(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 312 772**

(51) Int. Cl.:
G10L 21/00 (2006.01)
G10L 17/00 (2006.01)
H04K 1/00 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Número de solicitud europea: **03724113 .0**
(96) Fecha de presentación : **18.04.2003**
(97) Número de publicación de la solicitud: **1504445**
(97) Fecha de publicación de la solicitud: **09.02.2005**

(54) Título: **Equivalencia sólida e invariante de patrón de audio.**

(30) Prioridad: **25.04.2002 US 376055 P**

(45) Fecha de publicación de la mención BOPI:
01.03.2009

(45) Fecha de la publicación del folleto de la patente:
01.03.2009

(73) Titular/es: **Landmark Digital Services L.L.C.**
10 Music Square
East Nashville, Tennessee 37203, US

(72) Inventor/es: **Wang, Avery Li-Chun y**
Culbert, Daniel

(74) Agente: **Izquierdo Faces, José**

DESCRIPCIÓN

Equivalencia sólida e invariante de patrón de audio.

Campo de la invención

Esta invención se refiere generalmente al procesamiento de una señal de sonido en una gran base de datos de archivos de sonido o audio. Más en particular, se refiere a una técnica inventiva para determinar de manera rápida y precisa si dos muestras de audio coinciden, y es inmune a varias transformaciones entre las que se incluyen la variación de velocidad en una grabación. La técnica inventiva además permite una estimación precisa de las transformaciones.

Descripción de la técnica del contexto

La necesidad de un reconocimiento automático rápido y preciso para la música y otras señales de sonido continúa en crecimiento. La tecnología de reconocimiento de sonido previamente disponible a menudo sacrificaba la velocidad por la precisión o la inmunidad sonora. En algunas aplicaciones, era necesario calcular una regresión para calcular la pendiente de tiempo-tiempo dispersión-trazado en la presencia de ruido extremo, lo que introdujo un número de dificultades y disminuyó la actuación tanto de la velocidad como de la precisión. Por lo tanto, técnicas de reconocimiento de sonido que ya existían previamente eran incapaces de llevar a cabo un reconocimiento rápido y preciso en presencia de una significativa variación de velocidad en una grabación, por ejemplo, en el reconociendo de una grabación que está funcionando a una velocidad más rápida que la normal.

Además de la complejidad del problema existe un tipo de variación de velocidad cada vez más popular, la variación de tempo de corrección de tono, usada por los DJs en las emisoras de radio, clubs y en más sitios. En la actualidad, existe una técnica sólida y fiable que puede realizar un reconocimiento de sonido rápido y preciso a pesar de las variaciones de velocidad en la grabación y/o las variaciones de tempo de corrección de tono.

WO02/11123 describe un método para comparar dos muestras de audio o sonido igualando las impresiones digitales sonoras determinadas a partir de cada muestra de audio. Las impresiones digitales computadas son invariables en la extensión de tiempo y sus localizaciones relativas se emplean para determinar estadísticamente una compensación de tiempo entre dos muestras de audio.

"Musical Database Retrieval Based on Spectral Similarity" por Cheng Yang publicado como un Informe Técnico en Grupo sobre Bases de Datos en la Universidad de Stanford en 2001 describe la comparación de dos fragmentos de audio de respectivos vectores espectrales de audio. El método emplea una línea ajustada a puntos que representan ocurrencia en el tiempo de vectores correspondientes en ambos fragmentos de audio, con el fin de representar la similitud del archivo.

Resumen de la invención

La presente invención cumple la necesidad en la técnica de reconocimiento de audio proporcionando un método rápido e invariable para caracterizar la relación entre dos archivos de audio. El método inventivo es preciso incluso en presencia de ruido extremo, superando los inconvenientes mencionados de la tecnología existente.

De acuerdo con un aspecto de la invención, en lo sucesivo se proporciona un método de acuerdo con la reivindicación 1.

De acuerdo con otro aspecto de la invención, la técnica arriba descrita puede además mejorarse proporcionando un cálculo aproximado de valor relativo global con una localización de un pico en un eje del histograma. El valor relativo global, a su vez, puede perfeccionarse mediante una primera selección de una zona alrededor del pico de interés.

Incluso en otra realización, en la cual el valor de velocidad de grabación relativa se determina a partir de un pico del histograma, se calcula un valor de compensación relativo de tiempo para cada par de objetos con impresión digital equivalente. Se genera otro histograma en base a los valores de compensación relativos de tiempo. Si se encuentra un punto estadísticamente significativo en el segundo histograma, la relación entre las dos muestras de audio puede además caracterizarse por el pico, proporcionando además una mejora en la precisión de la invención.

Breve descripción de las figuras

Fig. 1 es una representación de un espectrograma de una muestra de audio analizada.

Fig. 2 es un diagrama ejemplar que muestra objetos con impresiones digitales que se generan a partir de una muestra de audio de acuerdo con un aspecto de la invención.

Fig. 3 ilustra dos muestras de audio que se comparan de acuerdo con los principios de la presente invención.

Figs. 4A-B muestran espectrogramas ejemplares con o sin pico estadísticamente significativo.

Figs. 5A-B muestran el movimiento de puntos frecuencia-tiempo cuando la velocidad de la grabación varía.

Figs. 6A-B muestran los tiempos correspondientes en una primera muestra de audio (sonido de muestra) y una segunda muestra de audio (sonido de base de datos) de detalles numerales correspondientes.

Figs. 7A-D ilustran pendientes rápidas y eficientes y técnicas de histogramas de la presente invención.

Descripción detallada

La presente invención permite indexar y buscar de manera rápida, sólida, invariable y en escalas una gran base de datos de archivos de audio y en particular es útil para aplicaciones en reconocimiento de patrones de audio.

Una operación de comparación rápida y eficiente entre dos archivos de muestra de audio es esencial en la construcción de un sistema de reconocimiento de audio comercialmente viable. De acuerdo con un aspecto de la invención, la relación entre dos muestras de audio puede caracterizarse en primer lugar por ciertas impresiones digitales de objetos que coinciden y que derivan de un espectrograma, como el mostrado en la Figura 1, de las respectivas muestras de audio. El espectrograma es una representación/análisis de tiempo-frecuencia que se genera tomando muestras 2*K en un punto en el tiempo en un marco de ventana corrediza y calculando una Transformación de Fourier, generando de este modo cajas de frecuencia K en cada marco. Los marcos pueden coincidir o montarse para mejorar la resolución temporal del análisis. Los parámetros particulares empleados dependen del tipo de muestras de audio que se están procesando. Preferiblemente se emplean archivos de audio con tiempo discreto con un rango de muestreo de 8 kilohercios, marcos con K=512, y un ritmo de 64 muestras.

Objetos con Impresiones Digitales

Después de generarse un espectrograma de cada muestra de audio, se escanea para buscar características locales, es decir, picos de energía local, tal y como se muestra en la Fig. 2. El proceso de comparación comienza extrayendo un conjunto de impresiones digitales de las correspondientes características locales de cada muestra de audio. En una realización ejemplar, una muestra de audio es una muestra de sonido desconocido que va a reconocerse y la otra muestra de audio es una grabación conocida almacenada en una base de datos. Cada objeto con impresión digital ocurre en una localización particular en la respectiva muestra de audio. En algunas realizaciones, cada objeto con impresión digital se localiza en alguna compensación de tiempo en un archivo de audio y contiene un conjunto de datos descriptivos sobre el archivo de audio junto con su respectivo coordinado temporal. Es decir, la información descriptiva contenida en cada objeto con impresión digital se calcula con dependencia de la muestra de audio cerca de la correspondiente compensación de tiempo. Esto se codifica en una pequeña estructura de datos. Preferentemente, la localización y la información descriptiva se determinan de modo que sean generalmente reproducibles incluso en el caso de presencia de ruido, distorsión, y otras transformaciones tales como variación en la velocidad de grabación. En este caso, la localización se determina dependiendo del contenido de la respectiva muestra de audio y cada objeto con impresión digital caracteriza una o más características locales de la respectiva muestra de audio en o cerca de la respectiva localización particular, por ejemplo, localización (t1, f1) o (t2, f2) tal y como se muestra en la Fig. 1.

En una realización ejemplar, cada objeto con impresión digital se caracteriza por su localización, un componente variante y un componente invariante. Cada característica local es un pico en el espectrograma y cada valor de frecuencia se determina a partir de una coordenada de frecuencia de un correspondiente pico en el espectrograma. Los picos se determinan buscando en las inmediaciones de cada coordenada tiempo-frecuencia y seleccionando los puntos que tienen mayor valor en magnitud en comparación con sus vecinos. Más específicamente, tal y como se muestra en la Fig. 2, se analiza una muestra de audio 210 en una representación de espectrograma 220 con regiones 221 y 222 de elevada energía demostrada. La información relativa a las regiones locales de energía 221 y 222 se extrae y resume en una lista 230 de objetos con impresiones digitales 231, 232, etc. Cada objeto con impresión digital incluye opcionalmente un campo de localización 242, un componente variante 252, y un componente invariante 262. Preferentemente, se selecciona una zona de tal modo que cada punto elegido sea el máximo en una unidad de bloque 21x21 en el centro de la misma. A continuación, se determina un valor relativo para cada par de objetos con impresión digital igualada. En algunas realizaciones, el valor relativo es un cociente o diferencia de logaritmo de valores paramétricos de las respectivas muestras de audio. A continuación se genera un histograma de los valores relativos. Si se encuentra un pico estadísticamente significativo en el histograma, entonces las dos muestras de audio pueden calificarse de sustancialmente iguales.

