



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 356 476**

51 Int. Cl.:
G10H 1/36 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **06709573 .7**

96 Fecha de presentación : **26.01.2006**

97 Número de publicación de la solicitud: **1849154**

97 Fecha de publicación de la solicitud: **31.10.2007**

54 Título: **Procedimiento y aparato para su uso en la modificación de sonido.**

30 Prioridad: **27.01.2005 GB 0501744**
27.01.2005 US 647555 P

45 Fecha de publicación de la mención BOPI:
08.04.2011

45 Fecha de la publicación del folleto de la patente:
08.04.2011

73 Titular/es: **SYNCHRO ARTS LIMITED**
13 Links Road
Epsom, Surrey KT17 3PP, GB

72 Inventor/es: **Bloom, Phillip Jeffrey;**
Ellwood, William, John y
Newland, Jonathan

74 Agente: **Ponti Sales, Adelaida**

ES 2 356 476 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

[0001] La presente invención se refiere a procedimientos y aparatos para la modificación de al menos una característica acústica de una señal de audio.

5 **[0002]** Es bien sabido que es difícil hablar o cantar con un audio o audio/vídeo de manera que la nueva actuación es una repetición sincronizada de manera precisa de las palabras originales del actor cantante. En consecuencia, una grabación de la nueva actuación es muy poco probable que tenga sus propiedades acústicas iniciales y detalladas sincronizadas con las de la pista de audio original. De modo similar, las características tales como el tono de un nuevo cantante podrían no ser tan precisas o intrincadamente variadas como las del cantante original. Hay muchos casos en la industria de la grabación de audio profesional y en los juegos de consumo de ordenador y en actividades donde se hace una grabación de sonido de una voz y el tono musical de la voz que se graba de nuevo se beneficiarían del ajuste del tono, significando generalmente corrección, para ponerla en sintonía con una grabación de voz original. Además, una grabación del canto de un aficionado normal, aunque sea entonado, no tendrá el estilo vocal experto ni las inflexiones de tono de un cantante profesional.

15 **[0003]** La figura 4 muestra las mediciones de tono de un cantante profesional (Tono de Guía 401) y un miembro del público (Nuevo Tono 402) cantando las mismas palabras en la misma pista musical. Las diferencias de tiempo entre los inicios y las compensaciones de las secciones correspondientes (pulsos) de las señales de voz (valores de tono que no son cero Hz), así como las posiciones de las secciones sordas o en silencio (en cero Hz) son frecuentes y significativas. La aplicación de datos de tono del Tono de Guía 401 directamente en los mismos momentos relativos a los datos del Nuevo Tono 402 sería claramente equivocada e inadecuada para una cantidad substancial del segmento mostrado. Este es un resultado típico e ilustra los problemas básicos a resolver.

20 **[0004]** El ajuste del tono musical nota por nota se puede aplicar de forma automática al canto grabado o en vivo mediante hardware comercialmente disponible y dispositivos de software, que generalmente sintonizan las notas entrantes a redes fijas especificadas de tonos de notas aceptables. En estos sistemas, cada nota de salida se puede corregir automáticamente, pero este enfoque a menudo puede conducir a resultados inaceptables o desagradables, ya que puede eliminar variaciones "humanas" naturales y deseables.

30 **[0005]** La base fundamental para la identificación de tonos de destino en estos dispositivos de software y hardware conocidos es una escala musical, que es básicamente una lista de frecuencias de las notas específicas con las que el dispositivo debe comparar primero la señal de entrada. La mayoría de dispositivos vienen con escalas musicales predeterminadas para escalas estándar y permiten la personalización de estas, por ejemplo para cambiar los tonos de destino o para dejar algunas notas de tono sin alteraciones.

35 **[0006]** Los dispositivos de software conocidos se pueden colocar en un modo automático, que es también en general cómo trabajan los dispositivos de hardware: el dispositivo detecta el tono de entrada, identifica la nota de la escala más cercana en una escala predeterminada especificada por el usuario, y cambia la señal de entrada de tal manera que el tono de salida coincida con el tono de la nota de la escala especificada. La velocidad a la que se cambia el tono de salida y vuelve al tono de destino, a veces descrita como "velocidad", es controlada para ayudar a mantener los contornos de los tonos naturales (es decir, tono en función del tiempo) con mayor precisión y de forma natural y permitir una mayor variedad de "estilos".

40 **[0007]** Sin embargo, el canto grabado de un aficionado no puede mejorarse mediante técnicas conocidas de ajuste automático para conseguir las variaciones de tono complejos y expertas encontradas en la actuación de un cantante profesional.

[0008] También se conocen procedimientos y sistemas de procesamiento de voz que realizan la corrección de tono y/u otras modificaciones vocales utilizando voces de destino u otras secuencias almacenadas de datos de parámetros de voz de destino para especificar las modificaciones deseadas. Estos procedimientos conocidos tienen una o más deficiencias significativas. Por ejemplo:

50 1. El tono de destino (u otra característica vocal) que se aplica a la señal de voz de entrada del usuario sigue rígidamente el tiempo de una pista de karaoke u otro acompañamiento de este tipo que el usuario canta - generalmente en tiempo real - y no se realiza ningún intento para alinear las correspondientes características vocales (patente US 5.966.687, patente japonesa 2003044066). Si la voz del usuario se inicia demasiado pronto en relación con el tiempo de los datos de la característica de destino (tono, por ejemplo), entonces la característica de destino se aplicará, erróneamente, a las últimas palabras o sílabas. Un problema similar se produce si la voz del usuario es tardía. Dentro de las frases, palabras o sílabas que están fuera de tiempo con la pista de música se le asignará el tono equivocado o cualquier otra característica para esa palabra o sílaba. De manera similar, cualquier segmento de voz que se producen cuando se espera recibir segmentos sordos no recibe ningún tono de destino almacenado u otra información característica de destino.

2. El tono de destino (o la otra característica vocal) que se aplica a la voz de entrada del usuario se basa en y sigue la detección de una secuencia de fonemas de entrada almacenados esperados o patrones de voz/sordos de manera similar o solamente vocales (por ejemplo, US 5.750.912). Estos procedimientos generalmente requieren la formación de usuarios o la introducción de características fijas de datos de fonemas y/o requerir una pronunciación lo suficientemente cercana de las mismas palabras para que se produzca una identificación precisa. Si no hay entrenamiento y el conjunto de fonemas del usuario difiere lo suficiente del conjunto almacenado no será reconocido, y el sistema no funcionará correctamente. Si los fonemas del usuario no son lo suficientemente largos, o son demasiado cortos, las notas de salida se pueden truncar o cortar. Si fonemas llegan demasiado pronto o demasiado tarde, el tono o la característica se podrían aplicar al fonema correcto, pero estarán fuera de tiempo con el acompañamiento musical. Si el usuario pronuncia mal el fonema o fonemas, el sistema puede fallar para mantener las coincidencias. Además, en una canción, un solo fonema a menudo dará una serie de múltiples y/o una serie continua de tonos en los que un sistema basado en fonemas es poco probable que aplique el tono correcto o cambios de las características. El reconocimiento preciso del fonema también requiere un tiempo de procesamiento diferente a cero - lo que podría retrasar la aplicación de las características correctas en un sistema en tiempo real. Los sonidos que no son vocales (por ejemplo, una flauta) no se pueden utilizar como señales de guía o entradas.

3. El modelo de tono de destino se basa en una serie de notas discretas que se describen típicamente por tablas (por ejemplo, datos Midi), que generalmente cuantifican tanto en el tono como el tiempo. En este caso, las modificaciones a la entrada de voz se limitan a las notas guardadas. Este enfoque conduce a un conjunto restringido de patrones vocales disponibles que se pueden generar. El control de las transiciones entre las notas, el vibrato y el glissando se limitará generalmente a descriptores secundarios basados en notas (es decir, Midi). Además, la voz de canto de tono corregido procesada puede tomar un sonido mecánico (monótono), y si el tono se aplica a la parte equivocada de una palabra a destiempo, entonces la canción sonará cantada extraña y posiblemente también fuera de tono.

4. El sistema está diseñado para funcionar en tiempo casi real (como en un sistema de karaoke en vivo) y para crear una salida en breve (es decir, en una fracción de segundo) después de que se haya recibido la entrada (que debe corregirse). Aquellos que utilizan fonemas o características similares (por ejemplo, la patente US 5.750.912) están restringidos a un espacio de tiempo muy localizado. Estos sistemas pueden salir del paso, provocando por ejemplo, que las vocales del cantante de karaoke coincidan con la parte equivocada del canto de destino de guía.

[0009] Por lo tanto, los problemas surgen en la modificación de una señal de sonido digitalizada basada en las características de otra señal de sonido digitalizada, donde las características correspondientes de la primera y segunda señales de sonido no se producen en las mismas posiciones relativas en el tiempo dentro de las señales respectivas.

[0010] La patente US 6.836.761 describe un aparato de conversión de voz en el que se almacenan los marcos temporales sucesivos de datos de una voz de destino, una señal de entrada de una voz de un usuario se analiza para extraer los datos del mismo tipo de una sucesión de marcos temporales de la señal, y una señal de salida se sintetiza mediante la utilización de datos de la voz de destino de acuerdo con los datos de marcos correspondientes de la señal de entrada.

[0011] En un artículo titulado "Speaker - specific pitch contour modelling and modification" en las páginas 885 a 888 de Acoustics, Speech and Signal Processing, 1998, Actas de la Conferencia Internacional IEEE 1998 en Seattle, WA, EE.UU., mayo de 1998, D.T. Chappell y J.H.L. Hansen describen la generación de una frase de salida mediante la comparación del contorno del tono con una frase de entrada hablada con los contornos del tono de un conjunto de frases almacenadas habladas por personas diferentes, la selección de la frase almacenada que tiene un contorno de tono que proporciona la menor diferencia de trayectoria de deformación de tiempo dinámica con el contorno de tono de la frase de entrada, y usando la diferencia de trayectoria de deformación de tiempo dinámica entre el contorno del tono de la frase seleccionada almacenada y el contorno del tono de la misma frase almacenada para un hablante deseado para la deformación temporal del contorno de tono esta último para producir la frase de salida.

[0012] Existe, por lo tanto, la necesidad de un procedimiento y un aparato que en primer lugar establezca una relación temporal detallada entre las características que varían con el tiempo de una nueva interpretación vocal y las características correspondientes de una interpretación vocal de guía. En segundo lugar, esta trayectoria de alineación temporal debe utilizarse como un mapa temporal para determinar y aplicar los ajustes de las características (por ejemplo, tono) correctamente a la nueva interpretación vocal en el momento adecuado de manera precisa. Cuando se hace correctamente, esto permite que los matices y la complejidad que se encuentran en la interpretación vocal de guía (por ejemplo, para el tono: vibrato, curvas de inflexión, deslizamientos, saltos, etc.) se impongan en la nueva interpretación vocal. Además, si se aplica la alineación temporal, se pueden controlar otras características como complemento o como alternativa al tono, por ejemplo características glotales (por ejemplo, voz

entrecortada y ronca), resonancias del tracto vocal, EQ, y otras.

5 **[0013]** Otro objetivo de esta invención es proporcionar procedimientos para modificaciones vocales que operan bajo condiciones de señal de entrada no ideales, especialmente cuando la nueva entrada (la voz del usuario, por ejemplo): (a) es de banda limitada y/o limitada en el rango dinámico (por ejemplo entrada a través de un sistema telefónico), (b) contiene ciertos tipos de ruido o distorsión; o (c) es de una persona con un acento diferente, sexo o edad a la voz de guía (de destino), o con una temporización muy diferente de entrega de palabras y fonemas si son iguales o diferentes de la señal de guía (de destino) e incluso con idiomas de entrada diferentes.

10 **[0014]** Un objetivo adicional es proporcionar un procedimiento que no requiere ningún tipo de información previa a cada señal que se almacena por ejemplo, respecto a la naturaleza fonémica de las señales, o el conjunto detallado de los estados de señal posibles que se podrían aplicar a la señal de salida. Así, un objetivo adicional relacionado es proporcionar un procedimiento que pueda funcionar con una señal de guía de audio y una nueva señal de audio, una o ambas de las cuales no se requiere que sea habla o canto.

15 **[0015]** Ya existen sistemas y procedimientos para la asignación de tiempo y la alineación de señales de audio. Un procedimiento y un aparato para determinar las diferencias de tiempo entre dos señales de audio y la alineación automática de tiempo de una de las señales de audio a la otra mediante la edición automática de la forma de onda ha sido descritos en la patente GB 2117168 y en la patente US 4591928 (Bloom et al.). Otras técnicas para la alineación de tiempo se describen en J Holmes y W
20 Holmes, (2001), "Speech synthesis and recognition, 2ª Edición", Taylor y Francis, Londres.

[0016] Las técnicas de cambio de tono y otras modificaciones vocales también están bien establecidas, siendo un ejemplo K. Lent (1989), "An efficient method for pitch shifting digitally sampled sounds", Computer Music Journal Vol. 13, N° 4, en las páginas 65 a 71.

25 **[0017]** La invención se define mediante las reivindicaciones 1 y 18 a continuación, referencia a las cuales debe hacerse ahora.

30 **[0018]** Realizaciones preferidas de la presente invención proporcionan procedimientos y aparatos para modificar automática y correctamente una o más características de la señal de una segunda señal de audio digitalizada para ser una función de las características especificadas en una primera señal de audio digitalizada. En estas realizaciones, las relaciones de tiempo relativo de las características especificadas en ambas señales se establecen primero. Sobre la base de estas relaciones de tiempo, se pueden aplicar correctamente las modificaciones detalladas y de tiempo crítico de las características de la señal. Para lograr esto, se genera una función de alineación del tiempo para crear una asignación entre las características de la primera señal y las características de la segunda señal y proporcionar una función de edición opcional de la segunda señal (del usuario).

