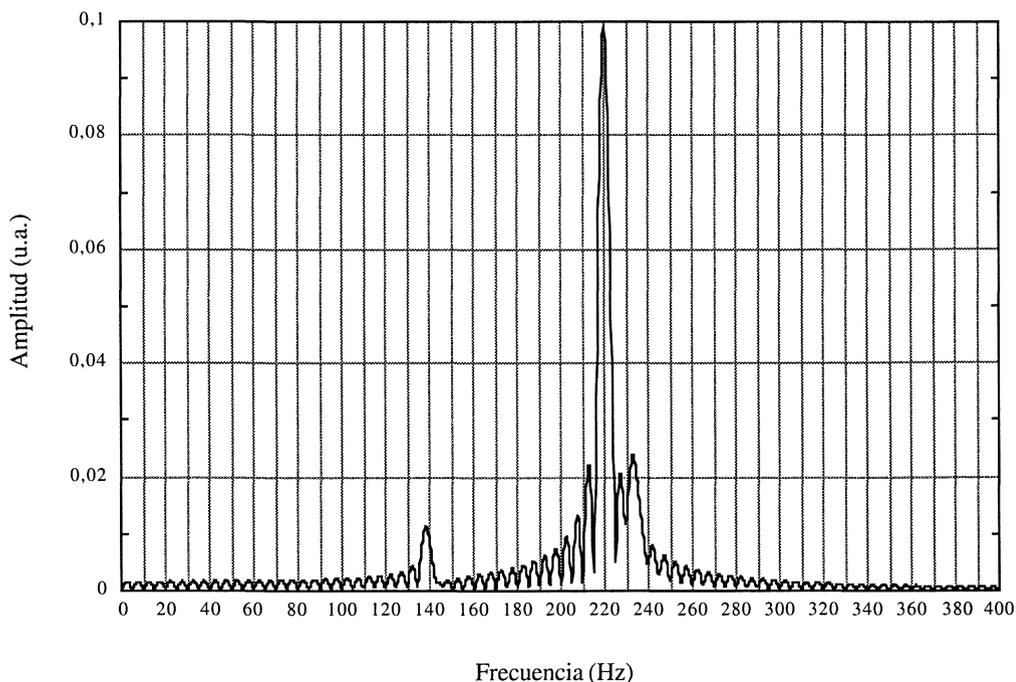


Fig. 12. Transformada de Fourier de un intervalo de 0,2 s de $s_4(t)$.

Observamos que el derrame espectral ha disminuido considerablemente y que la anchura de los lóbulos se ha reducido a la cuarta parte. El componente de 139 Hz, que antes estaba desplazado, se encuentra ahora en su sitio, claramente destacado y aproximadamente con su propia amplitud; y el componente de 233 Hz, antes solapado, aparece ahora netamente diferenciado en el lugar y con la amplitud que le corresponde. Generalizando, conforme aumentamos el tiempo sobre el que se realiza el análisis, la función adquiere un aspecto más picudo, hasta que para un tiempo infinito desaparecería el derrame espectral. Esto responde a la idea intuitiva de que a mayor intervalo de tiempo mayor precisión en la frecuencia y viceversa.

Podría parecer que una forma de minimizar el problema del derrame espectral consistiría en aumentar el tamaño del intervalo de análisis (la ventana de observación). Pero una ventana larga ocasiona una menor definición temporal, pues acontecimientos distintos en el tiempo son tratados como simultáneos, lo que en el caso de las señales musicales puede suponer una importante pérdida de información⁷. En efecto, la aplicación del análisis espectral a la música conlleva

7. También es posible elegir otras funciones distintas de la rectangular para la ventana de observación, funciones cuya transformada de Fourier presente una menor dispersión de la lobulización, atenuando los problemas derivados de los lóbulos laterales. No obstante, toda función que aminora la dispersión de la lobulización lo hace a costa de concentrarla sobre el lóbulo central, con lo que disminuye la capacidad de discriminar componentes próximos. Por ello, la elección del tipo de ventana más oportuno dependerá de la naturaleza del problema. El artículo de HARRIS, F. J. "On the Use of Windows