En referencia a la Fig. 3, las listas de objetos con impresiones digitales 310 y 320 se preparan respectivamente tal y como se ha descrito anteriormente para las muestras de audio 1 y 2, respectivamente. Se comparan los respectivos objetos con impresiones digitales 311 y 322 de cada lista. Los objetos con impresiones digitales que coinciden se emparejan, por ejemplo, empleando los respectivos componentes invariantes Inv e Inv' en el paso 351, y colocándolos en una lista en el paso 352. A continuación, en el paso 354, se genera un histograma de valores relativos. En el histograma se busca un pico estadísticamente significativo en el paso 355. Si no se encuentra ninguno en el paso 356, las muestras de audio 1 y 2 no coinciden, por ejemplo, el histograma 410 de la Fig. 4A. De manera alternativa, si se detecta un pico estadísticamente significativo, las muestras 1 y 2 coinciden, por ejemplo, el histograma 420 de la Fig. 4B.

La técnica que acaba de ser descrita puede además mejorar proporcionando un cálculo aproximado de un valor relativo global R con una localización en el pico del eje del histograma, tal y como se ilustra en el paso 361. En algunas realizaciones, R puede perfeccionarse mediante una primera selección de una zona alrededor del pico de interés. En la Fig. 1, esto se muestra como un área de interés 110 alrededor de una localización particular (t1, f1). A continuación, se calcula un promedio de los valores relativos en la zona seleccionada. El promedio puede ser una media ponderada con números de puntos en cada valor relativo en la zona seleccionada. En algunas realizaciones, R puede además perfeccionarse para generar un valor relativo de compensación de tiempo t'-R*t par cada pareja que coincide. Los pasos 362-364 muestran que, con estos valores relativos de compensación de tiempo, se genera un segundo histograma, permitiendo que se calcule un tiempo compensado.

Otros tipos de análisis tiempo-frecuencia pueden implementarse para extraer objetos con impresiones digitales, por ejemplo, la distribución Wigner-Wille o wavelets (ondas pequeñas). Así mismo, en lugar de picos en espectrograma, se pueden emplear otras características, por ejemplo coeficientes cepstrales. Además, las técnicas de súper-resolución pueden emplearse para obtener cálculos aproximados más precisos de tiempo y frecuencia de las coordenadas tiempo-frecuencia provistas por los picos del espectrograma. Por ejemplo, la interpolación parabólica en cubos o cajas de frecuencia podría usarse para aumentar la resolución de frecuencia. Pueden encontrarse descripciones ejemplares relacionadas en "PARSHL: Un Programa de Análisis/Síntesis para Sonidos No Armónicos en Base a una Representación Sinusoidal", Julius O. Smith II y Xavier Serra, Procedimientos de la Conferencia Internacional de Música en Ordenador (ICMC-87, Tokio), Asociación de Música en Ordenador, 1987, y en "Estimación Moderna Espectral: Teoría y Aplicación" por Steven M. Kay (Enero 1988) Prentice may, ambos aquí incorporados como referencias.

Equivalencia

En una operación de equivalencia, se comparan dos muestras de audio por medio de sus respectivos objetos con impresión digital. Tal y como se ha descrito anteriormente con referencia a la Fig. 3, se generan pares de objetos con impresiones digitales iguales, conteniendo cada par componentes que sustancialmente coinciden. Un modo de preparar los datos para permitir una rápida búsqueda es codificar los objetos con impresiones digitales en fichas numéricas, como números enteros sin firmar de 32 bits, y empleando fichas numéricas como una clave para clasificar y buscar. Las técnicas para una eficiente manipulación de datos son bien conocidas en el campo, por ejemplo "Art of Computer Programming, Volume 3: Sorting and Searching (2ª Edición)", por Donal Ervin Knuth (Abril 1998) Addison-Wesley, que aquí se incorpora como referencia.

En una realización ejemplar, cada objeto con impresión digital contiene un componente invariante y un componente variante. El componente variante hace referencia a los valores de radios de frecuencia correspondientes a los picos espectrales, así como los valores de radios de tiempo delta (es decir, la diferencia temporal) entre los picos espectrales son invariantes bajo el periodo de tiempo. Por ejemplo, en referencia a la Fig. 5A y 5B, si un espectrograma de una muestra de audio tiene algunos picos locales espectrales con coordenadas (t1, f1), (t2, f2), y (t3, f3), entonces el invariante para dos puntos es f2/f2, es decir f2'/f1'=f2/f1. Se dan invariantes adicionales para 3 puntos mediante f2/f1, (t3-t1)/(t2-t1), o (t3-t2)/(t2/t1), o cualquier otra combinación creada cambiando los puntos y/o funciones de computación de estas cantidades o combinaciones de estas cantidades. Por ejemplo, f2/f3 podría crearse dividiendo f2/f1 por f3/f1. Además, si la muestra de audio se extiende linealmente, simplemente reproduciéndola más rápido, de manera adicional la frecuencia y el tiempo delta experimentan una relación recíproca para que las cantidades como f1*(t2-t1) sean también invariantes. Pueden emplearse logaritmos de estas cantidades, sustituyendo la suma y la resta por la multiplicación y I a división. Para descubrir los radios de frecuencia y de extensión temporal, asumiendo que sean independientes, es necesario tener una cantidad de variante de frecuencia y de variante de tiempo.

Para realizar la equivalencia de manera más eficiente, empleamos la parte invariante para crear el índice de impresiones digitales y usamos valores próximos o exactos para la búsqueda. Realizar la búsqueda usando equivalencias próximas permite una solidez adicional contra la distorsión y el error concluyente, pero implica mayor coste si la búsqueda en componentes invariantes se vuelve una búsqueda de ámbito tridimensional. En la realización preferente, se precisa que el componente invariantes de los respectivos objetos con impresiones digitales se ajuste exactamente, dando lugar por lo tanto a un sistema que es muy rápido, con una menor compensación contra la sensibilidad de reconocimiento en presencia de ruido. Es importante señalar que este método funciona bien incluso si solamente una minoría de objetos con impresiones digitales en las correspondientes muestras de audio coincide correctamente. En el paso de detección de pico en el histograma, un pico puede ser estáticamente significativo incluso si tan sólo el 1-2% de los objetos con impresiones digitales coinciden correctamente y sobreviven.

El componente variante también puede emplearse para limitar el número de objetos con impresiones digitales equivalentes, además de o en lugar del componente variante. Por ejemplo, podríamos necesitar que un componente variante V de la primera muestra de audio coincidiera con un correspondiente V" de la segunda muestra de audio en un +/- 20%. En ese caso, podemos formar una representación de las fichas numéricas de tal modo que la parte superior (por ejemplo, los bits más significativos) contenga los componentes invariantes y la parte inferior (por ejemplo, los bits menos significativos) contenga los componentes variantes. Así, buscar una equivalencia aproximada se convierte en una búsqueda de ámbito sobre las fichas compuestas usando los valores más bajos y más altos del componente variante. El uso de un componente invariante en la equivalencia no es por lo tanto estrictamente necesario si la búsqueda se realiza usando un componente variante. Sin embargo, el uso de un componente invariante en el proceso de equivalencia es recomendado ya que ayuda a reducir el número de falsas coincidencias o parejas, por lo que hace más eficiente el proceso de realización de histogramas y reduce la cantidad de procesos generales.

Por otra parte, el propio componente variante nuevo puede ser o no parte del criterio de equivalencia entre dos objetos con impresiones digitales. El componente variante representa un valor que puede distorsionarse por una simple transformación paramétrica que va desde una grabación original a una grabación de muestra. Por ejemplo, los componentes variantes de frecuencia, como f1, f2, f3, y los componentes variantes de tiempo como (t2-t1), (t3-t1), o (t3-t2) pueden elegirse como componentes variantes para variación de velocidad en una grabación. En el caso de que haya una segunda muestra de audio, en la interpretación de equivalencia en una base de datos, con un espectrograma que incluye coordenadas (t1', f1'), (t2', f2') y (t3', f3'), correspondientes a los mismos puntos listados anteriormente para la primera muestra de audio. A continuación, el componente de frecuencia f1' podría tener un valor escalado $f1'=R_f*f1$, donde R_f es un parámetro de extensión lineal que describe lo rápido o despacio que se reproduce la primera muestra en comparación con la segunda. El componente variante de cada una de las dos muestras de audio equivalentes puede emplearse para calcular una aproximación del valor de extensión global, lo que describe un parámetro macroscópico, calculando el radio entre los dos valores de frecuencia, $R_f=2$ significa que la primera muestra de audio tiene la mitad de tono (frecuencia) de la segunda. Otra posibilidad es usar $R_t=(t2'-t1')/(t2-t1)$. En este caso, el valor relativo R es el radio de velocidad relativa de grabación, es decir, $R_t=2$ significa que la primera muestra de audio se reproduce el doble de rápido que la segunda muestra de audio.