35 **[0019]** Aplicaciones particulares de la presente invención incluyen la transferencia precisa de características de audio seleccionadas de una interpretación vocal digitalizada de un intérprete profesional a - y mejorando así - la interpretación de audio digitalizada de una persona menos profesional. Una aplicación específica de esta invención es ajustar automáticamente el tono de una nueva señal de audio ("Nueva Señal") generada por un miembro típico del público para seguir el tono de otra señal de audio
40 ("Señal de Guía") generada por un cantante profesional. Un ejemplo de esto es un sistema de reproducción y grabación de estilo karaoke usando videos digitalizados de música como fuente original en el que, durante la reproducción del audio original y del video correspondiente opcional, la voz del usuario se digitaliza y entra en el aparato (como Nueva grabación). Con este sistema, se puede crear una señal de voz de usuario modificada que tiene automáticamente el tiempo y el tono corregidos. Cuando la señal de voz modificada se reproduce de forma sincrónica con el vídeo original, la voz del usuario se puede reemplazar con precisión con la voz grabada del intérprete original en términos de afinación y tiempo,
45 incluyendo la sincronización de los labios. Durante la reproducción del video musical, el impacto de esta sustitución será incluso más efectivo si la señal de voz original reemplazada no es audible durante la reproducción con la grabación de voz del usuario modificada. La grabación de voz modificada se puede
50 combinar con la música reproducida original, tal como se describe en el documento WO 2004/040576.

[0020] Una aplicación adicional de esta invención es la creación de un archivo de sonido personalizado para su uso en sistemas de telefonía. En estas aplicaciones, el usuario canta o incluso habla para proporcionar una señal de voz que se graba y luego se mejora (por ejemplo con tono y tiempo
55 corregido para seguir las características de una versión de un cantante profesional) y, opcionalmente, se mezcla con una pista de fondo apropiada. La grabación de usuario mejorada resultante puede ponerse a disposición de los usuarios de teléfonos como tono de llamada personalizado o archivo de sonido para otros fines. El aparato que es una realización de la invención puede entonces tomar la forma de, por ejemplo, un servidor acoplado en un sistema de telecomunicaciones que comprende una red de telecomunicaciones y/o Internet, y puede utilizar un teléfono móvil como interfaz entre el aparato y los
60 usuarios. Además, o alternativamente, un teléfono móvil se puede adaptar para ser una realización de la invención. En este sistema, una señal de voz modificada, o datos que representan esta señal, producida

por una realización de la invención, puede transmitirse a un destinatario seleccionado a través de un sistema de entrega de tono de llamada utilizarse como tono de llamada u otra señal de sonido de identificación.

5 **[0021]** En realizaciones preferidas de la invención, la inclusión de la etapa de crear una función de asignación en función del tiempo entre las Señales de Guía y Nueva asegura que se realizan las modificaciones de la característica de la señal en el momento adecuado dentro de la Nueva Señal, independientemente de las diferencias substanciales entre las dos señales. La función de alineación de tiempo se utiliza para asignar los datos de la función de la característica de control al proceso de modificación de la señal deseada. El proceso de modificación accede a una Nueva Señal y la modifica
10 como se requiera. Esta acción crea una nueva tercera señal de audio a partir de la Nueva Señal. En consecuencia, la tercera señal tiene entonces las características de variación del tiempo deseadas determinadas por las características especificadas como características de control de la Señal de Guía.

15 **[0022]** En una realización, una segunda señal de audio, la Nueva Señal, se modifica con el tiempo (tiempo no lineal comprimido o expandido) utilizando la información de asignación de la función de alineación de tiempo, de manera que sus características que varían con el tiempo se alinean en el tiempo con una primera señal de audio, la Señal de Guía. Esta alineación del tiempo puede tener lugar antes o después de realizarse las modificaciones deseadas descritas anteriormente.

20 **[0023]** En una realización alternativa, el proceso de alineación del tiempo no se realiza en forma de onda nueva o modificada. En cambio, la trayectoria de deformación a lo largo del tiempo se utiliza para asignar las características de control de la primera señal (parámetros de control de audio de la Señal de Guía) a la segunda señal para modificar las partes apropiadas de la segunda forma de onda de la señal y mantener su temporización original.

25 **[0024]** Al realizar el procesamiento sin la limitación de procesamiento en tiempo real, se puede realizar un análisis detallado de las versiones almacenadas de la Señales de Guía y Nueva, y una cantidad estadísticamente significativa y substancial de las dos señales (por ejemplo, hasta 30 segundos o incluso las señales completas) se procesan antes de iniciar el proceso de alineación de tiempo y se realizan las decisiones críticas respecto a características de la señal a largo plazo.

30 **[0025]** En consecuencia, las discrepancias de tiempo a gran escala (por ejemplo, de varios segundos) se pueden acomodar y corregir y se puede realizar una alineación localizada óptima dentro de las palabras y frases. Además, las modificaciones de las características también se hacen "fuera de línea", permitiendo aplicar el procesamiento de la más alta calidad, así como una interpolación y/o suavizado de los datos relacionados con la modificación, para eliminar cualquier error graves evidente antes de la aplicación a la Nueva Señal.

35 **[0026]** Conjuntos de valores de características de salida para la Nueva Señal no tienen que ser predefinidos. Por ejemplo, si el tono de una Nueva Señal proporcionada por un usuario debe corregirse para que coincida con el tono de una Señal de Guía en forma de una grabación de un cantante profesional, los valores de tono aceptables no es necesario que se definan o establezcan. En cambio, la voz del usuario se ajustará a los valores que están presentes y se miden en la grabación de la Señal de Guía.

40 **[0027]** La Nueva Señal no tiene que limitarse para parecerse a la Señal de Guía o generarse mediante el mismo tipo de procesos acústicos que la Señal de Guía. Por ejemplo, el habla monótona podría se podría modificar en el tiempo y en el tono para seguir un instrumento de viento solista o un canto de pájaro. Mientras las dos señales tienen algunas características variables en el tiempo que puedan ser tratadas como relacionadas, un procedimiento de la invención puede crear una señal de salida con propiedades convenientemente modificadas. Además, las características de la Nueva Señal y
45 la Señal de Guía pueden estar desplazadas en las frecuencias entre sí. Por ejemplo, el tono de una señal puede ser una octava o más aparte de la otra señal.

[0028] También hay que indicar que una o ambas de las señales de audio pueden estar en las regiones de ultrasonido o de infra sonido.

50 **[0029]** Mediante la operación de una realización preferida de la presente invención, las variaciones de tono complejo y experto (y, opcionalmente, otras características) que se encuentran en la interpretación de un cantante profesional pueden transferirse con precisión a la voz digitalizada de un cantante usuario (por ejemplo, aficionado). Esto mejora muchos aspectos de la interpretación del usuario a nivel profesional.

55 **[0030]** Las realizaciones de la invención también se pueden aplicar en el campo de la sustitución automática de diálogos (ADR) para mejorar una interpretación grabada en un estudio ADR de un actor. Una realización se puede utilizar para modificar las características vocales de la grabación de estudio, tales como el tono, el nivel de energía y las características prosódicas para que coincidan o sigan las de la Señal de Guía original grabada en el estudio o el lugar con la imagen. Además, el actor en el

estudio puede ser un actor diferente del que grabó la Señal de Guía.

[0031] Además, la invención es flexible en el rango de los procesos que se pueden aplicar. Por ejemplo, en el caso de ajuste de tono, funciones adicionales de cambio de tono, tal como generación de armonía alineada en el tiempo, se pueden introducir como funciones de la función de ajuste del tono para crear señales de salida alternativa. Además, una característica medida en la Señal de Guía se puede asignar mediante una función arbitraria para controlar otra característica completamente diferente en la Nueva Señal.

[0032] Los procedimientos de esta invención se pueden implementar con programas de ordenador en un sistema informático, tal como un PC o una consola de juegos basada en ordenador con medios para la entrada y salida de audio.

[0033] Hay muchas permutaciones de las disposiciones de las secuencias de procesamiento que pueden implementarse, teniendo algunas ventajas sobre otras en ciertas situaciones. Los ejemplos a continuación se dan respecto al procesamiento del tono para ilustrar cómo las variaciones afectan a la complejidad de procesamiento y/o reducen el potencial para generar de artefactos de señal audible en la señal de salida. Observaciones y resultados similares se producirán al considerar las características de procesamiento que no sean el tono, tal como la sonoridad, el tono o la estructura formante.

[0034] Típicamente, en una realización, para empezar, las Señales Nueva y de Guía se muestrean y se almacenan digitalmente. A continuación, un análisis robusto de corto plazo de las características independientes del hablante extrae los perfiles de modulaciones de las características de las señales en las dos señales. Las mediciones de la energía espectral se realizan cada 10 ms en sucesivos "marcos" de ventanas de las señales, proporcionando algoritmos de compensación del ruido y del nivel (por ejemplo, tal como se describe en la patente US 4.591.928). Este análisis se realiza sobre toda la señal de entrada para maximizar la precisión y la solidez del procesamiento. Otras mediciones de las características de corto plazo se pueden usar alternativamente, ejemplos de las cuales se pueden encontrar en L.R. Rabiner y R.W. Schafer (1978) "Digital Processing of Speech Signals", Prentice Hall.

[0035] Tomando el ejemplo de la determinación de tono, el resto de etapas del procesamiento de la señal principal que se realizan en el sistema informático de las señales grabadas y sus datos de las características de la señal medida son:

Procedimiento 1

[0036]

(a) Las secuencias de características dependientes del tiempo de la Señal de Guía y de la Nueva Señal se procesan en un algoritmo de coincidencia de patrones que determina y produce una función óptima de trayectoria de alineación del tiempo como una secuencia de datos. Esta trayectoria óptima correlaciona los marcos de la Nueva Señal en los marcos de la Señal de Guía.

(b) Los datos de la trayectoria de alineación de tiempo se utilizan para editar la nueva señal y generar una Nueva Señal de que está alineada en el tiempo con la Señal de Guía.

(c) La Señal de Guía se divide en marcos consecutivos discretos y se mide el tono de cada marco. Los valores de la secuencia de medición del tono se suavizan para proporcionar el contorno del tono de la Señal de Guía.

(d) El procesamiento en la etapa (c) se repite para el alineado (editado) de la Nueva Señal para generar su contorno del tono.

(e) Cada valor de contorno del tono de la Señal de Guía se divide por el valor del contorno del tono correspondiente a la Nueva Señal alineada y ajustada para cambios de octava para generar un contorno de corrección que es un conjunto de valores que dan el factor de corrección a aplicar a cada marco de la Nueva Señal alineada. Este contorno de corrección se suaviza para eliminar cualquier error manifiesto.

(f) Un algoritmo de cambio de tono se utiliza para cambiar el tono de la Nueva Señal alineada con los valores de acuerdo con el contorno de corrección suavizado de la etapa (e) y así generar una Nueva Señal coincidente en el tiempo y el tono con la Señal de Guía dada.

[0037] El Procedimiento 1 utiliza dos algoritmos de edición en cascada y mide el tono de la Nueva Señal después de haber sufrido una etapa de edición. Así, la calidad de la salida generada en el Procedimiento 1 depende de la calidad de salida de la señal editada de la etapa (b). En consecuencia, las imperfecciones introducidas durante la edición en esa señal pueden degradar la calidad de las salidas de las etapas (d) y (f). Esto podría provocar pequeños errores ocasionales en el tono corregido y, posiblemente, crear una sutil rugosidad en la salida generada.

Procedimiento 2

[0038] Para reducir el riesgo de errores, otra realización combina las etapas anteriores (b) y (f) para formar una sola fase de edición. También cualquier característica de la Nueva Señal (en este ejemplo, el tono) se mide desde la Nueva Señal sin modificar, y no desde una versión alineada en el tiempo (editada). Esto se logra mediante el cálculo de la inversa de la trayectoria de alineación del tiempo. La trayectoria inversa correlaciona cada marco de la Nueva Señal sin editar a su marco correspondiente de la Señal de Guía. A partir de esta correlación se calcula un contorno de corrección de tono para la Nueva Señal que está alineado en el tiempo para la Señal de Guía. En efecto, la Señal de Guía se alinea en el tiempo con la Nueva Señal antes de calcular el contorno de corrección del tono.

10 **[0039]** Las siguientes etapas resumen este procedimiento.

(a) Las secuencias de características dependientes del tiempo de la Señal de Guía y de la Nueva Señal se procesan en un algoritmo que coincide con el patrón que determina y produce una función de trayectoria de alineación de tiempo óptima como una secuencia de datos que de forma óptima correlaciona los marcos de la Nueva Señal con los marcos de la Señal de Guía.

15 (b) Los datos de la trayectoria de alineación de tiempo se utilizan para producir una función de trayectoria inversa que correlaciona los marcos de la Señal de Guía para los marcos correspondientes de la Nueva Señal.

20 (c) La Señal de Guía se segmenta en marcos discretos y se mide el tono de cada marco. Los valores de secuencia de medición del tono se suavizan para proporcionar el contorno de tono de la Señal de Guía.

(d) El procesamiento en la etapa (c) se repite para la Nueva Señal (sin editar) para generar su contorno de tono.