va exigencias mutuamente contradictorias, pues requiere simultanear una alta resolución en el tiempo y en la frecuencia. Por lo que respecta al tiempo, las señales musicales son estacionarias sólo de un modo aproximado y para un intervalo de tiempo muy breve. En una pieza musical las notas pueden sucederse a gran velocidad (en un pasaje rápido se pueden dar más de doce notas por segundo). Además, los parámetros sonoros de las notas pueden evolucionar con mucha rapidez, tanto por la propia acústica de los instrumentos como por la manipulación expresiva del ejecutante. En lo que concierne a la frecuencia, la música muchas veces concentra un gran número de componentes en bandas reducidas del espectro. Y diferencias de muy pocos hertzios pueden ser significativas, especialmente en los graves, debido al comportamiento logarítmico de nuestro sistema auditivo (por ejemplo, entre la_2 , 110 Hz, y sib_2 , 116,5 Hz, hay solamente 6,5 Hz, diferencia que se va duplicando conforme ascendemos de octava). Así pues, es necesario encontrar una solución de compromiso entre tiempo y frecuencia, y elegir el tamaño de la ventana de observación en función de las peculiaridades del fragmento musical y del objetivo que se pretenda con el análisis. Junto a ello, hay que tener en cuenta las posibles perturbaciones que se han podido producir entre los componentes del fragmento musical analizado, especialmente a la hora de realizar las mediciones⁸.

* * *

A lo largo de este artículo hemos podido comprobar que la espectrografía, incluso con los problemas que acabamos de señalar, puede ser una herramienta útil para el estudio sistemático de la interpretación musical. La espectrografía nos permite “ver” lo que oímos, transformar en representación espacial el discurrir temporal de la ejecución musical, obteniendo así una especie de partitura icónica de la interpretación (el espectrograma puede ser para el análisis de la interpretación lo que para el análisis de la obra es la partitura). Aporta, además, la posibilidad de establecer medidas, de cuantificar los parámetros sonoros de una ejecución musical, y de esa manera objetivar sus rasgos estilísticos y expresivos. No obstante, la aplicación del análisis espectral a la música no es una ciencia exacta, más bien se trata de un “arte” que requiere cierta pericia en su manejo.

for Harmonic Analysis with the Discrete Fourier Transform” *Proceedings of the IEEE* 66-1 (1978), pp. 51-83, es la referencia clásica para una discusión completa sobre la utilización de funciones de ventana en el análisis espectral.

8. Existen otras técnicas alternativas a la transformada de Fourier para realizar el análisis espectral. Alguna de ellas pueda dar resultados mejores ante determinadas situaciones concretas; sin embargo, no hay una técnica única, válida para cualquier situación, que proporcione a la vez una buena resolución y una fiabilidad en sus resultados. Excede al propósito de este trabajo una discusión de sus posibles aplicaciones al análisis de la interpretación musical. Para una exposición detallada del conjunto de las técnicas de análisis espectral véase el libro de MARPLE, S. L., *Digital Spectral Analysis*, Prentice Hall, New York, 1987.

Bibliografía

- ALLEN, J. B., "Short-term Spectral Analysis, Synthesis, and Modification by the Discrete Fourier Transform", *IEEE Transactions on Acoustics, Speech, and Signal Processing* 25(3), 1977, pp. 235-238.
- ALLEN, J. B. y RABINER, L. R., "A Unified Approach to Short-time Fourier Analysis and Synthesis", *Proceedings of the IEEE* 65(11), 1977, 1558-1564.
- HARRIS, F. J., "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform", *Proceedings of the IEEE* 66(1), 1978, pp. 51-83.
- KAY, S. M. y S. L. MARPLE, "Spectrum Analysis A Modern Perspective", *Proceedings of the IEEE* 69 (11), 1981, pp. 1380-1419.
- MARPLE, S. L., *Digital Spectral Analysis*, Prentice Hall, New York, 1987.
- SERRA, X. y SMITH, J., "Spectral Modelling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic plus Stochastic Decomposition", *Computer Music Journal* 14 (4), 1990, pp. 12-24.
- ROADS, C., *The Computer Music Tutorial*, The MIT Press, Cambridge, Massachusetts, 1996.
- TEMPELAARS, S., *Signal Processing, Speech and Music*, (Studies on New Music Research), Swets & Zeitlinger, Lisse, Holanda, 1996.
- ROADS, C., POPE, S. T.; PICCIALI, A. (eds.), *Musical Signal Processing*, (Studies on New Music Research), Swets & Zeitlinger, Lisse, Holanda, 1997.