Si $R_f=1/R_t$, es decir, $f'/f'=(t2-t1)/(t2'-t1')$, entonces las dos muestras de audio están relacionadas por una extensión lineal de tiempo debido a la relación recíproca tiempo-frecuencia para dichas muestras de audio. En este caso, podemos usar en primer lugar el método del histograma aquí descrito para formar un R_f aproximado del radio relativo de frecuencia relativa usando los correspondientes componentes variantes de frecuencia, y de nuevo para formar un R_t aproximado de la velocidad relativa de grabación, y llevar a cabo a continuación una comparación para detectar si la relación de la grabación es lineal o no lineal.

En general, el valor relativo se calcula a partir de objetos con impresiones digitales equivalentes usando los correspondientes componentes variantes de la primera y segunda muestra de audio. El valor relativo podría ser un simple radio de frecuencias o tiempos delta, o cualquier otra función que dé como resultado un cálculo aproximado un parámetro global empleado para describir el trazado entre la primera y la segunda muestra de audio. Pero en general, puede emplearse cualquier función con 2 entradas F ( ), por ejemplo, R=F (vl,vl'), donde vl y vl' son respectivas cantidades variantes. Es mejor si F ( ) es una función continua para que ocurran pequeños errores en la medida de vl y vl' en la salida R.

Histogramas

Tal y como aquí se describe, se genera un histograma sobre un conjunto de valores relativos calculados a partir de una lista de parejas equivalentes de objetos con impresiones digitales. A continuación se busca un histograma para un pico. La presencia de un pico estadísticamente significativo en el histograma indica que se ha dado una posible coincidencia o equivalencia. En particular, este método busca en el histograma un grupo valores relativos en lugar de diferencias de compensaciones de tiempo, como (t1'-t1). De acuerdo con un principio de la presente invención, un histograma sirve para formar cubos de valores totales, correspondiendo cada cubo a un valor particular a lo largo del eje independiente del histograma. Para fines de la invención, generar un histograma puede llevarse a cabo simplemente clasificando la lista de valores relativos. Por lo tanto, un modo rápido y eficaz de detectar el pico de un histograma de una lista de valores es clasificar la lista en orden ascendente y escanear a continuación el mayor grupo de unidades que tengan valores iguales o similares.

Significado Estadístico

Tal y como se ha establecido anteriormente, con la presente invención, dos muestras de audio pueden coincidir correctamente incluso si solamente el 2% de los objetos con impresiones digitales sobrevive a todas las distorsiones y coinciden correctamente. Esto es posible anotando la comparación entre las dos muestras de audio. En concreto, se elige una zona alrededor del pico del histograma y se cuentan todas las parejas equivalentes que caen en la zona, dando como resultado una puntuación. Además, puede calcularse una puntuación ponderada descontando la contribución de parejas que son de puntos más lejanos al centro del pico.

Un modo de calcular el criterio límite es asumir que la probabilidad de distribución del resultado de una ruta no equivalente cae con una cola exponencial. El modelo se aplica a la distribución real medida de resultados de rutas no equivalentes. A continuación se calcula la distribución cumulativa de probabilidad del resultado más elevado sobre una base de datos de rutas N (por ejemplo, tomado como la potencia Nth de la distribución cumulativa de probabilidad de un único resultado no equivalente). Una vez que se conoce la curva de probabilidad y se elige un nivel máximo de positivos falsos (por ejemplo, 0.5%), puede elegirse el umbral numérico y emplearse para determinar si el pico del histograma tiene un número estadísticamente significativo de parejas equivalentes.

Estimación Hiperfina

Una vez que se encuentra un pico en el histograma estadísticamente significativo, puede calcularse una estimación "hiperfina" de elevada resolución del valor relativo global (como la velocidad relativa de grabación). Esto se lleva a cabo eligiendo una zona alrededor del pico, por ejemplo, incluyendo un intervalo de aproximadamente 3 a 5 cubos centrados en el histograma del pico, y calculando una media de los valores relativos en la zona. Empleando esta técnica, podemos encontrar velocidad relativa de grabación exacta en un 0.05%. Con la derivación de compensación

aquí descrita la compensación global de tiempo puede calcularse con una precisión mejor que un milisegundo, lo que es más preciso que la resolución de tiempo de los marcos de espectrograma mencionados.

Regresión Sólida

En el caso de que las muestras realmente coincidan, puede verse una línea diagonal en una zona de dispersión donde las muestras equivalentes tiene las correspondientes coordenadas de tiempo (t',t) de objetos con impresiones digitales trazadas o marcadas una contra la otra, tal y como se muestra en la Fig. 6A. El objetivo es encontrar la ecuación del elemento de la regresión, que se determina mediante la pendiente y la compensación de la línea en presencia de una elevada cantidad de ruido. La pendiente indica la velocidad relativa de grabación, y la compensación es la compensación relativa desde el inicio de una muestra de audio hasta el inicio de la segunda. Técnicas convencionales de regresión, como mínimos cuadrados ponderados, se encuentran disponibles por ejemplo en "Numerical Recipes in C: The Art of Scientific Computig (2ª edición)", por William H. Press, Brian P. Flannery, Saul A. Teukolsky, y William T. Vetterling (Enero 1993), Cambridge Universtiy Press, que aquí se incorpora como referencia. Desafortunadamente, estas técnicas convencionales sufren sensibilidad desproporcionada, donde un punto demasiado alejado puede desviar drásticamente los parámetros estimados de regresión. En la práctica, los puntos a menudo están dominados por un punto más alejado, provocando que sea muy complicado detectar la correcta línea diagonal. Pueden emplearse otras técnicas para regresión sólida para superar el problema del punto más alejado y para encontrar una relación lineal entre los puntos en presencia de ruido, pero estas técnicas tienden a ser lenta e iterativas y existe la posibilidad de quedarse atascadas en un punto óptimo local. Existe una amplia variedad de técnicas en la bibliografía para buscar un elemento de regresión lineal no conocido. El kit de herramientas Matlab, disponible en Mathsworks, contiene una variedad de rutinas software para análisis de regresión.

La presente invención proporciona un método inventivo para calcular la velocidad relativa de una grabación (o, equivalentemente, la reciprocidad del tono relativo, en el caso de una relación de grabación lineal) que soluciona el problema de buscar una línea de regresión en el trazado de dispersión tiempo-tiempo incluso si la pendiente de la pareja no se iguala, por ejemplo, Fig. 6B. El uso de un histograma de velocidades relativas locales de grabación, tal y como aquí se establece, tiene la ventaja de contar con la información no considerada previamente y proporciona una ventaja inesperada al solucionar rápida y eficientemente el problema de regresión.

Para encontrar la compensación, se asume que los correspondientes puntos en el tiempo tienen la relación

$$\text{compensación} = t1' - R_t * t1,$$

donde R_t se obtiene tal y como se ha descrito anteriormente. Esto es la compensación de tiempo y sirve para normalizar los sistemas de coordenada de tiempo entre dos muestras de audio. Esto también puede verse como una transformación de corte en el trazado de dispersión tiempo-tiempo que provoca que la línea diagonal de la pendiente no conocida de la Fig. 7A sea vertical en la Fig. 7C. El histograma 720 de la Fig. 7B muestra un pico de radios de velocidad relativa acumulada de grabación que indican el radio global relativo de velocidad de grabación R. A continuación se dan nuevos valores mediante la fórmula de compensación, y se genera un nuevo histograma 740 tal y como se observa en la Fig. 7D. El pico del nuevo histograma 740 ofrece un cálculo aproximado de la compensación global, lo que puede pulirse mediante el uso de un promedio de los valores en la zona del pico, tal y como se ha descrito anteriormente.

En resumen, la primera fase de realización de histograma proporciona un modo para calcular la velocidad relativa de grabación, y así mismo determina si existe alguna equivalencia. La segunda fase de realización de histograma asegura que las muestras de audio candidatas a equivalencia tengan un número significativo de objetos con impresiones digitales que se alinean también de manera temporal. La segunda fase de realización de histograma también sirve como un segundo criterio de análisis independiente y ayuda a disminuir la probabilidad de positivos falsos, proporcionando de este modo un criterio más fuerte y sólido para decidir si las dos muestras de audio coinciden. La segunda fase de la realización de histograma puede realizarse de manera opcional solamente si existe un pico estadísticamente significativo en el primer histograma, ahorrando de este modo recursos y esfuerzos computacionales. Opcionalmente puede realizarse una optimización adicional, por ejemplo, para reducir el grupo computacional, en vez de computar el segundo histograma sobre todas las parejas de objetos con impresiones digitales equivalente en la lista, el segundo histograma puede generarse empleando solamente las parejas equivalentes correspondientes al pico del primer histograma.