25 (e) Usando la función de trayectoria inversa para alinear el contorno de tono de la Señal de Guía para el contorno de tono de la Nueva señal, cada valor de contorno de tono de la Señal de Guía correlacionada se divide por el valor del contorno del tono correspondiente a la Nueva Señal y ajustado para cambios de octavas para generar un contorno de corrección alineado que es un conjunto de valores que dan el factor de corrección a aplicar a cada marco de la Nueva Señal. Este contorno de corrección alineado se suaviza para eliminar cualquier error importante.

30 (f) Usando la función de trayectoria de alineación en el tiempo y el contorno de corrección alineado suavizado, la Nueva Señal se edita utilizando un algoritmo de procesamiento que cambia su tono y comprime el tiempo o expande el tiempo en la Nueva Señal tal como sea necesario para generar una señal de salida que se alinea en el tiempo y en el tono con la Señal de Guía.

35 (g) O, como alternativa a la etapa (f), el del contorno de corrección alineado suavizado podría aplicarse sin la alineación del tiempo de la Nueva Señal con la Señal de Guía. Esto mantendría la temporización original de la Nueva Señal, pero aplicaría la corrección del tono a los marcos correctos de la Nueva Señal, aunque la Nueva Señal no se haya alineado en el tiempo con la Señal de Guía.

[0040] Cualquier forma del Procedimiento 2 proporciona una corrección de tono del sonido más fiable y natural en todas las palabras y frases, que puede seguir y recrear fielmente cualquier matiz sutil, tal como el vibrato y otros detalles.

40 **Procedimiento 3**

[0041] Aunque el Procedimiento 2 sólo edita la Nueva Señal una vez, utiliza una técnica de procesamiento que modifica la alineación del tono y del tiempo a la vez. Variando la secuencia de las etapas ligeramente es posible procesar por separado el cambio de tono y la modificación del tiempo sin utilizar el Procedimiento 1. Aunque esto introduce dos etapas de edición, los algoritmos de procesamiento especializados más adecuados se pueden elegir por separado para cada etapa.

[0042] Las siguientes etapas resumen el tercer procedimiento:

50 (a) Las secuencias de características dependientes del tiempo de la Señal de Guía y de la Nueva Señal se procesan en un algoritmo de coincidencia del patrón que determina y produce una óptima función de trayectoria de alineación con el tiempo como una secuencia de datos que, de forma óptima, correlaciona los marcos de la Nueva Señal con los marcos de la Señal de Guía.

(b) La Señal de Guía se segmenta en marcos discretos y se mide el tono de cada marco. Los valores de la secuencia de medición de tono se suavizan para proporcionar el contorno del tono de la Señal de Guía.

(c) El procesamiento en la etapa (b) se repite para la Nueva Señal (sin editar) para

generar su contorno de tono.

(d) Usando la función de trayectoria de alineación con el tiempo, el contorno del tono de la Nueva Señal se alinea efectivamente con el tiempo para el contorno de tono de la Señal de Guía.

5 (e) Cada valor del contorno del tono de la Señal de Guía se divide por el contorno del tono de la Nueva Señal alineada con el tiempo, y el resultado se ajusta a los cambios de octavas. Esto genera un contorno de corrección alineado que contiene los factores de corrección a aplicar a cada marco de una Nueva Señal alineada con el tiempo. Este contorno de corrección alineado se suaviza para eliminar cualquier error importante.

10 (f) Los datos de la trayectoria de alineación en el tiempo se utilizan para editar la Nueva Señal y generar una Nueva Señal de que está alineada con el tiempo con la Señal de Guía.

(g) Usando un algoritmo de cambio de tono, el tono de la Nueva Señal alineada con el tiempo cambia por el contorno de corrección alineado suavizado generado en la etapa (e). Esto da una Nueva Señal editada alineada en el tiempo y en tono con la Señal de Guía dada.

15 **[0043]** El Procedimiento 3 utiliza la función de trayectoria de alineación con el tiempo original y no la inversa. Además, tiene la ventaja como en el Procedimiento 2 que se mide el tono de la Nueva Señal sin modificar y no el de una versión alineada con el tiempo (editada). Sin embargo, no puede modificar el tono de la Nueva Señal (etapa g) sin primero generar una versión alineada en el tiempo (etapa f).

20 **[0044]** En otras realizaciones, otras características de una señal de sonido, además del tono, se pueden modificar para seguir las de una Señal de Guía, una vez que se ha creado una función de alineación en el tiempo. Los tipos adicionales de características modificables de sincronización con el tiempo incluyen modificaciones de las características de la señal de sonido tales como la sonoridad instantánea, la igualización, los patrones de los formantes o de la resonancia del habla, la reverberación y las características del eco, e incluso las propias palabras, previendo que esté disponible un mecanismo adecuado para el análisis y la modificación de las características especificadas.

[0045] En la presente invención, una señal de vídeo no es necesaria, y la señal de entrada de audio puede requerirse no solamente para acompañar o reemplazar a otra señal de audio.

30 **[0046]** En una realización preferida de la invención, se incluyen unos medios para determinar una función de alineación con el tiempo o de trayectoria de deformación en el tiempo, que puede proporcionar una correlación del tiempo óptimo y suficientemente detallada entre las características variables en el tiempo de una segunda (Nueva) señal de audio correspondiente con las características que varían con el tiempo de una primera (Guía) señal de audio. Esta correlación asegura que las alteraciones variables en el tiempo se basan en las características especificadas en la porción de la señal de Guía (control) que corresponde a la porción apropiada de la Nueva Señal que se modificada. Las mediciones de las características específicas que varían en el tiempo utilizadas para determinar la alineación en el tiempo se hacen cada T segundos, en porciones o ventanas cortas de las formas de onda de la señal de la muestra, siendo cada ventana de duración T', y T' puede ser diferente de T. Las mediciones se realizan en una base de marco a marco sucesivo, usualmente con la superposición de las ventanas de la muestra. Este es un análisis de la señal de "corto tiempo", tal como se describe en L.R. Rabiner y R.W. Schafer (1978) "Digital Processing of Speech Signals", Prentice Hall.

45 **[0047]** Debe indicarse que las características medidas para el proceso de alineación en el tiempo es probable que sean características diferentes de las características que se alteran y de las características utilizadas como control. Una relación funcional entre las características a alterar y los parámetros usados como control debe estar definida. Por ejemplo, una relación simple que se describe más detalladamente a continuación, modifica el tono de una Nueva Señal para que coincida con una Señal de Guía, con ajustes para mantener el rango de tono natural de la persona que crea la Nueva Señal. Esta definición de la función de modificación, y otras definiciones, también pueden variar con el tiempo si se desea. La función de modificación puede programarse como una agrupación de datos de valores de salida frente a valores de entrada, o como una función matemática o como un conjunto de reglas de procesamiento en el sistema informático de procesamiento de audio. Téngase en cuenta que esta función no depende necesariamente de la propia señal y así puede que la señal no necesite ningún análisis. En etapas adicionales, la característica especificada que se modifica en la segunda señal y la característica de control especificada en la primera señal se miden como funciones del tiempo. Estas mediciones se almacenan como datos.

55 Breve descripción de los dibujos

[0048]

La figura 1 es un diagrama de bloques de un sistema informático adecuado para su uso

en la implementación de la presente invención.

La figura 2 es un diagrama de bloques que muestra los componentes de software adicionales que se pueden añadir al ordenador en la figura 1 para implementar la presente invención.

5 La figura 3 es un diagrama de bloques de una realización de la presente invención, que muestra las señales y los módulos de procesamiento utilizados para crear una señal de salida de audio con ajustes de tono sobre una señal de entrada con diferentes características de tono y sincronización.

10 La figura 4 es un gráfico que muestra un ejemplo típico de las mediciones de tono en función del tiempo para una voz de Guía grabada por un cantante profesional y las mismas mediciones en una Nueva Señal grabada de un usuario inexperto cantando la misma canción con el mismo acompañamiento musical.

La figura 5 es un gráfico que representa una función de deformación en el tiempo o trayectoria de alineación.

15 La figura 6 es un gráfico que muestra en el eje de frecuencia izquierdo el tono de la Señal de Guía y el tono de la Nueva Señal Alineada de la figura 4 (antes de la corrección de tono) y el Factor de Corrección del tono suavizado calculado en el eje vertical derecho.

La figura 7 es un gráfico del tono de la Señal de Guía y el tono de la Nueva Señal Corregida que se muestra sin corregir en la figura 6.

20 La figura 8 es un diagrama de bloques de otra realización de la presente invención que muestra las señales y los módulos de procesamiento utilizados para crear una señal de salida de audio con las modificaciones de las características generales de la señal basadas en las características de alineación en el tiempo de una señal de entrada arbitraria.

La figura 9A es un diagrama de bloques de otra realización de acuerdo con el procesamiento de invención, en el cual las características de la Nueva Señal se modifican con o sin alineación en el tiempo simultánea de una Señal de Guía.

25 La figura 9B es un diagrama de bloques de otra realización de acuerdo con el procesamiento de la presente invención, en la cual la trayectoria de alineación en el tiempo se utiliza para crear una Nueva Señal Alineada en el Tiempo y para proporcionar una función de correlación para determinar con precisión las modificaciones que deben realizarse en la Nueva Señal Alineada en el Tiempo.

30 La figura 10 (a) es una representación gráfica de un ejemplo de las posiciones relativas y las formas de las ventanas de análisis utilizadas para crear una señal $s''(n)$ usando síntesis de superposición y de adición.

35 La figura 10 (b) es una representación gráfica de un ejemplo de las posiciones relativas y las formas de las ventanas de síntesis usadas para crear una señal $s''(n)$ usando síntesis de superposición y de adición.

La figura 11 es un diagrama de bloques de otra realización de la invención que utiliza un sistema de telecomunicaciones.

Descripción detallada de la invención

40 **[0049]** Son bien conocidos sistemas informáticos capaces de grabar sonido de entrada mientras se reproduce simultáneamente señales de sonido y/o vídeo a partir de archivos de audio y video por ordenador. Los componentes de un sistema de PC típico y ambiente que pueden soportar estas funciones se presentan en la figura 1 de los dibujos adjuntos, y este sistema se puede utilizar con el software en la figura 2 como la base para proporcionar el entorno de hardware y software para múltiples realizaciones de la presente invención.

45 **[0050]** En la figura 1 se muestra un sistema informático convencional 100 que consiste en un ordenador 110 con una CPU (Unidad Central de Procesamiento) 112, memoria RAM (Memoria de Acceso Aleatorio) 118, hardware de interfaz de usuario que típicamente incluye un dispositivo de señalización 120 tal como un ratón, un teclado 125, y una pantalla de visualización 130, un dispositivo de almacenamiento interno 140 tal como un disco duro o memoria RAM adicional, un dispositivo 160 para acceder a los datos en un medio de almacenamiento fijos o removible 165 tal como un CD-ROM o DVD ROM, y
50 opcionalmente un módem o una interfaz de red 170 para proporcionar acceso a Internet 175. El dispositivo de señalización 120 controla la posición de un cursor mostrado en la pantalla (no representado) y la selección de las funciones que aparecen en la pantalla 130.

[0051] El ordenador 110 puede ser cualquier ordenador doméstico o de negocios

convencional, tal como un PC o Apple Macintosh, o alternativamente una "máquina de juegos" dedicada, tal como Microsoft® Xbox™ o Sony Playstation 2™, siendo entonces el dispositivo de señalización 120 un dispositivo controlador del juego. Algunos componentes que se muestran en la figura 1 pueden estar ausentes de una máquina de juegos particular. La figura 2 ilustra otro software que se puede instalar en el ordenador 110.

[0052] Un usuario puede obtener de un CD-ROM, Internet u otros medios, un archivo de datos digitales 115 que contiene un audio y un clip de vídeo de acompañamiento opcional que, por ejemplo, podría estar en un formato común tal como el formato de película AVI o QuickTime® y que, por ejemplo, se copia y almacena en el disco duro 140 o en la memoria RAM. El ordenador 110 tiene un sistema operativo 135 conocido tal como el proporcionado por cualquiera de las versiones disponibles de Microsoft® Windows® o Mac® OS, software de audio y hardware en forma de una tarjeta de sonido 150 o hardware equivalente en la placa base del ordenador, conteniendo un ADC (Convertidor Analógico-Digital) al que está conectado un micrófono 159 de grabación y que contiene un DAC (Convertidor Digital-Analógico) al que están conectados uno o más altavoces 156 para la reproducción de audio.

[0053] Tal como se muestra en la figura 2, este sistema operativo 135 generalmente se suministra con software de grabación y edición de audio 180 que soporta la grabación de audio a través de la tarjeta de sonido 150 y funciones de edición, tal como el programa de aplicación "Grabador de sonidos" que se suministra con Windows®. El programa de grabación y/u otros programas pueden usar la tarjeta de sonido 150 para convertir una señal analógica de audio de entrada en datos de audio digitales y grabar esos datos en un fichero informático en el disco duro 140. Un software reproductor de audio y vídeo 190, tal como Windows Media Player se incluye con Windows® y/o cualquier otro software puede utilizarse para la reproducción de archivos de vídeo compuesto digital y audio o archivos de sólo audio a través de la tarjeta de sonido 150, también hardware y software de vídeo integrado, la pantalla 130 y los altavoces 156. Los archivos de vídeo y audio compuestos consisten en datos de vídeo y una o más pistas en paralelo sincronizadas de datos de audio. Alternativamente, los datos de audio pueden mantenerse como archivos separados asignados para almacenar múltiples corrientes de datos de audio. Los datos de audio pueden ser datos de voz, tales como diálogo o canto, música instrumental, "efectos de sonido", o cualquier combinación de éstos. Los bloques 180 y 190 también pueden, en concierto con 135 y 110, representar el software y el hardware que pueden implementar los sistemas de procesamiento de señales que se describirá aquí. Realizaciones alternativas distribuidas del sistema de hardware y software en 100 y 110 pueden utilizarse, siendo un ejemplo cuando los principales elementos del sistema informático 100 se proporcionan al usuario mediante un servidor remoto. En tal caso, los transductores de entrada y salida 159 y 156 podrían proporcionarse en el extremo del usuario mediante teléfonos o micrófonos y altavoces conectados al sistema del PC del usuario, con señales de audio analógicas o digitales transmitidas entre el usuario y 100 a través de una red de sistema de teléfono y/o Internet. El usuario puede controlar remotamente la operación del sistema mediante numerosos procedimientos incluyendo un teclado táctil de un teléfono, un teclado de ordenador, entrada de voz, o por otros medios.

[0054] Una realización de esta invención en la forma de un sistema de karaoke de consumidor en tiempo no real permite a un miembro del público grabar su voz cantando una canción pop con un vídeo musical en un sistema basado en un ordenador. Cuando la voz grabada del usuario se modifica y posteriormente se reproduce, la voz modificada sincroniza los labios con los movimientos de la boca del cantante original y tiene la misma variación de tono que la voz del cantante reemplazado en el vídeo musical. El sistema de la figura 2 permite la reproducción de audio del intérprete original que canta una canción con o sin vídeo que lo acompañe. El usuario puede reproducir la canción y el sistema digitalizará y grabará (guardará) la voz del usuario en el disco duro del ordenador u otro dispositivo de memoria. Como hay un requisito para medir con precisión las características de la voz del cantante original, es mejor tener esa señal de la voz separada de la pista de música de fondo. Esto se puede conseguir de una manera más eficaz mediante la solicitud de una grabación aislada de la voz a la compañía u organización de grabación que proporciona el contenido de la grabación.

[0055] En la presente realización se utiliza una primera señal, la Señal de Guía, que es una grabación digitalizada del cantante que interpreta una canción de forma aislada (por ejemplo, la pista vocal solamente transferida desde una grabación de múltiples pistas de la sesión de grabación original), preferiblemente sin procesamiento añadido, tal como eco o reverberación. Estas Señales de Guía digitalizadas, $g(n)$, se pueden proporcionar al sistema del usuario en CD o DVD-ROM 165 o a través de Internet 175. Alternativamente, en otras realizaciones, las características requeridas de una Señal de Guía (tanto para la alineación en el tiempo como para controlar la modificación de las características) se pueden analizar previamente en el mismo u otro sistema para extraer los datos requeridos. Estos datos se pueden introducir al sistema 100 para su uso como archivos de datos a través de 165, 175 o a través de otros procedimientos de transferencia de datos. Los almacenes de datos y los módulos de procesamiento de la realización se muestran en la figura 3.

[0056] El usuario, al ejecutar el programa de grabación de sonido y de reproducción, reproduce la canción deseada con el cantante original audible o no audible y canta al mismo tiempo. El canto del usuario se digitaliza y se graba en un archivo de datos en un almacén de datos 310. Esta señal

digitalizada es la segunda señal, es decir, la Nueva Señal, $s(n)$.

[0057] La realización de la figura 3 realiza el Procedimiento 1 descrito anteriormente. El objetivo es corregir el tono y la temporización de la Nueva Señal del usuario para imitar el tono y la temporización de la Señal de Guía. En este caso, la función de la Señal de Guía que se utiliza como función de control y la característica que se modifica en la Nueva Señal es la misma característica, es decir, el contorno de tono de la señal respectiva. Un proceso de seguimiento de las diferencias entre las mediciones del tono de la Nueva Señal alineada en el tiempo y las mediciones del tono de la Señal de Guía se utiliza para calcular una función de ajuste del tono para hacer un tono de la Nueva Señal modificado que sigue a la Señal de Guía. Se asume aquí que la Nueva Señal, $s(n)$ es similar en el fraseo, el contenido y la duración de la Señal de Guía, $g(n)$. Para una aplicación de tipo karaoke en tiempo no real, esto es una suposición razonable, porque el usuario está normalmente tratando de imitar la interpretación original vocal en el tiempo, el tono y las palabras.

[0058] El Procedimiento 1 se realiza aquí sobre los datos de audio digitales en tiempo no real de la siguiente manera.

15 Descripción y Medición de la Señal de Entrada

[0059] Es muy poco probable que la Nueva Señal y la Señal de Guía estén adecuadamente alineadas en el tiempo sin procesamiento. La patente US 4591928 (Bloom et al.) describe las diferencias entre los patrones de energía de señales de habla no alineadas en el tiempo, pero similares, y el uso de mediciones relacionadas con la energía, tales como salidas de banco de filtros como entrada a un proceso de alineación en el tiempo.

[0060] La figura 4 muestra una serie de tiempo $P_g(M)$ indicada en lo sucesivo como un contorno de tono 401, obtenido mediante la medición del tono de una Señal de Guía de una cantante profesional femenina, en función del número de marco M de medición del tono, en el que $M = 0, 1, 2, \dots, N$, y una serie de tiempo $P_s(M)$ se muestra como un contorno de tono 402 de una Nueva Señal típica de aficionado (voz masculina) antes de la alineación en el tiempo a lo largo de la misma escala de tiempo. Las diferencias en los contornos de tono de ambas señales, así como su desalineación en el tiempo son evidentes. La primera serie, $P_g(M)$, que no está alineada en el tiempo con la segunda serie, $P_s(M)$, no puede utilizarse directamente como una función de control o de tono objetivo para la segunda señal sin generar errores significativos y audibles.

[0061] Un punto de datos mostrado como cero HZ en un contorno de tono 401 ó 402 indica que el marco de medición de tono correspondiente contiene silencio o voz sorda. Las mediciones que no son cero indican la medición del tono de la señal respectiva en ese marco.

[0062] En la figura 4, los segmentos de valor que no es cero (pulsos) de sonido de voz en el contorno de tono de la Nueva Señal 402 generalmente van detrás de las características correspondientes en el contorno de tono de la Señal de Guía 401 y tienen duraciones diferentes. También los sonidos de voz de dos contornos de tono están en octavas diferentes. Además, la variación de rango de tono en cada pulso del contorno de tono de la Señal de Guía 401 es mucho más amplia que en el pulso correspondiente en el contorno de tono de la Nueva Señal 402. Esto se espera porque el contorno de tono de la Señal de Guía 401 se toma de un cantante profesional. Son estos detalles y la temporización del contorno de tono de la Señal de Guía 401 los que se imparten al canto grabado del usuario aficionado.

Alineación en el Tiempo de la Nueva Señal

[0063] En la figura 3, en la forma de onda de la Nueva Señal de muestra, $s(n)$, leída desde el almacén de datos 310, se alinea primero en el tiempo con la Señal de Guía, $g(n)$, leída desde el almacén de datos 312, utilizando una técnica tal como la descrita en la patente US 4591928 para crear una señal de audio intermedia, la Nueva Señal alineada en el tiempo, $s'(n)$, que se almacena, por ejemplo, en el disco 330. Esto asegura que los detalles de los patrones de energía en $s'(n)$ se producen a los mismos momentos relativos que los de la Señal de Guía. Además, se asegura que cualquier sincronización de los labios requerida será efectiva y cualquier transferencia de características de la Señal de Guía a la Nueva Señal no necesita ninguna correlación de tiempo adicional. La frecuencia de muestreo utilizada en la creación de la Nueva Señal $s(n)$ y la Señal de Guía $g(n)$ en este ejemplo es de 44,1 kHz.

[0064] El proceso de alineación en el tiempo descrito en la patente US 4.591.928 mide las características de energía espectral (por ejemplo, una salida filterbank) cada 10 ms, y genera una alineación en el tiempo o la trayectoria de "deformación del tiempo" con un punto de trayectoria cada 10 ms que asocia características espectrales similares en la Nueva Señal con las características correspondientes más cercanas en la Señal de Guía.

[0065] La figura 5 muestra un ejemplo de una trayectoria de deformación del tiempo, $w(k)$, $k = 0, 1, 2, \dots$ en el que cada marco de características de la Nueva Señal tiene un número de marco j y cada marco de características de la Señal de Guía tiene un número de marco k , siendo el intervalo de muestras

- de marco de T segundos, donde $T = 10$ ms. Esta trayectoria de deformación se crea dentro de un módulo de procesamiento de alineación con el tiempo 320, y esta trayectoria se utiliza para controlar la edición (es decir, Compresión del Tiempo/Expansión) de la Nueva Señal $s(n)$ en el módulo 320 en la creación de la Nueva Señal alineada en el tiempo $s'(n)$ almacenada en el disco 330. Tal como se describe en la patente 4.591.928, la Nueva Señal alineada en el tiempo, $s'(n)$, se crea mediante el módulo 320 creando una versión editada de $s(n)$ en la que porciones de $s(n)$ se han repetido o eliminado según $w(k)$ y retroalimentación de errores de la temporización adicional desde el sistema de edición, que está restringida a hacer ediciones de tono sincrónicas cuando hay sonido de voz.

Generar el Contorno del Tono de la Nueva Señal

- 10 **[0066]** Un contorno de tono en bruto, $Ps'(M)$, de la Nueva Señal alineada, $s'(n)$, se crea a partir de mediciones de $s'(n)$ tomadas usando un análisis de movimiento en la ventana Hann en marcos de medición de tono discretos consecutivos, donde M es el número de marco y $M = 1, 2, 3, \dots$. Para obtener mediciones precisas de tono, se recomienda que la longitud de la ventana de análisis sea de 2,5 a 3,0 veces la longitud del período menor que se mide. Por lo tanto, en la realización actual, para medir el tono
- 15 tan bajo como de 72 Hz, con un período de aproximadamente 0,0139 s, se usa una ventana de análisis de muestra 1536 (a 44,1 kHz de frecuencia de muestreo) (o aproximadamente 35 ms). El intervalo de muestreo de un marco de medición de campo es de 10 ms. La ventana de análisis del módulo estimador de tono 340 se centra en cada marco de medición de tono de las muestras. Para cada marco de medición de tono, se hace una estimación del tono usando uno de los procedimientos bien conocidos para la
- 20 estimación de tono (por ejemplo, correlación automática, filtro de peine, etc.). Descripciones detalladas de estas técnicas se pueden encontrar en referencias tales como Wolfgang Hess (1983) "Pitch Determination of Speech Signals, Algorithms and Devices", Springer-Verlag, R.J. McAulay y T.F. Quatieri. (1990); "Pitch estimation and voicing detection based on a sinusoidal model", Proc. Int. Conf. on Acoustics, Speech and Signal Processing, Albuquerque, NM, páginas 249-252, y T.F. Quatieri (2002) "Discrete-Time Speech Signal Processing: Principles and Practice", Prentice Hall.

- 25 **[0067]** Las mediciones se pueden tomar sin la superposición de las ventanas de análisis, pero la superposición de datos de ventanas sucesivas de entre el 25 y el 50% se recomienda generalmente. En esta realización, la velocidad de marcos de medición de M es de 100 Hz (es decir, intervalos de 10 ms), que proporciona una superposición suficiente y también convenientemente es la misma que la
- 30 velocidad de medición de la función de alineación en el tiempo. Con el fin de realizar la primera y las pocas últimas mediciones del tono correctamente, en las que la ventana de análisis se extiende necesariamente más allá de las muestras de datos disponibles, tanto en el inicio como el final de la señal se rellenan con hasta una longitud de una ventana de análisis de muestras de magnitud cero antes de tomar esas mediciones del tono.

- 35 **[0068]** Para crear un contorno de tono suavizado final, $P's'(M)$ para la Nueva Señal alineada en el tiempo, las mediciones de tono de los marcos individuales se suavizan en un módulo de filtro 350 usando un filtro mediano de 3 puntos seguido por un filtro de promedio. Además, los marcos de silencio y sordos de las Nuevas Señales alineadas en el tiempo $s'(n)$ están marcados en $P's'(M)$ como que tienen un tono cero.

40 Generar el Contorno del Tono de Guía

- [0069]** De manera similar, en un módulo estimador de tono 345 se crea un contorno de tono $Pg(M)$ de la Señal de Guía $g(n)$, utilizando los mismos procedimientos y parámetros que los descritos para crear el contorno de tono $Ps'(M)$, y se suaviza en un módulo de filtro 355 para crear un contorno de tono suavizado $P'g(M)$ para la Señal de Guía.