Sincronización de Grabaciones Múltiples

La presente invención también puede implementarse para introducir y alinear el tiempo de grabaciones de audio no sincronizadas. Por ejemplo, suponemos que un grabador DAT y un grabador de cintas funcionaron independientemente con diferentes micrófonos en localizaciones o ambientes ligeramente diferentes. Si más tarde se desea combinar las dos grabaciones a partir de las respectivas grabaciones en una mezcla, las dos rutas pueden sincronizarse usando la técnica sólida de regresión aquí descrita para obtener la compensación temporal. De este modo, incluso si las grabaciones no sincronizadas operan a velocidades ligeramente diferentes, la velocidad relativa puede determinarse con un elevado grado de precisión, permitiendo que una grabación se compense con respecto a la otra. Este hecho

resulta especialmente útil si se descubre que una de las grabaciones se ha corrompido y necesita complementarse con otra fuente. Por lo tanto, la alineación temporal y la sincronización tal y como aquí se describen permiten una mezcla transparente.

Búsqueda de Bases de Datos

Debido a que el método de comparación es extremadamente rápido, es posible pre-procesar una gran base de datos de muestras de audio en respectivas listas de objetos con impresiones digitales. Tal y como el experto en la técnica apreciará, una muestra de audio no conocida puede por lo tanto pre-procesarse en su propia lista respectiva de objetos con impresiones digitales usando las técnicas de procesamiento de datos disponibles. Las técnicas arriba mencionadas sobre equivalencia, histogramas, y detección de pico pueden llevarse a cabo a continuación empleando los objetos con impresiones digitales pre-tratados en la base de datos para encontrar una equivalencia.

A pesar de que la presente invención y sus ventajas han sido descritas con detalle, debería entenderse que la presente invención no se limita o define por lo que aquí se muestra o establece. En particular, los dibujos y descripciones aquí adjuntas muestran tecnologías relacionadas con la invención, muestran ejemplos de la invención, y proporcionan ejemplos para usar la invención y no pretenden limitar la presente invención. Pueden establecerse métodos, técnicas y sistemas conocidos sin dar detalles, para evitar confundir los principios de la invención. Como un experto en la técnica apreciará, la presente invención puede implementarse, modificarse, o sino alterarse sin partir de los principios y espíritu de la presente invención. Por ejemplo, los métodos, técnicas y pasos aquí descritos pueden implementarse o sino realizarse de una forma mediante instrucciones ejecutables por un ordenador en un medio legible de ordenador. De manera alternativa, la presente invención puede implementarse en un sistema de ordenador que tenga un cliente y un servidor. El cliente envía la información, por ejemplo, objetos con impresiones digitales, necesaria para la caracterización de la relación entre la primera y la segunda muestra de audio al servidor donde se lleva a cabo la caracterización. Por consiguiente, el alcance de la invención debería determinarse por las siguientes reivindicaciones.

REIVINDICACIONES

1. Un método para caracterizar una relación de una primera y una segunda muestra de audio, que consiste en los siguientes pasos:

- generar un primer conjunto de objetos con impresiones digitales (310) para la primera muestra de audio, ocurriendo cada objeto con impresión digital en una respectiva localización en la primera muestra de audio, estando la respectiva localización determinada en dependencia del contenido de la primera muestra de audio, y estando caracterizado cada objeto con impresión digital por una o más características de la primera muestra de audio en o cerca de cada respectiva localización;

- generar una segundo conjunto de objetos con impresiones digitales (320) para la segunda muestra de audio, ocurriendo cada objeto con impresión digital en una respectiva localización en la segunda muestra de audio, estando la respectiva localización determinada en dependencia del contenido de la segunda muestra de audio, y estando caracterizado cada objeto con impresión digital por una o más características de la segunda muestra de audio en o cerca de cada respectiva localización;

- emparejar objetos con impresiones digitales (352) haciendo coincidir un primer objeto con impresión digital (311) de la primera muestra de audio con un segundo objeto con impresión digital (322) de la segunda muestra de audio que sea sustancialmente similar al primer objeto con impresión digital; donde cada objeto con impresión digital tiene un componente invariante (262) y un componente variante (252) en la localización, y el primer y segundo objeto con impresión digital en cada pareja equivalente de objetos con impresión digital tienen componentes invariantes que coinciden;

- generar, en base al paso de emparejamiento, una lista de parejas de objetos con impresiones digitales (352);

- determinar el valor relativo para cada pareja de objetos con impresión digital equivalente usando los componentes variantes (252),

- generar un histograma del valor relativo (354); y

- buscar un pico estadísticamente significativo en el histograma (355), caracterizando el pico la relación entre la primera y segunda muestra de audio que incluye un factor de elasticidad.

2. El método de acuerdo con la reivindicación 1 en el que la relación entre la primera y segunda muestra de audio se caracteriza por una sustancial coincidencia si se encuentra un pico estadísticamente significativo.

3. El método de acuerdo con la reivindicación 1 o 2, que además comprende el paso de calcular un valor global relativo con una localización de un pico en un eje del histograma, caracterizando además el valor global relativo la relación entre la primera y la segunda muestra de audio.

4. El método de acuerdo con la reivindicación 3, que además comprende el paso de determinar un cálculo aproximado hiperfino del valor global relativo, donde el paso de determinación comprende:

- seleccionar una zona alrededor del pico, y calcular un promedio de los valores relativos en la zona vecina.

5. El método de acuerdo con la reivindicación 1 en el que el componente invariante se genera usando:

(i) un radio entre el primer y el segundo valor de frecuencia, determinándose cada valor de frecuencia respectivamente a partir de una primera y una segunda característica local cerca de la respectiva localización de cada objeto con impresión digital;

(ii) un producto entre un valor de frecuencia y un valor de tiempo delta, determinándose el valor de frecuencia a partir de una primera característica local, y determinándose el valor de tiempo delta entre la primera característica local y una segunda característica local cerca de la respectiva localización de cada objeto con impresión digital; o

(iii) un radio entre un primer y un segundo valor de tiempo delta, determinándose el primer valor de tiempo delta a partir de una primera y una segunda característica local, determinándose el segundo valor de tiempo delta a partir de la primera y tercera característica local, estando cada característica local cerca de la respectiva localización de cada objeto con impresión digital.

6. El método de acuerdo con la reivindicación 5 en el cual cada característica local es un pico de espectrograma y cada valor de frecuencia se determina a partir de una coordenada de frecuencia de un correspondiente pico de espectrograma.

7. El método de acuerdo con la reivindicación 1, en el cual cada objeto con impresión digital tiene un componente variante, y el valor relativo de cada pareja de objetos con impresión digital equivalente se determina usando respectivos componentes variantes del primer y segundo objeto con impresión digital.

8. El método de acuerdo con la reivindicación 7 en el cual el componente variante es un valor de frecuencia determinado a partir de una característica local cerca de la respectiva localización de cada objeto con impresión digital de tal modo que el valor relativo de una pareja de objetos con impresión digital equivalente se califique de radio de respectivos valores de frecuencia del primer y segundo objeto con impresión digital y el pico en el histograma caracterice la relación entre la primera y segunda muestra de audio que se califican de un tono relativo, o, en el caso de extensión lineal, una velocidad relativa de grabación.

9. El método de acuerdo con la reivindicación 8, donde el radio de los respectivos valores de frecuencia se caracteriza por ser una división o una diferencia de logaritmos.

10. El método de acuerdo con la reivindicación 8, en el que cada característica local es un pico de espectrograma y cada valor de frecuencia se determina a partir de una coordenada de frecuencia de un correspondiente pico en el espectrograma.

11. El método de acuerdo con la reivindicación 7, en el que el componente variante es un valor temporal delta determinado a partir de una primera y una segunda característica local cerca de la respectiva localización de cada objeto con impresión digital de tal modo que el valor relativo de una pareja de objetos equivalentes se califique como el radio de respectivos valores temporales delta variantes y el pico en el histograma caracterice la relación entre la primera y la segunda muestra de audio que se caracterizan por una relativa velocidad de grabación, o, en el caso de expansión lineal, un tono relativo.

12. El método de acuerdo con la reivindicación 11, donde el radio de los respectivos valores temporales deltas variantes se caracteriza por ser una división o una diferencia de logaritmos.

13. El método de acuerdo con la reivindicación 11, en el cual cada característica local es un pico en el espectrograma y cada valor de frecuencia se determina a partir de una coordenada de frecuencia de un correspondiente pico en el espectrograma.

14. El método de acuerdo con la reivindicación 7, que además incluye los pasos de:

- determinar un tono relativo para la primera y segunda muestra de audio usando los respectivos componentes variantes, donde cada componente variante es un valor de frecuencia determinado a partir de una característica local cerca de la respectiva localización de cada objeto con impresión digital;

- determinar una velocidad relativa de grabación para la primera y segunda muestra de audio usando los respectivos componentes variantes, donde cada componente variante es un valor temporal delta determinado a partir de una primera y segunda característica local cerca de la respectiva localización de cada objeto con impresión digital; y

- detectar si el tono relativo y una reciprocidad de la velocidad relativa de grabación son sustancialmente diferentes, en cuyo caso la relación entre la primera y segunda muestra de audio se califica como no lineal.