45 Calcular el Ajuste del Tono

- [0070]** El proceso siguiente es el cálculo del ajuste del tono o el factor de corrección para cada marco de la Nueva Señal alineada con el tiempo. Esto se hace mediante un módulo de ajuste de tono 370 y tiene en cuenta la relación entre el tono de la Señal de Guía con el tono de la Nueva Señal alineada en el tiempo y cualquier cambio deseado en la octava. El cálculo se realiza para cada par de marcos de
- 50 medición de tono que tienen el mismo número de marco M. Un filtro de paso bajo en el módulo 370 suaviza entonces los factores de corrección. Hay dos etapas: determinación de octava y cambio de tono de la Nueva Señal. Hay dos opciones principales que se consideran en relación con el ajuste del tono: a) el ajuste del tono de salida para ser el mismo que el tono de la Señal de Guía o b) mantener el rango del tono de la Nueva Señal de entrada de manera que la voz ajustada suene lo más natural. El ajuste de la octava para lograr este último efecto se describirá ahora. Un módulo de ajuste de octava 358 calcula un
- 55 multiplicador de octavas, Q, que se mantiene constante durante la duración de la señal. Esto pone de relieve la necesidad de analizar la totalidad o al menos una cantidad substancial de la Nueva Señal antes de poder establecer este valor. Para cada marco de análisis del tono M de la Nueva Señal alineada en el tiempo, las estimaciones del tono sin suavizar para el marco M desde los módulos estimadores de tono
- 60 340 y 345 se utilizan para calcular una corrección de tono local, $C_L(M)$, donde M es el número de marco, limitando el cálculo de aquellos marcos donde se expresan la Nueva Señal alineada en el tiempo y su

correspondiente marco de Señal de Guía, es decir, ambos marcos tienen un tono válido. En esos marcos, el factor de corrección de tono local $C_L(M)$, que haría el tono del marco M de la Nueva Señal alineada en el tiempo el mismo que el tono del marco M de la Señal de Guía, viene dada por

$$C_L(M) = P_g(M)/P_s'(M) \quad (1)$$

- 5 **[0071]** Cada relación $C_L(M)$ se redondea entonces a su octava más próxima seleccionando las potencias de 2 de acuerdo con la siguiente tabla:

Relación $C_L(M)$	Octava	Comentario
0,5. hasta 0,75	0,5	La Nueva Señal es una octava mayor
0,75 hasta 1,5	1,0	La Nueva Señal es la misma octava
1,5 hasta 3	2,0	La Nueva Señal es una octava menor
3,0 hasta 6,0 etc.	4,0	La Nueva Señal es dos octavas menor

- 10 **[0072]** Todos los valores resultantes de las octavas se introducen en un histograma y entonces se selecciona el valor de corrección de octava, Q, que se produce con mayor frecuencia. Q no es una función del tiempo en este caso, pero puede ser en realizaciones alternativas. Si se desea, Q podría multiplicarse por otro factor para alcanzar cualquier desplazamiento deseado en la frecuencia del tono. El cálculo de Q se realiza en un módulo 358. El valor de corrección de la octava Q se suministra a un módulo de ajuste del tono 370 y se utiliza en la ecuación (2) a continuación para producir un factor de corrección de tono corregido con octava, C(M) donde

15
$$C(M) = P'g(M)/(Q \cdot P's(M)) \quad (2)$$

donde

C(M) es el factor de corrección de tono en el marco M de las señales, y

P's'(M) y P'g(M) son el tono estimado suavizado en el marco M de la Nueva Señal alineada en el tiempo y la Señal de Guía, respectivamente.

- 20 **[0073]** Para generar una señal de corrección de tono, el factor de corrección de tono C(M) se calcula a partir de la ecuación (2) sobre todos los marcos de la Nueva Señal alineada en el tiempo, de modo que el registro del tono de la Nueva Señal alineada modificada en el tiempo se acerque más a el de la Nueva Señal original.

- 25 **[0074]** Si no existe ningún tono de la Señal de Guía correspondiente en un marco M, (es decir, la Señal de Guía es sorda o la Nueva Señal alineada en el tiempo es ligeramente más larga que la Señal de Guía) el último valor del factor de corrección en M-1 se vuelve a utilizar. También sería posible utilizar la extrapolación para obtener una mejor estimación en este caso.

- 30 **[0075]** Ejemplos de los valores de procesamiento de corrección resultantes son: un factor de corrección, C(M), de 1,0 significa que no hay cambios en s'(n) en el marco M; 0,5 significa bajar el tono en una octava, 2,0 significa elevar el tono en una octava, y así sucesivamente.

Cambio del tono de la Nueva Señal

- 35 **[0076]** Cada valor de C(M) en la señal de corrección del tono proporciona el multiplicador de corrección necesario para un marco M correspondiente de las muestras de la Nueva Señal alineada en el tiempo, s'(n). En este ejemplo, la velocidad del marco C(M) se elige para ser la misma que la utilizada mediante el algoritmo de alineación en el tiempo, que es de 100 marcos por segundo o fps. En otras palabras C(M) tendrá un centenar de muestras por cada segundo de s'(n).

- 40 **[0077]** Para funcionar correctamente, algunos algoritmos de cambio de tono deben tener una velocidad muy inferior a la del algoritmo de alineación con el tiempo; es decir, el intervalo de muestreo (marco de análisis) es mucho más largo. Por ejemplo, las técnicas de cambio de tono de dominio del tiempo suelen tener una velocidad de marco de 25 a 30 fps si van a trabajar hasta frecuencias de 50 a 60 Hz. Sin embargo, su velocidad de marco no tiene que ser constante a través de toda la señal, y la velocidad se puede variar, por ejemplo, con el tono fundamental de la señal s'(n). En la presente realización, sin embargo, se usa una velocidad de marcos fija en el cambio de tono.

[0078] En la presente realización, las respectivas velocidades de marco para el cálculo del factor de corrección de tono $C(M)$ y el funcionamiento del algoritmo de cambio de tono son diferentes, y por lo tanto, se usa interpolación lineal para derivar una estimación de la corrección de tono necesaria en el centro de cada marco de análisis del algoritmo de cambio de tono a partir de las muestras $C(M)$ más cercanas en el tiempo a ese centro. Este factor de corrección interpolado se deriva de la siguiente manera:

[0079] Un marco M de la señal de corrección de tono tiene una longitud igual a L_c muestras de la nueva señal $s(n)$, donde L_c está dada por:

$$L_c = \text{velocidad muestreo de la Nueva Señal } s(n) / \text{velocidad de marco de } C(M) \quad (3)$$

[0080] El número de muestras a lo largo de $s'(n)$ en el centro de cada uno de los marcos de análisis del algoritmo de cambio de tono en el que se requiere una estimación de la corrección de tono se determina de la siguiente manera.

[0081] Si $N_c(\text{Fps}-1)$ es el número de la muestra a lo largo de $s'(n)$ en el centro del marco de análisis del cambio de tono $\text{Fps}-1$, entonces la muestra número $N_c(\text{Fps})$ en el centro del marco siguiente, Fps , es:

$$N_c(\text{Fps}) = N_c(\text{Fps}-1) + L_s(\text{Fps}, \text{To}(\text{Fps}-1)) \quad (4)$$

donde:

Fps es el número de marco de análisis de cambio de tono, $\text{Fps} = 0, 1, 2, \dots$ y

$L_s(\text{Fps}, \text{To}(\text{fps}-1)) = \text{velocidad de muestreo de la Nueva Señal} / \text{velocidad de marco del algoritmo de cambio de tono}$.

[0082] En este caso general, L_s es una función del número del marco Fps $\text{To}(\text{Fps}-1)$, la duración del período del tono $\text{Fps}-1$, para permitir una velocidad de marco variable en el tiempo. En esta realización, L_s se mantiene constante y se ajusta a 1536 muestras, es decir, 34,83 ms.

[0083] Los valores iniciales para el número de muestras a lo largo de $s'(n)$ en los centros del marco de análisis de cambio de tono antes del primer marco calculado, $N_c(-1)$, y el primer marco calculado, $N_c(0)$, dependen del algoritmo de cambio de tono. En esta realización $N_c(-1) = 0,5 * \text{To}(-1)$ y $N_c(0) = 0$.

[0084] Usando $N_c(\text{Fps})$ y L_c , los números de marco de corrección de tono $F_c(M)$ de $C(M)$, que unen o incluyen la muestra en el centro de un marco de análisis específico Fps en el algoritmo de cambio de tono son:

$$F_c(\text{Fps}) = N_c(\text{Fps}) / L_c \quad (5)$$

donde:

$/$ representa división entera,

$F_c(\text{Fps})$ es el marco de $C(M)$ que se produce justo antes o en el centro del marco del algoritmo de cambio de tono Fps , y

L_c es tal como se definió anteriormente.

[0085] Si $F_c(\text{Fps})$ es el marco de corrección de tono que se produce justo antes o en el centro del marco del algoritmo de cambio de tono entonces $(F_c(\text{Fps})+1)$ será el siguiente marco de corrección de tono que se produce después de su centro.

[0086] La interpolación lineal entre las correcciones de tono $C(F_c(\text{Fps}))$ y $C(F_c(\text{Fps})+1)$ proporciona un factor de corrección interpolado $C_s(\text{Fps})$ en el centro del marco de análisis de cambio del tono para controlar el cambio de tono:

$$C_s(\text{Fps}) = C(F_c(\text{Fps})) * (1 - \alpha) + \alpha * C(F_c(\text{Fps})+1) \quad (6)$$

donde:

$$\alpha = (N_c(\text{Fps}) - L_c * F_c(\text{Fps})) / L_c$$

y donde

$/$ representa la división entera,

y otros símbolos son como se describe anteriormente.

[0087] El valor del factor de corrección interpolado $Cs(Fps)$ se suaviza mediante un filtro simple de paso bajo para convertirse en $C'(Fps)$ y se representa como la salida del módulo 370 que se suministra al módulo cambiador de tono 380. Para la corrección del tono, la Nueva Señal alineada en el tiempo $s''(n)$ se procesa en marcos Fps correspondientes a los marcos del algoritmo de cambio de tono. Cada uno de estos marcos, Fps , de la Nueva Señal alineada en el tiempo $s'(n)$ varía de forma dinámica en el tono mediante su factor de corrección de suavizado en el módulo 380 y la Nueva Señal con el tono corregido y alineada en el tiempo resultante, $s''(n)$, se escribe en el disco 390 para la reproducción posterior con el fondo musical y, opcionalmente, el video musical correspondiente, si está disponible. Esta señal de salida, $s''(n)$ tendrá la alineación en el tiempo y la corrección de tono requeridas para reproducirse como un reemplazo para la Señal de Guía $g(n)$ o sincrónicamente con la misma. Un ejemplo del contorno alineado en el tiempo y con el tono corregido 701 que se observa en $s''(n)$ como resultado de multiplicar los valores del tono de la Nueva Señal alineada en el tiempo $s'(n)$ por los valores del factor de corrección correspondientes mostrados en la figura 6 se muestra en la figura 7. La mayoría de los detalles del contorno del tono de la Señal de Guía 401 aparecen ahora en este ejemplo de un contorno de tono modificado computarizado 701.

[0088] El cambio de tono realizado por el módulo 380 para crear la forma de onda de la señal de salida alineada en el tiempo y con el tono corregido, $s''(n)$ en el almacén 390 se puede conseguir utilizando cualquiera de los procedimientos estándar de cambio de tono, tales como TDHS, PS-OLA, FFT, que se describen en referencias tales como K. Lent (1989), "An efficient method for pitch shifting digitally sampled sounds", Computer Music Journal Vol. 13, N° 4, en las páginas 65 a 71;. N. Schnell, G. Peeters, S. Lemouton, P. Manoury, y X. Rodet (2000), "Synthesizing a choir in real-time using Pitch Synchronous Overlap Add (PSOLA)", International Computer Music Conference, en las páginas 102-108; J. Laroche y M. Dolson (1999), "New Phase-Vocoder Techniques for Pitch-Shifting, Harmonizing and other Exotic Effects", Proc IEEE 1999 Taller sobre aplicaciones de procesamiento de señales de audio y acústica en las páginas 91 a 94;. G. Peeters (1998), "Analyse-Synthese de sons musicaux par la methode PSOLA", Actas de la Jornadas de Informática Musical, Agelonde, Francia, y V. Goncharoff y P. Gries (1998), "An algorithm for accurately marking pitch pulses in speech signals", Actas de la IASTED Conferencia Internacional de Procesamiento de Señales e Imágenes (SIP'98), 28-31 de octubre.

[0089] En esta realización, un algoritmo de dominio de tiempo substancialmente tal como se describe en D. Malah (1979) "Time Domain Algorithms for Harmonic Bandwidth Reduction and Time Scaling of Speech Signals", IEEE Transactions Acoustics, Speech and Signal Processing, volumen 27, N° 2, páginas 121-133, se utiliza en el módulo 380 para cambiar el tono de la señal $s'(n)$.

[0090] En cada marco Fps de $s'(n)$ se mide el período del tono, que se define aquí como $To(Fps)$. Por motivos de simplicidad a partir de ahora, aunque las variables con base en cálculos que incluyen $To(Fps)$ también son variables de Fps , el parámetro Fps no se hace explícito en esas expresiones.

[0091] En esta realización, la Nueva Señal alineada en el tiempo $s'(n)$ se descompone en una secuencia de muestras de ventana $s'(u,n)$ de la señal multiplicando $s'(n)$ con $h(p)$, una función de ventana de análisis 801 (mostrada en la figura 10(a)) que cambia periódicamente en el tiempo, de modo que:

$$s'(u,n) = h(n) * s'(n-ta(u)) \quad (7)$$

donde

$h(p)$ es la ventana de análisis de cambio de tono para muestras de longitud P , cuya longitud en el tiempo es igual al doble del periodo de tono medido del marco Fps , es decir, $2 * To(Fps)$. En esta realización $h(p)$ es una ventana de Hann de P muestras.

$ta(u)$ es el caso de análisis u -ésimo que se fija en una velocidad de tono sincrónica para marcos de voz, tal que $ta(u) - ta(u-1) = To(Fps)$, donde $u = 0, 1, 2 \dots$. Para los marcos sordos $ta(u)$ se establece en una velocidad constante de 10 ms. También podría establecerse en el último valor válido de To desde un marco con voz.

[0092] A partir de la corrección de tono suavizada de $C's(Fps)$ se calcula el nuevo período de salida $To'(Fps)$ de la señal corregida. Para las señales sordas, en el marco Fps , $To'(Fps) = To(Fps)$. Para las señales con voz en el marco Fps ,

$$To'(Fps) = To(Fps)/C's(Fps) \quad (8)$$

[0093] A partir de este procesamiento, se genera una secuencia 802 de ventanas de síntesis a corto plazo $ts(v)$, que está sincronizada con el nuevo período de salida $To'(Fps)$ de tal manera que

$$ts(v)-ts(v-a) = To'(Fps) \quad (9)$$

donde:

ts(v) es el caso de síntesis v-ésimo en el marco de salida.