15. El método de acuerdo con la reivindicación 1, donde R es un valor para la velocidad relativa de grabación determinada a partir del pico del histograma de los valores relativos, que además comprende los siguientes pasos:

- para cada pareja de objetos con impresiones digitales equivalentes en la lista, determinar un valor temporal relativo compensado, t-R*t', donde t y t' son localizaciones en el tiempo con respecto al primer y segundo objeto con impresión digital;

- generar un segundo histograma de los valores relativos de tiempo compensados; y

- buscar un pico estadísticamente significativo en el segundo histograma de los valores relativos de tiempo compensado, caracterizándose además el pico por la relación entre la primera y segunda muestra de audio.

16. Un programa de ordenador para llevar a cabo un método de acuerdo con cualquiera de las reivindicaciones precedentes.

17. Un sistema de ordenador que incluya medios para llevar a cabo cada paso de un método de acuerdo con las reivindicaciones 1 a 15, e incluyendo un cliente para enviar la información necesaria para la caracterización de la relación entre la primera y segunda muestra de audio a un servidor que realice la caracterización.

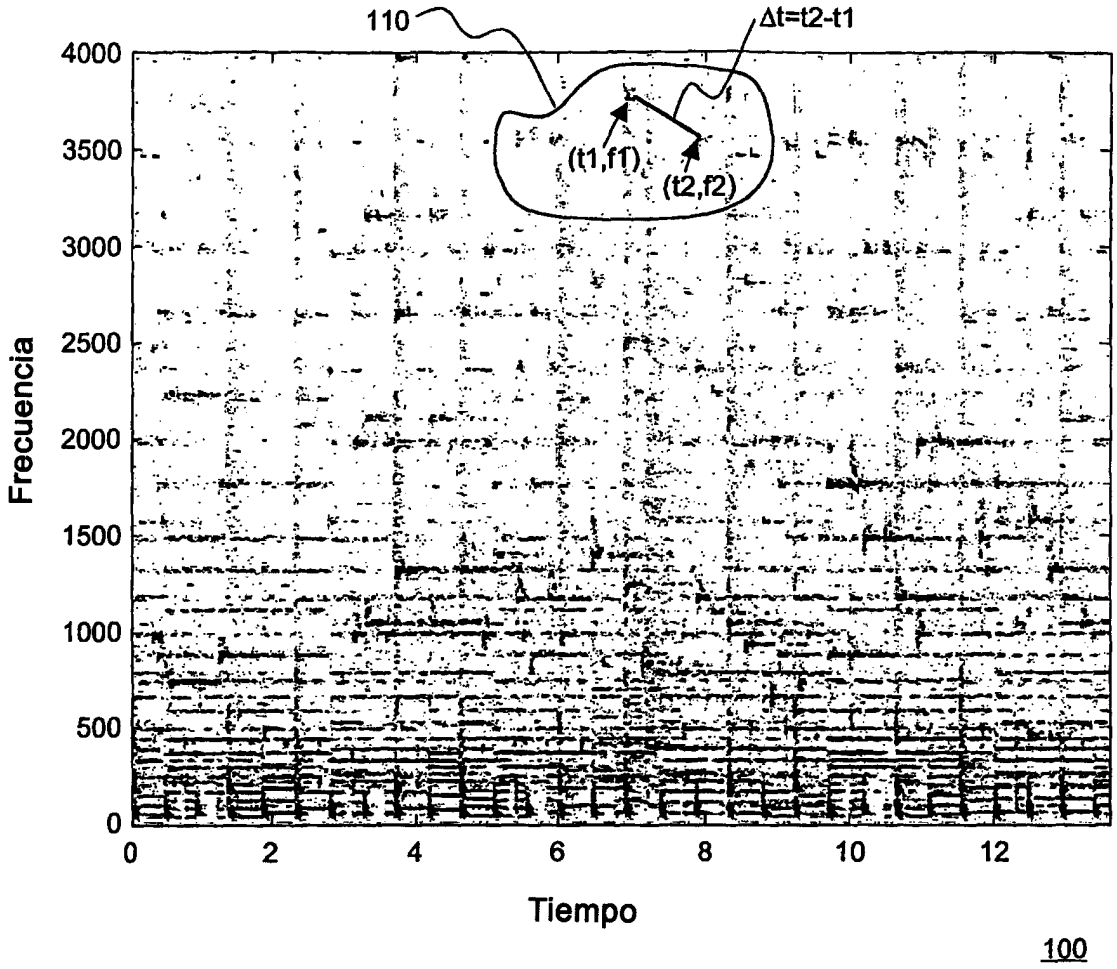
110
Δt=t2-t1
(t1,f1)
(t2,f2)
4000
3500
3000
2500
2000
1500
1000
500
0
Frecuencia
0
2
4
6
8
10
12
Tiempo
100

FIG. 1

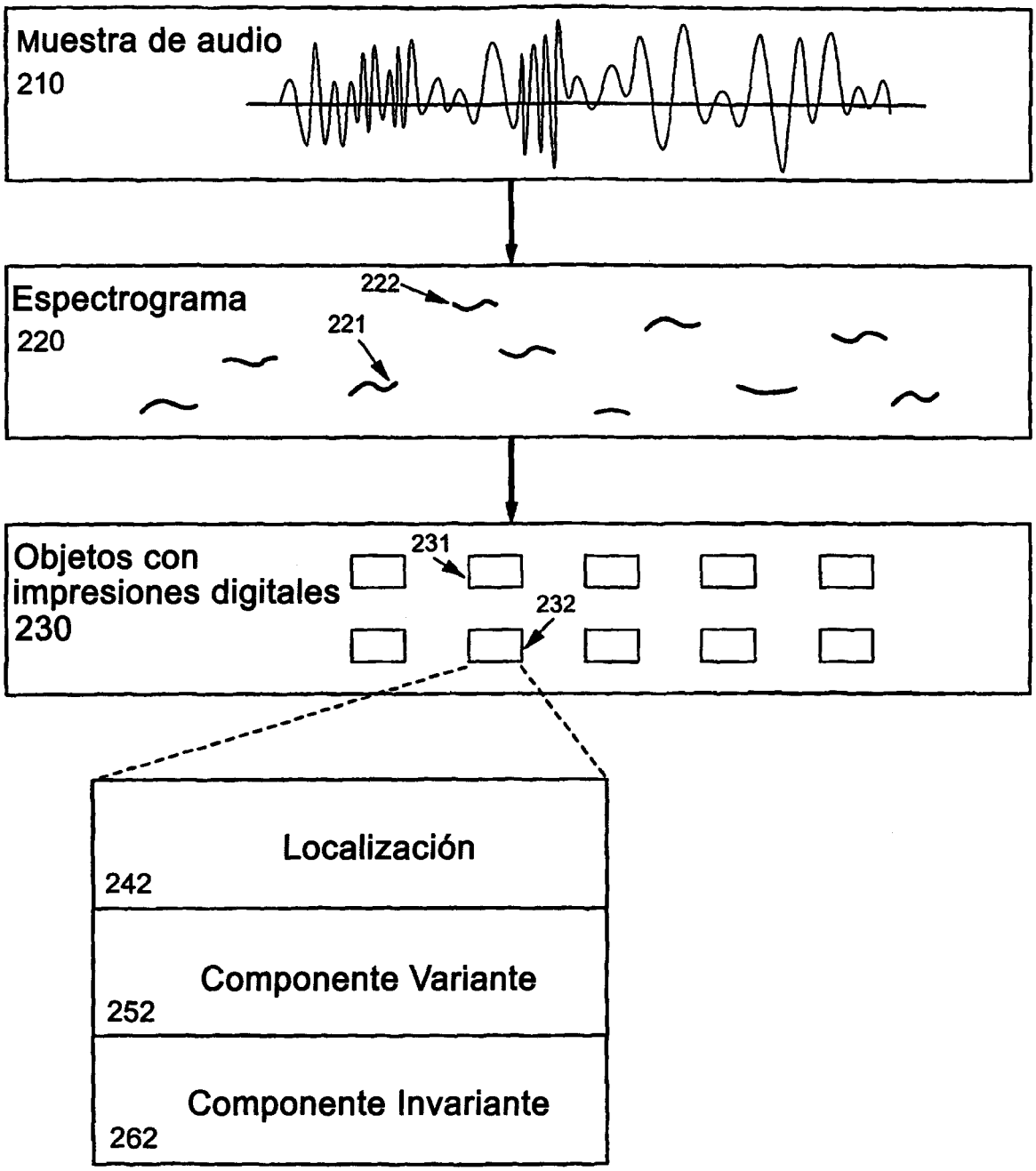
Muestra de audio
210
Espectrograma
220
222
221
Objetos con
impresiones digitales
230
231
232
Localización
242
Componente Variante
252
Componente Invariante
262

FIG. 2

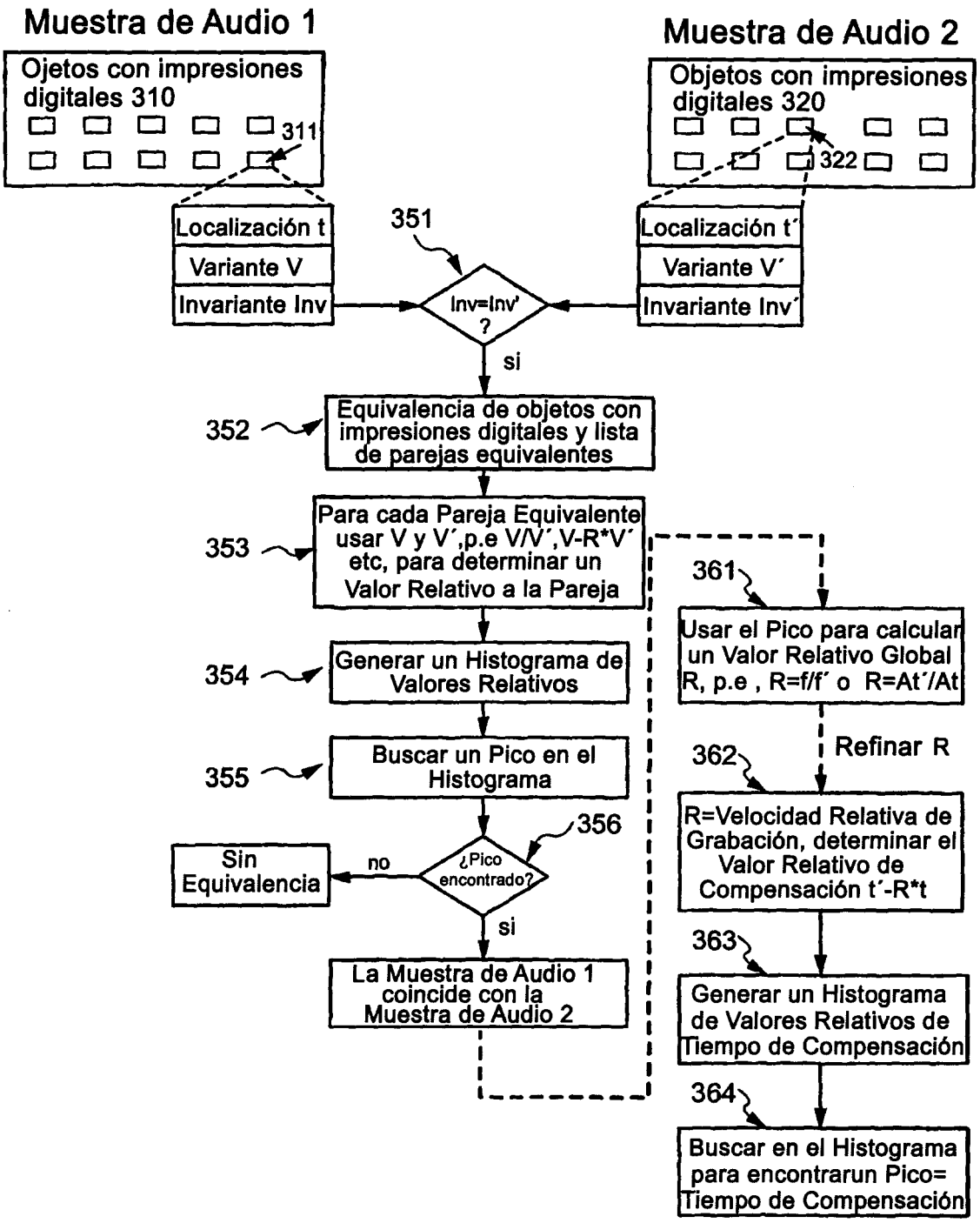
Muestra de Audio 1
Ojetos con impresiones digitales 310
311
Localización t
Variante V
Invariante Inv
Muestra de Audio 2
Objetos con impresiones digitales 320
322
Localización t´
Variante V´
Invariante Inv´
351
Inv=Inv´ ?
si
352
Equivalencia de objetos con impresiones digitales y lista de parejas equivalentes
353
Para cada Pareja Equivalente usar V y V´,p.e V/V´,V-R*V´ etc, para determinar un Valor Relativo a la Pareja
354
Generar un Histograma de Valores Relativos
355
Buscar un Pico en el Histograma
356
¿Pico encontrado?
no
Sin Equivalencia
si
La Muestra de Audio 1 coincide con la Muestra de Audio 2
361
Usar el Pico para calcular un Valor Relativo Global R, p.e , R=f/f´ o R=At´/At
Refinar R
362
R=Velocidad Relativa de Grabación, determinar el Valor Relativo de Compensación t´-R*t
363
Generar un Histograma de Valores Relativos de Tiempo de Compensación
364
Buscar en el Histograma para encontrarun Pico= Tiempo de Compensación

FIG. 3

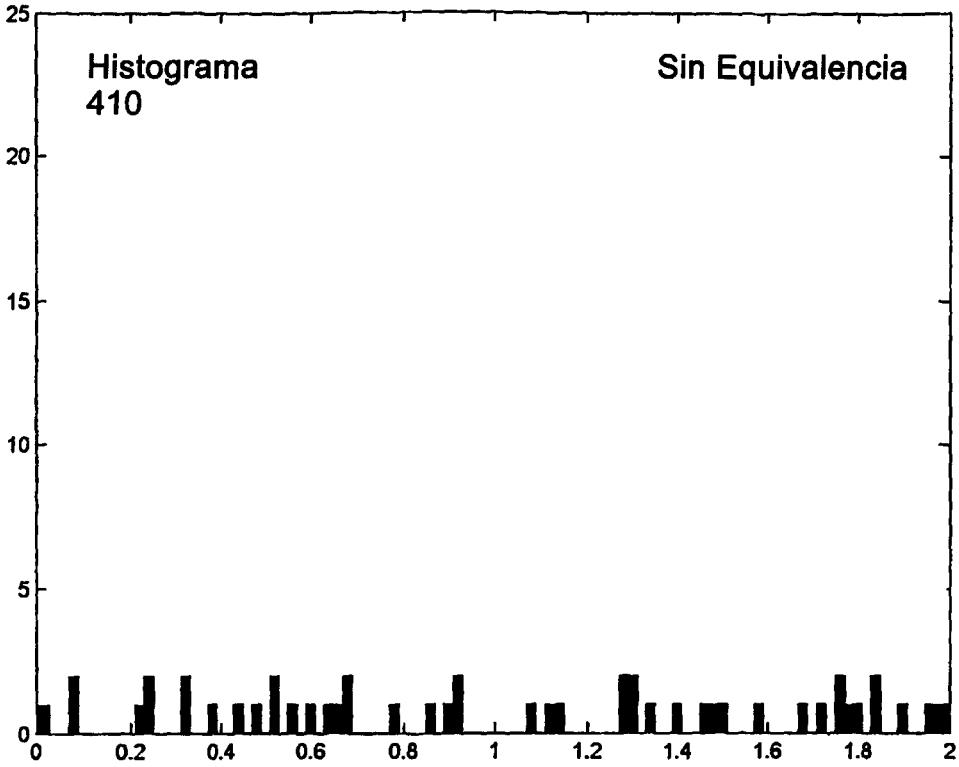
Histograma
410
Sin Equivalencia
25
20
15
10
5
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

FIG. 4A

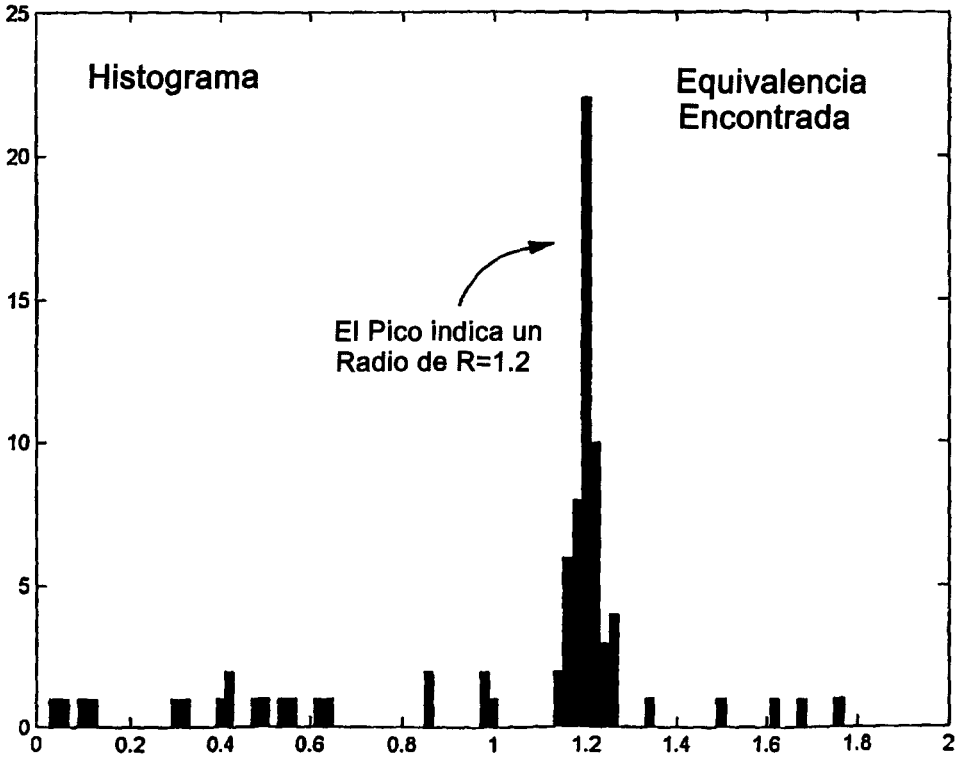
Histograma
Equivalencia
Encontrada
El Pico indica un
Radio de R=1.2
25
20
15
10
5
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