5 **[0094]** Tal como se muestra en las figuras 10 (a) y (b), se selecciona para cada ts(v) esa ventana ta(u) de datos s'(n) que es la más cercana en el tiempo. La ventana seleccionada ta(u) de datos s'(n) se añade entonces a una memoria intermedia de la secuencia de salida (no representada) para generar una corriente de señal de salida s''(n) un marco en un momento mediante el procedimiento conocido de superposición y adición que combina todas las ventanas de síntesis a corto plazo, ts(v) de un marco Fps. En efecto, las muestras de ventana s'(u,n) se recombinan con un período de tono de To'(Fps) más que con un período de To(Fps).

[0095] Otras realizaciones se describirán ahora.

10 **[0096]** Además del tono, que incluye curvas de vibrato y de inflexión, muchas otras características de las señales del sonido se pueden medir y se pueden modificar. Ejemplos son la sonoridad instantánea, las características glotales, los formantes del habla o los patrones de resonancia, ecualización, reverberación, y características de eco. Además, las Señales Nueva y de Guía no se limitan necesariamente a tener similitudes prosódicas, rítmicas y acústicas.

15 **[0097]** En la figura 8 se muestra una operación de análisis de características que actúa en la Nueva Señal y la Señal de Guía en los módulos 840 y 850, respectivamente, para crear fs(N) y fg(M). Estos se indican en negrita como vectores de características, especificando las características seleccionadas medidas en los marcos N y M, respectivamente. Estos vectores no tienen que ser de las mismas características. Aunque fg(M) debe contener al menos una característica, fs(N) puede, en una
20 realización adicional, ser un vector nulo y sin función.

[0098] Una función de ajuste de características, A(fs(N), fg(M), M), debe proporcionarse y aquí se entra al sistema como una especificación de procesamiento desde una fuente 865. Esta función define la relación deseada entre los vectores de características de las dos señales en los marcos N y M, donde éstos pueden o no ser el mismo marco, el tiempo transcurrido, tal como se representada por el parámetro
25 del marco M, y el proceso de modificación de la señal variable en el tiempo implementado en el software y aplicado en el módulo 870. Esta función y las variaciones generalmente se definirían y serían entradas por el programador del sistema y, en consecuencia, se puede presentar como un conjunto de preselecciones y/o variaciones de oferta definidas por el usuario que se pueden seleccionar por el usuario del sistema.

30 **[0099]** Un ejemplo del uso de dos características diferentes en A(fs(N), fg(M), M), es tener el volumen de la Señal de Guía controlando la frecuencia central de un proceso de filtro de paso de banda en movimiento en la Nueva Señal con la condición de que la Nueva Señal contenga energía en el movimiento de la banda del filtro de paso de banda. Hacer A una función de M también generaliza el proceso para incluir posibles modificaciones basadas en el tiempo a la función.

35 **[0100]** Otra forma de realización, que utiliza el Procedimiento 2 descrito anteriormente, se muestra en la figura 9A, en la que una forma de onda de la Nueva Señal no se genera como una primera etapa. En su lugar, los datos alineados en el tiempo, obtenidos como en la realización de las figuras 3 y 8 en un módulo 920, se utilizan para distorsionar el tiempo en un módulo 960, las características medidas de la Señal de Guía a los tiempos apropiados en la Nueva Señal. El módulo 970 hace las modificaciones
40 alineadas con el tiempo en la Nueva Señal. Una alineación en el tiempo opcional de se puede realizar en la Nueva Señal modificada en el módulo del proceso de modificación de las características 970 al mismo tiempo (combinando el procesamiento de los módulos 970 y 975 en un algoritmo), o en un módulo de proceso posterior 975 sobre la señal modificada de las características. Más detalles de este enfoque se proporcionan a continuación.

45 **[0101]** La inversa de la función de alineación en el tiempo en la figura 5 correlaciona los marcos coincidentes de la Señal de Guía en el marco k en cada marco de la Nueva Señal en el marco j. Si Fs es un número de marco de la Nueva Señal y W(Fs) es la función de deformación del tiempo (inversa) (o función de correlación) generada por el módulo de proceso de alineación en el tiempo 920 entonces

50
$$\mathbf{Fag(Fs)} = \mathbf{W(Fs)} \quad (10)$$

donde Fag es el número de marco correspondiente de la Guía alineada en el tiempo.

[0102] A partir de esta correlación se genera una versión alineada en el tiempo o distorsionada de la función de ajuste de características y se utiliza en el módulo de ajuste 960 en la figura 9A.

55 **[0103]** Por ejemplo, volviendo a la aplicación en la corrección de tono, una versión distorsionada de la función de corrección de tono, basada en la ecuación (1), se calcula como:

$$\mathbf{C(Fs)} = \mathbf{Pg(Fag(Fs))/Ps(Fs)} \quad (11)$$

[0104] A partir de (10) y (11)

$$C(Fs) = Pg(W(Fs))/Ps(Fs) \quad (12)$$

donde C(Fs) es el factor de corrección del marco Fs de la Nueva Señal.

[0105] Ps(Fs) es el tono estimado del marco Fs de la Nueva Señal. W(Fs) es el marco correspondiente en la Guía a partir de la función de deformación. El procesamiento adicional de C(Fs) tal como se describió previamente, incluyendo las modificaciones de octava (si se desea) se realiza en el módulo de ajuste 960 que entonces proporciona una función de modificación, basada en la ecuación (2), dada por

$$C(Fs) = P'g(W(Fs))/(Q*P's(Fs) \quad (13)$$

[0106] Esta función de modificación se aplica a s(n) en el módulo de modificación 970 en una base de marco a marco para producir una salida modificada, s*(n).

[0107] El procesamiento mostrado en la figura 9A se generaliza como en la descripción de la figura 8 para permitir que cualesquiera características de la señal que se especifican para el análisis y la modificación, pero se diferencia en que la salida modificada s*(n) en el almacén 980 no está alineada en el tiempo con la Señal de Guía, pero tiene en su lugar la temporización de la Nueva Señal original s(n). La alineación en el tiempo de la salida modificada s*(n) en la Señal de Guía g(n) se puede lograr para la modificación del tono en un proceso único, en el que las modificaciones de las características en el módulo 970 y la alineación con el tiempo en un módulo 975 se ejecutan de forma simultánea. Descripciones de los procedimientos de implementación, por ejemplo, de la modificación del tono y el tiempo simultáneas (lo que puede reducir artefactos potenciales de procesamiento y mejorar la eficiencia computacional) se encuentran en referencias tales como J. McAulay y T. Quatieri (1992), "Shape Invariant Time-Scale and Pitch Modification of Speech", IEEE Trans. Sig. Processing, IEEE Trans. Sig. Processing, marzo, vol. 40 N° 3, páginas 497-510 y D. O'Brien y A. Monaghan (1999), "Shape Invariant Pitch Modification of Speech Using a Harmonic Model", EuroSpeech 1999, páginas 1059-1062. Estas referencias asumen un cambio arbitrario de tono constante o un cambio de tono constante basado en mediciones de la señal original para determinar la cantidad de cambio a aplicar. Por ejemplo, si se detectan marcos sordos en la forma de onda de voz original, es una práctica normal desactivar, o al menos reducir, cualquier modificación en el tiempo o el tono aplicada durante ese marco.

[0108] Opcionalmente, la función de alineación en el tiempo normal, también se puede aplicar a un proceso de edición no lineal en el módulo 975 para crear una señal s*(n), que es una versión alineada en el tiempo de la Nueva Señal con las características modificadas s*(n).

[0109] Otra realización, que realiza el Procedimiento 3, se muestra en la figura 9B, en la que una señal alineada en el tiempo s'(n) en un módulo de almacenamiento 982 se crea mediante el módulo 975 utilizando la trayectoria original de alineación en el tiempo creada en el módulo 920. En esta disposición, un contorno de característica de la Nueva Señal se produce mediante el módulo 840 desde la Nueva Señal sin modificar s(n), y un contorno de característica de la Señal de Guía se produce mediante el módulo 850. En el módulo 960, la ecuación:

$$C(M) = P'g(M)/Q*P's(w(M)) \quad (14)$$

donde w(M) es la trayectoria de distorsión del tiempo generada por el módulo 920, se implementa para producir el contorno de modificación de características C(M). Este contorno de modificación se aplica en el módulo 972 a la Nueva Señal alineada en el tiempo para crear la Nueva Señal alineada en el tiempo y con características modificadas, s*(n), en el módulo de almacenamiento de salida 987.

[0110] En otras realizaciones, la Señal de Guía puede estar formada por una serie de diferentes señales individuales en lugar de una señal continua, o múltiples Señales de Guía (por ejemplo, armonías vocales) pueden utilizarse para generar múltiples partes vocales de una sola Nueva Señal.

[0111] En otras realizaciones, las características en la Nueva Señal no tienen que medirse o entrar en los cálculos de ajuste de las características de la Nueva Señal y simplemente se pueden modificar con base en mediciones de una característica o características de la Señal de Guía. Un ejemplo de esto podría ser la aplicación de reverberación o EQ a la Nueva Señal en función de esas características en la Señal de Guía.

[0112] Se apreciará que los módulos de procesamiento utilizados en las realizaciones descritas anteriormente serán módulos de software cuando se implementan en un sistema tal como el sistema 100 de las figuras 1 y 2, pero pueden ser, en implementaciones alternativas, módulos de hardware o una mezcla de módulos de hardware y software.

[0113] Una aplicación de la invención es la creación de archivos de sonido personalizados con una voz de usuario que puede proporcionar, por ejemplo, un tono de llamada de teléfono en un teléfono

móvil o un sistema telefónico basado en ordenador. Otros ejemplos incluyen la sustitución de cualquiera de los tonos de llamada u otros sonidos que se pueden presentar al que llama o al receptor de llamadas durante una llamada telefónica u otro intercambio de datos. Estos intercambios se pueden realizar a través de redes telefónicas, sistemas VoIP (Voz sobre Protocolo de Internet), u otro sistema de entrega de mensajes. Otros ejemplos incluyen la generación de archivos de sonido personalizados para cualquier dispositivo o sistema que pueda utilizar un mensaje personalizado pregrabado.

[0114] La figura 11 muestra una realización de la invención para permitir a un usuario generar, enviar y recibir estos archivos de sonido. En funcionamiento, el usuario inicia una llamada telefónica desde un teléfono fijo 1110 o un teléfono móvil 1120 a través de una red de telecomunicaciones 1140. Un convertidor apropiado 1150 recibe la señal de la red de telecomunicaciones 1140 y la convierte en señales de audio digital y tonos operativos de comando, y estos son procesados mediante un ordenador del servidor 1160. El ordenador del servidor 1160, opcionalmente, puede proporcionar respuesta de voz interactiva (IVR) desde un módulo 1165 para dar al usuario opciones e información de retorno sobre operaciones.

[0115] El ordenador del servidor 1160 se pueden implementarse en uno o más ordenadores e incorpora módulos de procesamiento de audio 1170 para la implementar los procesos tal como se describe en la figura 3 u 8 ó 9A ó 9B. El ordenador 1160 accede a un módulo de almacenamiento 1180 para almacenar archivos de música de audio y una base de datos para hacer referencia a esos archivos de canciones. El ordenador 1160 también almacena en un módulo de almacenamiento 1185 grabaciones de audio de usuario procesadas y originales y una base de datos para hacer referencia a esas grabaciones.

[0116] El ordenador del servidor 1160 interpreta las señales de comandos del teclado u otras para iniciar las operaciones. Por ejemplo, con el teclado del teléfono en esta implementación, el usuario puede indicar al ordenador 1160:

(a) Seleccionar una "pista", por ejemplo, una porción de una canción (almacenada en el módulo 1180);

(b) Transmitir la pista seleccionada a través del convertidor 1150 y la red 1140 al auricular del teléfono 1110 ó 1120 para que el usuario oiga y vuelva a oír.

(c) Registrar la voz del usuario, mientras que la pista seleccionada se vuelve a reproducir a través del auricular del teléfono 1110 ó 1120 y el usuario canta con el micrófono del auricular;

(d) Volver a reproducir el registro procesado de la voz del usuario mezclada con la pista de fondo apropiada (por ejemplo, una versión de la pista sin la voz del cantante original).

[0117] En la etapa (c), la voz del usuario se registra en el módulo de almacenamiento 1185, se procesa a través del módulo de procesamiento 1170, implementando el procesamiento, tal como se muestra en la figura 3 u 8 ó 9A ó 9B y el resultado se almacena en el módulo 1185.

[0118] Por último, el usuario introduce entonces un número de teléfono móvil del destinatario con el teclado de su teléfono 1110 ó 1120. El ordenador 1160 envía entonces un mensaje de datos al número del destinatario usando un sistema de entrega de tono de llamada 1190 tal como el sistema "WAP push". Este mensaje de datos da al destinatario la información necesaria para descargar el audio procesado a su teléfono móvil u otro dispositivo.