FIG. 4B

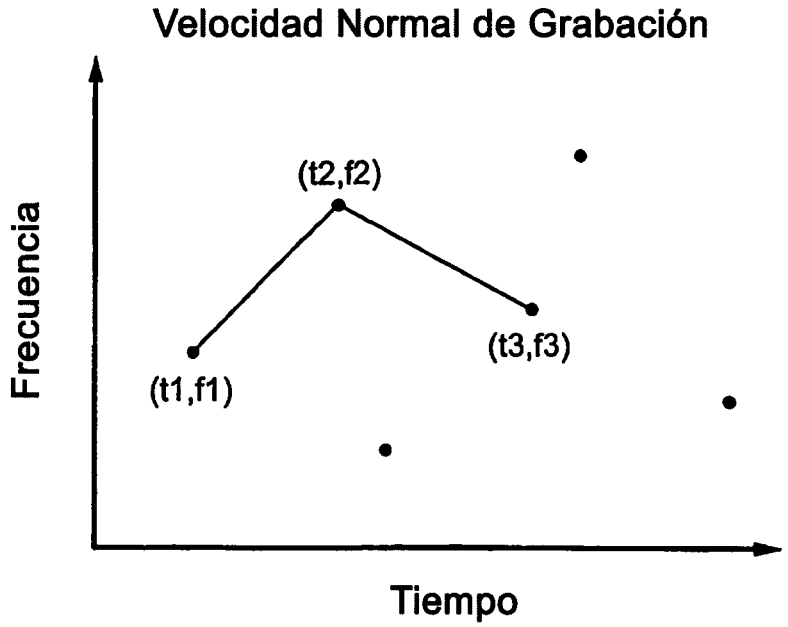
Velocidad Normal de Grabación
Frecuencia
(t2,f2)
(t3,f3)
(t1,f1)
Tiempo

FIG. 5A

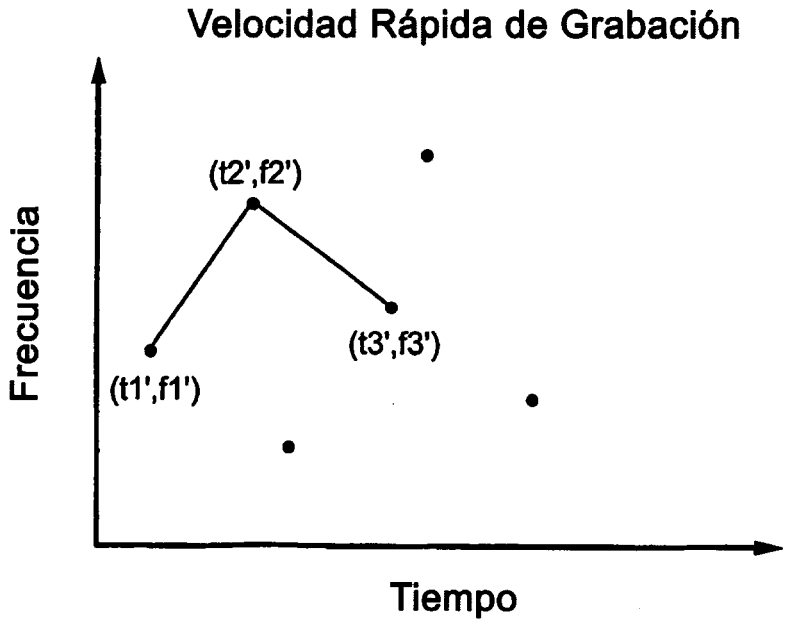
Velocidad Rápida de Grabación
Frecuencia
(t2',f2')
(t3',f3')
(t1',f1')
Tiempo

FIG. 5B

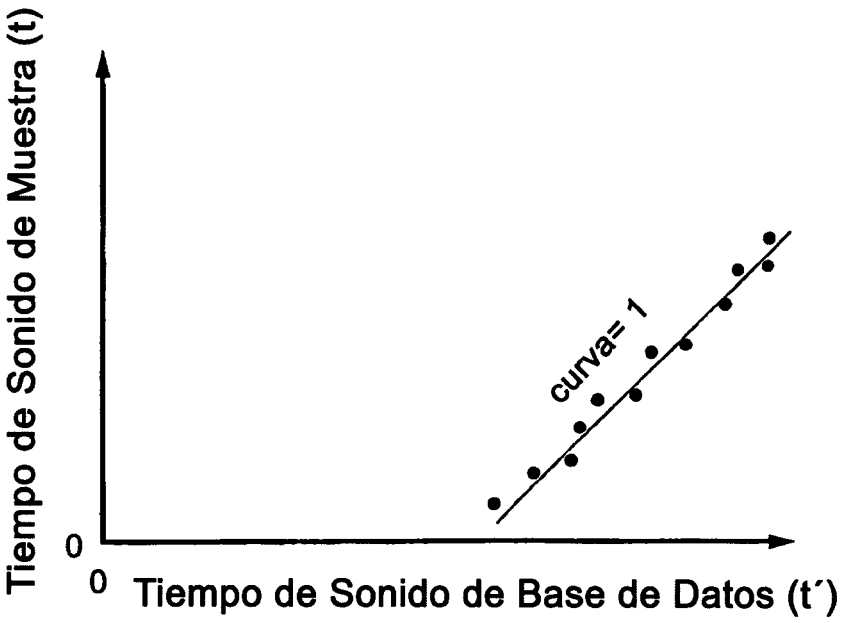
Tiempo de Sonido de Muestra (t)
curva= 1
0
0
Tiempo de Sonido de Base de Datos (t´)

FIG. 6A

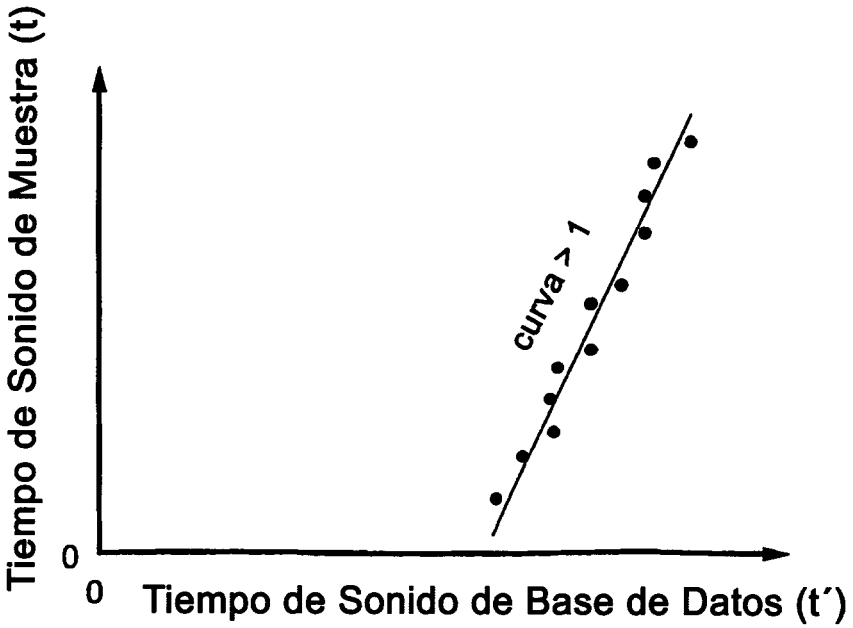
Tiempo de Sonido de Muestra (t)
curva > 1
0
0
Tiempo de Sonido de Base de Datos (t´)