[0119] En una implementación alternativa, un ordenador del usuario 100 con micrófono 159 y altavoz 156 se utiliza para tener acceso al ordenador del servidor 1160 directamente a través de Internet 175 o mediante una llamada de teléfono utilizando software VOIP 1135. El usuario puede ir entonces por el mismo procedimiento descrito anteriormente, pero escucha y graba a través del ordenador 100 y envía comandos introducidos con el teclado 125 (no mostrado) del ordenador 100 al ordenador del servidor 1160. El usuario puede especificar por último un teléfono móvil mediante su número para recibir el archivo de sonido creado a través del sistema de entrega 1190. El archivo de sonido también se puede utilizar en el ordenador del usuario 100 u otro ordenador especificado (tal como el ordenador de un amigo) como tono de llamada u otro archivo de sonido de identificación en el sistema VOIP del ordenador especificado.

[0120] En otra implementación alternativa en la que el usuario accede al ordenador del servidor 1160 a través de Internet, algunos o todos los módulos de procesamiento de las figuras 3, u 8, o 9A o 9B se pueden descargar en el ordenador del usuario 100 tal como se representa por un módulo 1130. Un archivo de sonido resultante de la utilización del módulo 1130 con o sin la ayuda de un módulo de procesamiento de audio en el ordenador del servidor 1160 y almacenado en el ordenador del usuario 100 o en el módulo de almacenamiento 1185 se puede enviar a través de Internet 175 o la red de telecomunicaciones 1140 a un teléfono de destino solicitado u otro ordenador personal.

[0121] En otras realizaciones, los procesos se pueden implementar en su totalidad o en parte,

en teléfonos o cualquier otro dispositivo que contenga un sistema informático y memoria y los medios para la entrada y la salida de las señales de audio necesarias.

5 **[0122]** En una realización adicional, las señales de vídeo (tales como vídeos de música) puede proporcionarse desde el ordenador del servidor 1160 con los archivos de las canciones de audio que el usuario recibe. El usuario puede reproducir estas señales de audio y video y hacer grabaciones de sonido tal como se describió anteriormente. El archivo procesado, mezclado con la pista de fondo y el video sincronizado, se entrega al teléfono designado, ordenador personal u otro dispositivo capaz de reproducir un archivo de audio/visual.

10 **[0123]** Los archivos de audio de canciones no se limitan a canciones y pueden ser cualquier grabación de sonido, incluyendo el habla, efectos de sonido, música o cualquier combinación de los mismos.

REIVINDICACIONES

1. Procedimiento para modificar al menos una característica acústica de una señal de audio, comprendiendo el procedimiento:

5 comparar una primera y una segunda señales de audio de muestra para determinar los datos de alineación del tiempo de las diferencias de temporización entre los tiempos de aparición de características dependientes del tiempo en la segunda señal y los tiempos de aparición de características dependientes del tiempo en la primera señal; medir en posiciones seleccionadas a lo largo de la primera señal al menos una característica acústica de la primera señal para producir a partir de la misma una secuencia de mediciones de características de la primera señal;

10 procesar la secuencia de las mediciones de características de la primera señal para producir una secuencia de datos de modificación de características; y

 aplicar la secuencia de datos de modificación de características a la segunda señal para modificar al menos una característica acústica de porciones seleccionadas de la segunda señal de acuerdo con los datos de alineación en el tiempo.

15 2. Procedimiento según la reivindicación 1, en el que el procedimiento incluye la etapa de medir en posiciones seleccionadas a lo largo de la segunda señal dicha por lo menos una característica acústica de la segunda señal para producir a partir de la misma una secuencia de mediciones de características de la segunda señal, y la etapa de procesamiento de la secuencia de las mediciones de la primera señal incluye la comparación de las mediciones de características de la primera señal con las mediciones de características de la segunda señal y la determinación de los datos de modificación de las características a partir de dicha comparación.

20 3. Procedimiento según la reivindicación 1 ó 2, en el que dicha etapa de aplicar los datos de modificación de las características incluye las etapas de utilizar los datos de alineación en el tiempo para producir a partir de la segunda señal de muestra una segunda señal alineada en el tiempo y aplicar los datos de modificación de las características a la segunda señal alineada en el tiempo.

 4. Procedimiento según la reivindicación 2 ó 3, en el que dicha etapa de procesamiento incluye la etapa de usar los datos de alineación en el tiempo con las mediciones de características de la primera señal para producir los datos de modificación de características alineadas en el tiempo con las mediciones de características de la segunda señal.

30 5. Procedimiento según cualquier reivindicación anterior, en el que la etapa de aplicación de los datos de modificación de las características incluye la modulación de los datos de modificación de las características de acuerdo con una función predeterminada para modificar dicha por lo menos una característica acústica de dichas porciones seleccionadas de la segunda señal conjuntamente mediante los datos de modificación de las características y la función predeterminada.

35 6. Procedimiento según cualquier reivindicación anterior, en el que dicha al menos una característica acústica de la primera señal es el tono.

 7. Procedimiento según cualquier reivindicación anterior, en el que dicha al menos una característica acústica de la segunda señal es el tono.

40 8. Procedimiento según cualquier reivindicación anterior, en el que dichas características que dependen del tiempo de la primera y segunda señales son mediciones de energía espectral de muestra.

45 9. Procedimiento según la reivindicación 1, en el dicha al menos una característica acústica de la primera señal es el tono y dicha al menos una característica acústica de la segunda señal es el tono, y dicha etapa de procesamiento incluye la etapa de determinar, a partir de los valores de la relación de medición del tono de la primera señal para la medición del tono alineada en el tiempo de la segunda señal, un factor multiplicador e incluir así dicho factor en dicha etapa de aplicación de los datos de modificación de características para cambiar el rango de frecuencia de cambios de tono en la segunda señal en las porciones de señal seleccionadas modificadas.

50 10. Procedimiento según la reivindicación 9, que también incluye la etapa de escalado de dicho factor multiplicador por una potencia de 2 para cambiar el tono en dichas porciones de señal seleccionadas modificadas de acuerdo con una selección de dicha potencia de 2.

55 11. Procedimiento según la reivindicación 2, en el que la etapa de medir en posiciones seleccionadas a lo largo de la segunda señal incluye las etapas de usar los datos de alineación en el tiempo para producir a partir de la segunda señal de muestra una segunda señal alineada en el tiempo en la que los tiempos de aparición de dichas características que dependen del tiempo de la segunda señal de muestra son substancialmente coincidentes con los tiempos de aparición de dichas características

dependientes del tiempo en la primera señal de muestra, y la medición de la por lo menos una característica acústica en la segunda señal alineada en el tiempo en posiciones a lo largo de la segunda señal alineada en el tiempo seleccionada para estar relacionada en el tiempo con dichas posiciones seleccionadas a lo largo de la primera señal de muestra.

5 **12.** Procedimiento según la reivindicación 2, en el que dicho al menos una característica acústica de la primera señal de muestra es el tono, dicha al menos una característica acústica de la segunda señal de muestra es de tono, dicha etapa de aplicar los datos de modificación de características incluye las etapas de usar los datos de alineación en el tiempo para producir a partir de la segunda señal de muestra una segunda señal alineada en el tiempo y aplicar los datos de modificación de características de la segunda señal alineada en el tiempo para producir una segunda señal alineada en el tiempo con el tono modificado.

10 **13.** Procedimiento según la reivindicación 12, en el que la etapa de aplicar los datos de modificación de características incluye la modulación de los datos de modificación de características de acuerdo con una función predeterminada para modificar el tono en dichas porciones seleccionadas de la segunda señal conjuntamente mediante los datos de modificación de características y la función predeterminada.

15 **14.** Procedimiento según la reivindicación 13, en el que la función predeterminada es una función de los valores de la relación de la medición del tono en la primera señal de muestra y la medición del tono correspondiente en la segunda señal de muestra a lo largo de la segunda señal de muestra.

20 **15.** Procedimiento según la reivindicación 1, en el que dicha etapa de aplicación incluye producir así datos que representan una segunda señal modificada.

16. Procedimiento según la reivindicación 15, que también comprende la etapa de suministrar los datos que representan la segunda señal modificada al aparato de telecomunicaciones.

25 **17.** Procedimiento según la reivindicación 16, en el que dicha etapa de suministro incluye transmitir los datos que representan la segunda señal modificada a través de un sistema de suministro de tonos de llamada.

18. Aparato para modificar al menos una característica acústica de una señal de audio, comprendiendo el aparato:

30 medios para comparar una primera y una segunda señales de audio de muestra para determinar los datos de alineación en el tiempo de diferencias temporales entre los tiempos de aparición de características dependientes del tiempo en la segunda señal y los tiempos de aparición de características dependientes del tiempo en la primera señal;

35 medios para medir en posiciones seleccionadas a lo largo de la primera señal al menos una característica acústica de la primera señal para producir a partir de la misma una secuencia de las mediciones de características de la primera señal;

 medios para procesar la secuencia de las mediciones de las características de la primera señal para producir una secuencia de datos de modificación de características; y

40 medios para aplicar la secuencia de datos de modificación de características a la segunda señal para modificar al menos una característica acústica de porciones seleccionadas de la segunda señal de acuerdo con los datos de alineación en el tiempo.

45 **19.** Aparato según la reivindicación 18, que también incluye medios para medir en posiciones seleccionadas a lo largo de la segunda señal dicha por lo menos una característica acústica de la segunda señal para producir a partir de la misma una secuencia de las mediciones de características de la segunda señal, y en el que los medios para procesar la secuencia de las mediciones de la primera señal incluyen medios para comparar las mediciones de características de la primera señal con las mediciones de características de la segunda señal y determinar los datos de modificación de características a partir de esta comparación.

50 **20.** Aparato según la reivindicación 18 ó 19, en el que dichos medios para aplicar los datos de modificación de características incluye medios para utilizar los datos de alineación en el tiempo para producir a partir de la segunda señal de muestra una segunda señal alineada en el tiempo y aplicar los datos de modificación de características a la segunda señal alineada en el tiempo.

55 **21.** Aparato según la reivindicación 19 ó 20, en el que dichos medios de procesamiento incluyen medios para utilizar los datos de alineación en el tiempo con las mediciones de características de la primera señal para producir los datos de modificación de características en alineación con el tiempo con las mediciones de características de la segunda señal.

22. Aparato según la reivindicación 18, en el que los medios para aplicar los datos de modificación de características incluyen medios para modular los datos de modificación de características de acuerdo con una función predeterminada para modificar dicha por lo menos una característica acústica de dichas porciones seleccionadas de la segunda señal conjuntamente mediante los datos modificación de características y la función predeterminada.

23. Aparato según la reivindicación 18, en el que dicha al menos una característica acústica de la primera señal es el tono.

24. Aparato según la reivindicación 18, en el que dicha al menos una característica acústica de la segunda señal es el tono.

25. Aparato según la reivindicación 18, en el que dichas características que dependen del tiempo de la primera y segunda señales son mediciones espectrales de energía de muestra.

26. Aparato según la reivindicación 18, en el que dicho al menos una característica acústica de la primera señal es el tono y dicha al menos una característica acústica de la segunda señal es el tono, y dichos medios de procesamiento incluyen medios para determinar, a partir de los valores de la relación de la medición del tono de la primera señal y la medición del tono alineado en el tiempo de la segunda señal, un factor multiplicador y así incluir dicho factor en la aplicación de los datos de modificación de las características para modificar los cambios del rango de frecuencia del tono en la segunda señal en las porciones de la señal seleccionada modificada.

27. Aparato según la reivindicación 26, que también incluye medios para escalar dicho factor multiplicador por una potencia de 2 para cambiar el tono en las porciones de la segunda señal modificada seleccionada de acuerdo con una selección de dicha potencia de 2.

28. Aparato según la reivindicación 19, en el que los medios para medir en posiciones seleccionadas a lo largo de la segunda señal incluyen medios para utilizar los datos de alineación en el tiempo para producir a partir de la segunda señal de muestra una segunda señal alineada en el tiempo en la que los tiempos de aparición de dichas características que dependen del tiempo de la segunda señal de muestra son substancialmente coincidentes con los tiempos de aparición de dichas características dependientes del tiempo en la primera señal de muestra, y medios para medir la al menos una característica acústica en segunda señal alineada en el tiempo en posiciones a lo largo de la segunda señal alineada en el tiempo seleccionadas para estar relacionadas en el tiempo con dichas posiciones seleccionadas a lo largo de la primera señal de muestra.

29. Aparato según la reivindicación 28, en el que dichas posiciones seleccionadas que están relacionadas en el tiempo son sustancialmente coincidentes en el tiempo con dichas posiciones seleccionadas a lo largo de la primera señal de muestra.

30. Aparato según la reivindicación 19, en el que dicha al menos una característica acústica de la primera señal de muestra es el tono, dicha al menos una característica acústica de la segunda señal de muestra es el tono, dichos medios para aplicar los datos de modificación de características incluyen medios usar los datos de alineación en el tiempo para producir a partir de la segunda señal de muestra una segunda señal alineada en el tiempo y aplicar los datos de modificación de características en la segunda señal alineada en el tiempo para producir una segunda señal alineada en el tiempo con el tono modificado.

31. Aparato según la reivindicación 30, en el que los medios para aplicar los datos de modificación de características incluyen medios para modular los datos de modificación características de acuerdo con una función predeterminada para modificar el tono en dichas porciones seleccionadas de la segunda señal conjuntamente mediante los datos de modificación de las características y la función predeterminada.

32. Aparato según la reivindicación 31, en el que la función predeterminada es una función de los valores de la relación de medición del tono en la primera señal de muestra y la medición del tono correspondiente en la segunda señal de muestra a lo largo de la segunda señal de muestra.