FIG. 6B

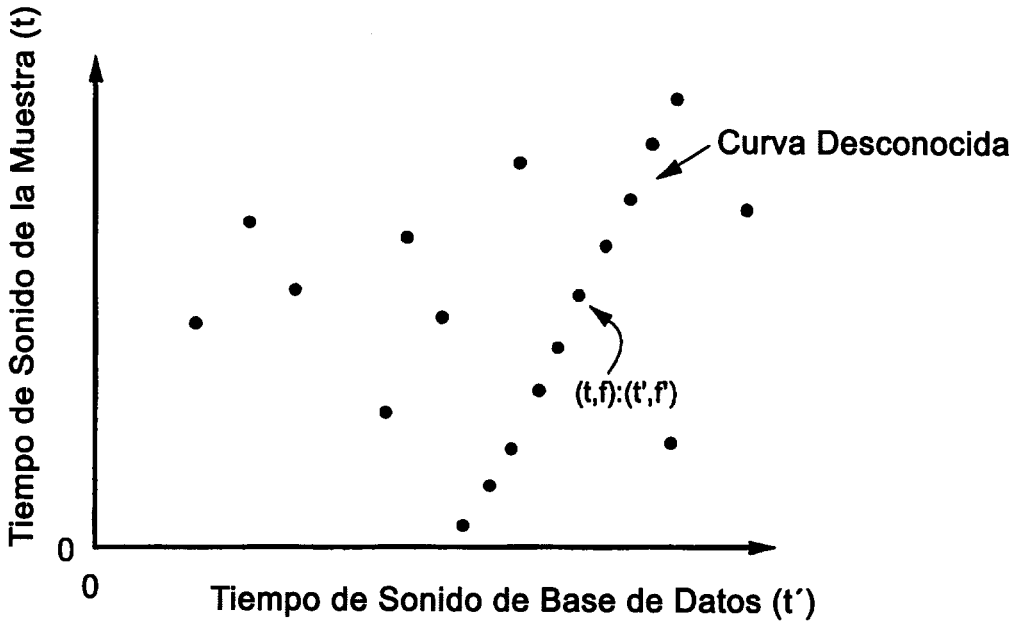
Tiempo de Sonido de la Muestra (t)
Curva Desconocida
(t,f):(t',f')
0
0
Tiempo de Sonido de Base de Datos (t´)

FIG. 7A

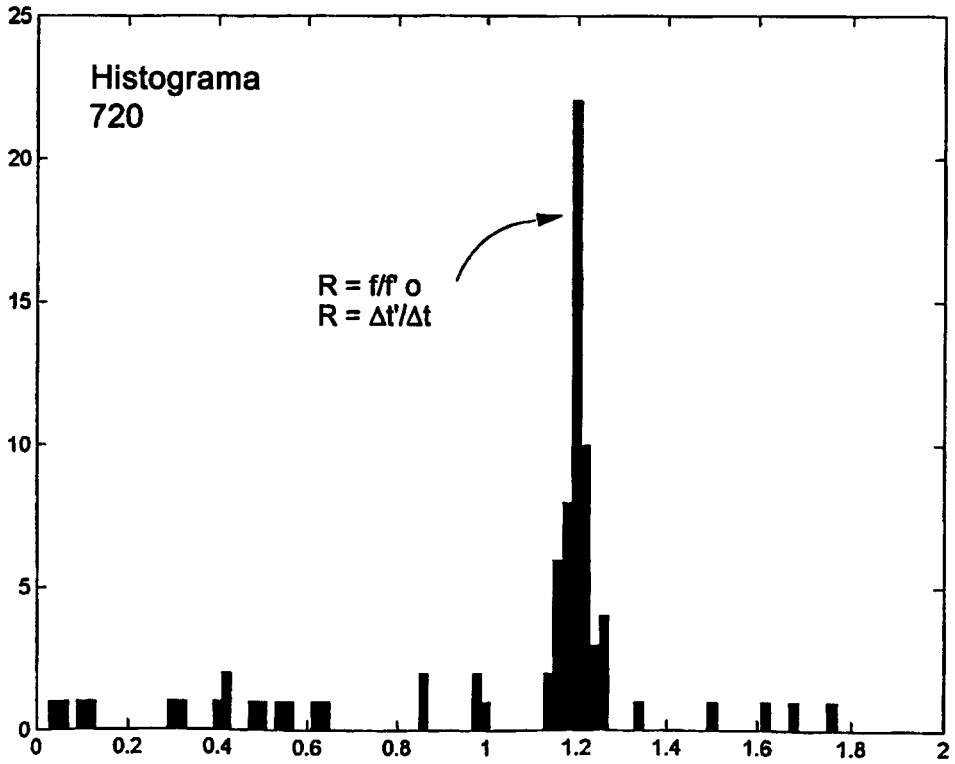
Histograma
720
R = f/f' o
R = Δt'/Δt
25
20
15
10
5
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

FIG. 7B

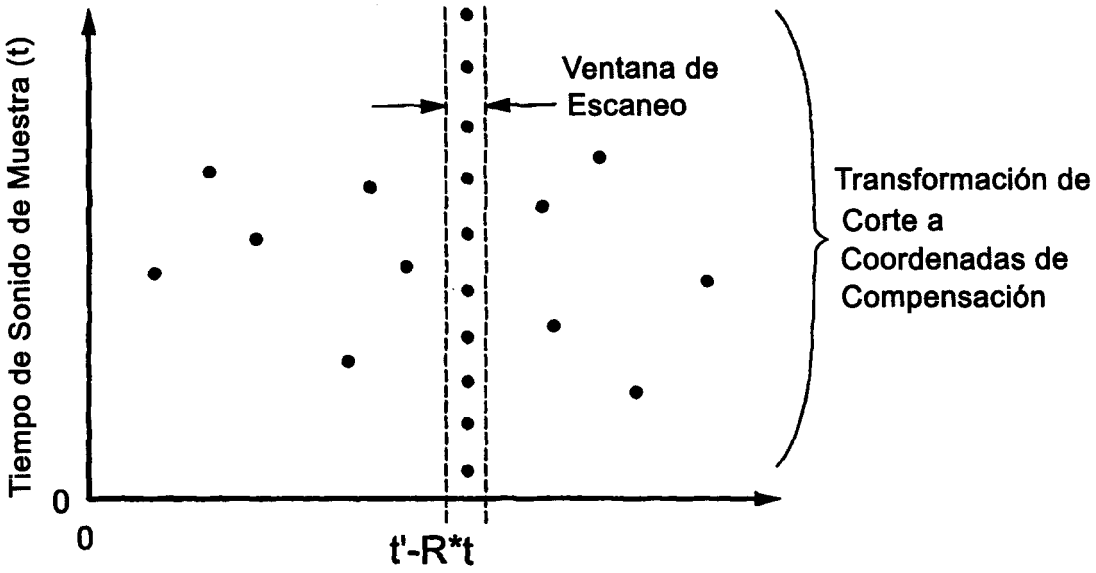
Tiempo de Sonido de Muestra (t)
Ventana de Escaneo
Transformación de Corte a Coordenadas de Compensación
0
0
t'-R*t

FIG. 7C

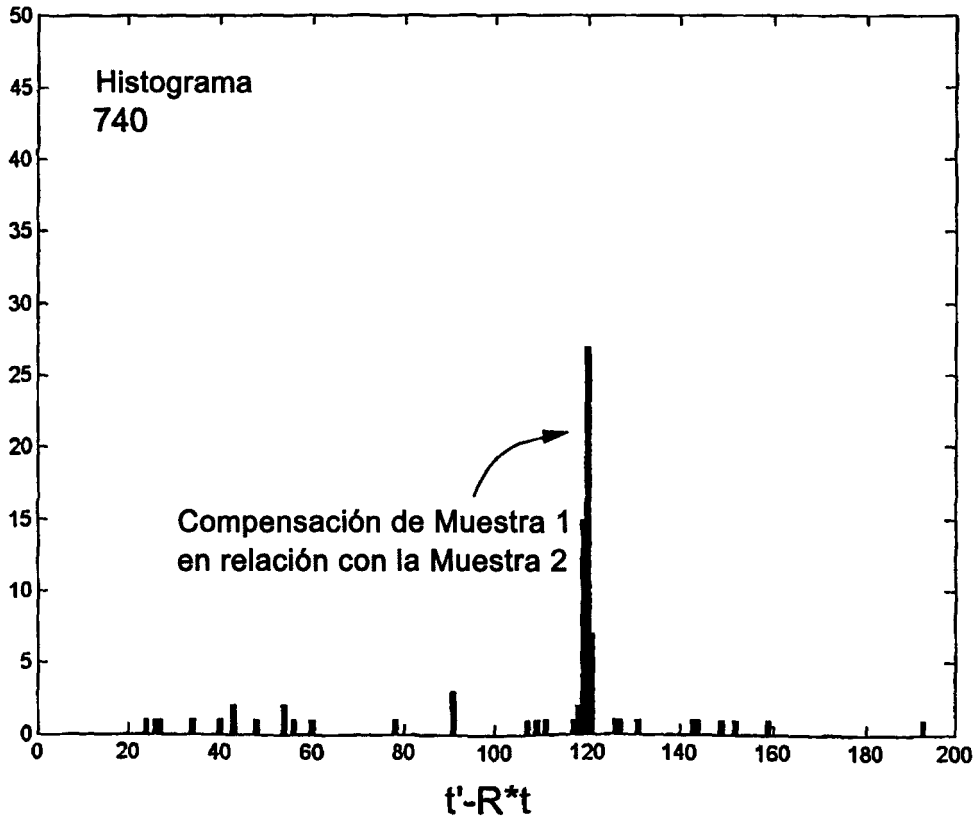
Histograma
740
Compensación de Muestra 1 en relación con la Muestra 2
50
45
40
35
30
25
20
15
10
5
0
0
20
40
60
80
100
120
140
160
180
200
t'-R*t

FIG. 7D