33. Aparato según la reivindicación 18, en el que los medios de comparación comprenden un módulo de alineación del tiempo dispuesto para recibir una nueva señal de audio y una señal de audio de guía y para producir a partir de las mismas una nueva señal alineada en el tiempo;

los medios de medición comprenden un primer módulo de medición de las características acústicas dispuesto para recibir la señal de audio de guía y medir al menos una característica acústica de la señal de audio de guía;

los medios de procesamiento comprenden una calculadora de ajuste de las características acústicas acoplado al primer módulo de medición de características acústicas y dispuesto para calcular un factor de modificación de características acústicas; y los medios de aplicación

- comprenden un modulador de características acústicas acoplado al módulo de alineación en el tiempo para recibir la nueva señal alineada en el tiempo y la calculadora de ajuste de características acústicas para recibir el factor de modificación de características acústicas y dispuesto para modificar dicha al menos una característica acústica de la nueva señal alineada en el tiempo de acuerdo con el factor de modificación de características acústicas.
- 5
- 34.** Aparato según la reivindicación 33, en el que un módulo de funciones de procesamiento está acoplado a la calculadora de ajuste de características para suministrar al mismo una función de la señal, y la calculadora de ajuste de características está adaptada para calcular el factor de modificación de características acústicas en dependencia de la función de la señal.
- 10
- 35.** Aparato según la reivindicación 33 ó 34, en el que un segundo módulo de medición de características acústicas está acoplado al módulo de alineación en el tiempo y está dispuesto para medir por lo menos una característica acústica de la nueva señal alineada en el tiempo; y la calculadora de ajuste de las características acústicas está acoplada al segundo módulo de medición de las características acústicas.
- 15
- 36.** Aparato según la reivindicación 33, en el que un segundo módulo de medición de características acústicas está dispuesto para recibir la nueva señal de audio y para medir dicha por lo menos una característica acústica de la nueva señal de audio, y en el que la calculadora de ajuste de las características acústicas está acoplada al segundo módulo de medición de las características acústicas y al módulo de alineación con el tiempo y está adaptado para alinear las características acústicas medidas de la nueva señal de audio a las características acústicas medidas de la señal de audio de guía.
- 20
- 37.** Aparato según la reivindicación 18, en el que los medios de comparación comprenden un módulo de alineación en el tiempo dispuesto para recibir una nueva señal de audio y una señal de audio de guía y para producir a partir de las mismas datos de alineación en el tiempo;
- 25
- los medios de medición comprenden un primer módulo de medición de las características acústicas dispuesto para recibir la señal de audio de guía y medir al menos una característica acústica de la señal de audio de guía;
- los medios de procesamiento comprenden una calculadora de ajuste de las características acústicas acoplados al módulo de alineación en el tiempo y al primer módulo de medición de las características acústicas y dispuesto para calcular los valores alineados en el tiempo de un factor de modificación de características acústicas; y los medios de aplicación comprenden un modulador de características acústicas acoplado a recibir la nueva señal de audio y a la calculadora de ajuste de características acústicas para recibir los valores alineados en el tiempo del factor de modificaciones de las características acústicas y dispuesto para modificar dicha al menos una característica acústica de la nueva señal de audio, de acuerdo con los valores alineados en el tiempo del factor de modificación de las características acústicas para producir una nueva señal de audio modificada.
- 30
- 38.** Aparato modificación de señales de audio según la reivindicación 37, en el que un alineador de tiempo está acoplado con el modulador de características acústicas para recibir la nueva señal de audio modificada y con el módulo de alineación temporal para recibir los datos de alineación en el tiempo y está dispuesto para producir una nueva señal modificada alineada en el tiempo de acuerdo con dicha nueva señal de audio modificada y los datos de alineación en el tiempo.
- 40
- 39.** Aparato de modificación de señales de audio según la reivindicación 37 ó 38, en el que un segundo módulo de medición de las características acústicas está colocado para recibir la nueva señal de audio y medir al menos una característica acústica de la nueva señal de audio; y la calculadora de ajuste de las características acústicas está acoplada al segundo módulo de medición de las características acústicas.
- 45
- 40.** Aparato según la reivindicación 37, en el que el primer módulo de medición de características acústicas comprende un primer módulo de medición de tono acoplado al módulo de alineación en el tiempo y está colocado para medir el tono en la nueva señal alineada en el tiempo;
- 50
- un segundo módulo de medición del tono está dispuesto para recibir la señal de audio de guía y para medir el tono de la señal de audio de guía;
- la calculadora de ajuste de características acústicas comprende una calculadora de ajuste de tono acoplada al primer y segundo módulos de medición de tono y está dispuesta para calcular un factor de corrección de tono; y el modulador de características acústicas comprende un modulador de tono acoplado al módulo de alineación temporal para recibir la nueva señal alineada en el tiempo y la calculadora de ajuste de tono para recibir el factor de corrección de tono y está dispuesta para modificar el tono en la nueva señal alineada en el tiempo de acuerdo con el factor de corrección del tono.
- 55
- 41.** Aparato según la reivindicación 19, en el que dichos medios de comparación, dichos

medios de medición, dichos medios de procesamiento, y dichos medios de aplicación están incorporados en un aparato de telecomunicaciones.

42. Aparato según la reivindicación 41, en el que el aparato de telecomunicaciones comprende un ordenador de servidor adaptado para acoplarse a una red de telecomunicaciones.

5

43. Aparato según la reivindicación 41, en el que el aparato de telecomunicaciones comprende un teléfono móvil.

44. Aparato según la reivindicación 41, en el que el aparato de telecomunicaciones está adaptado para suministrar datos que representan una segunda señal modificada a un sistema de suministro de tonos de llamada.

10

45. Aparato según la reivindicación 43, en el que el teléfono móvil está adaptado para proporcionar datos que representan una segunda señal modificada a un sistema de suministro de tonos de llamada.

Figura 1

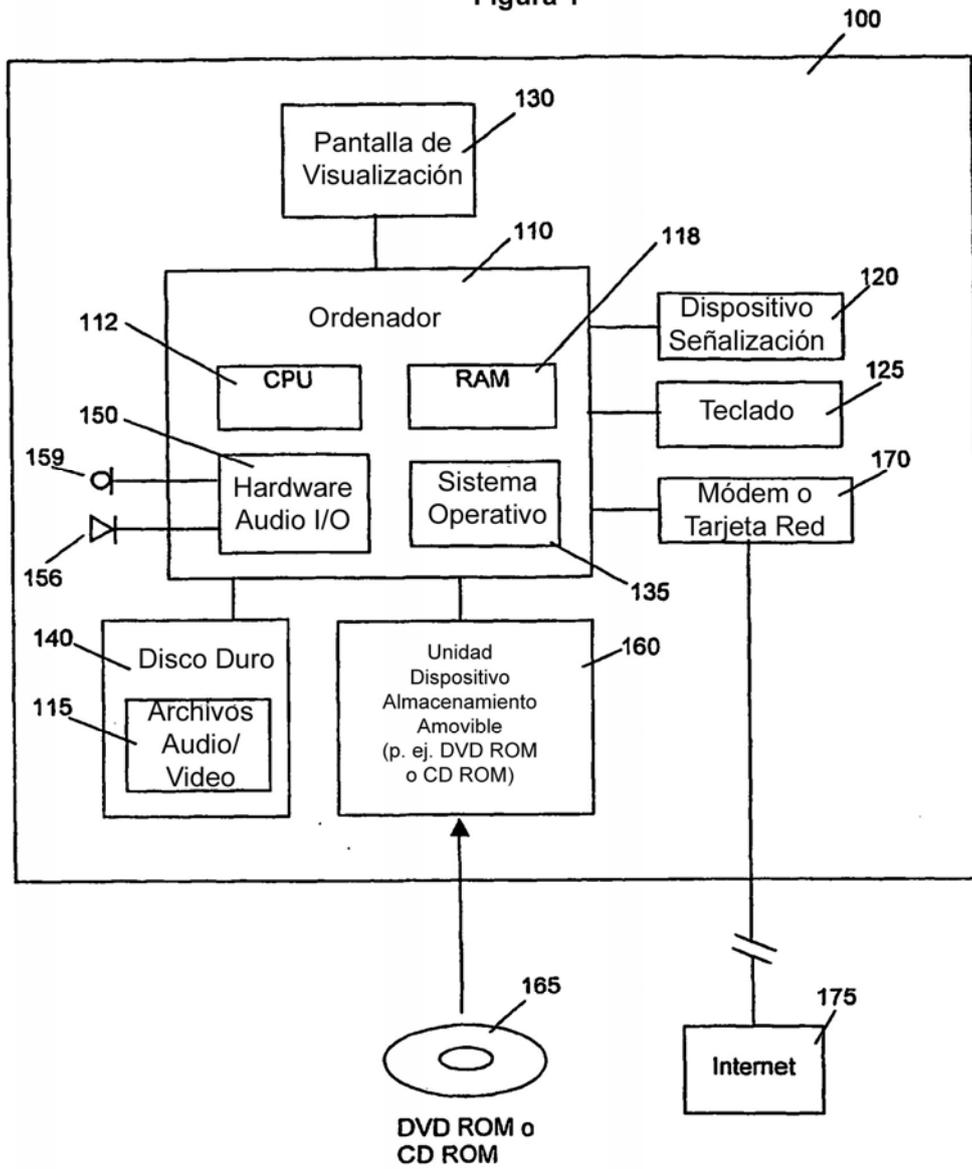
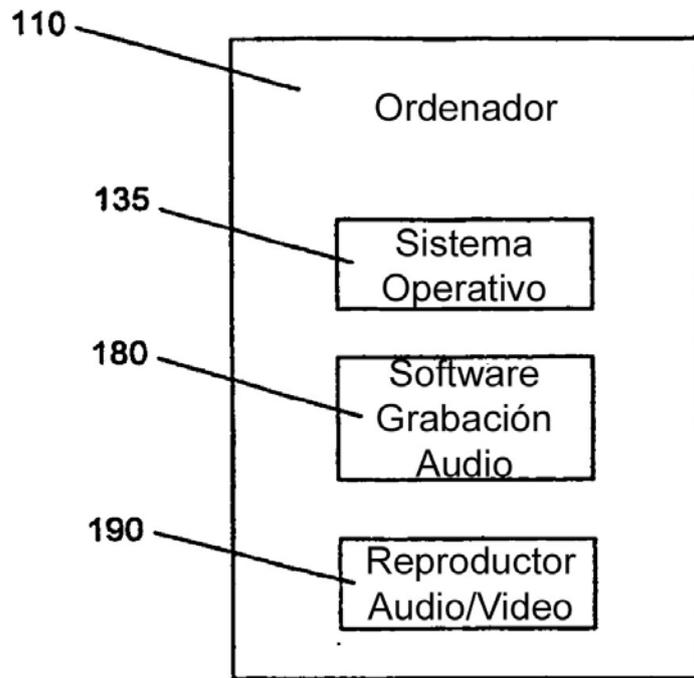


Figura 2



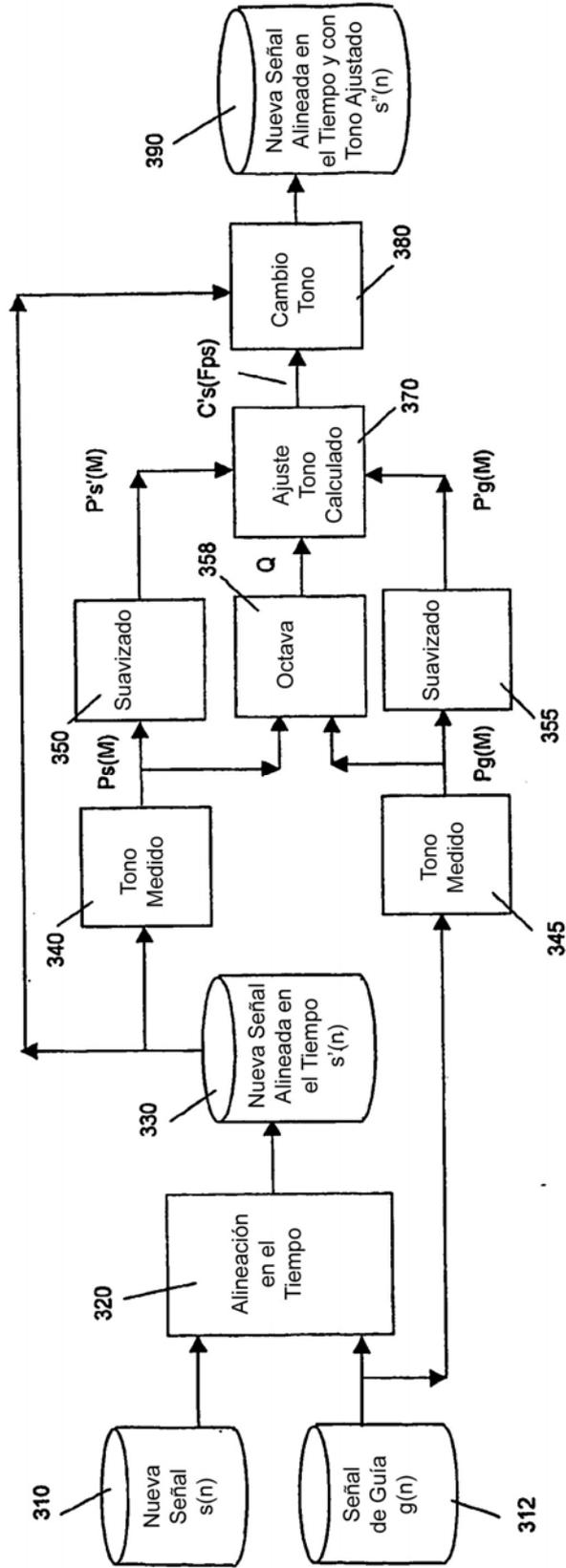


Figura 3

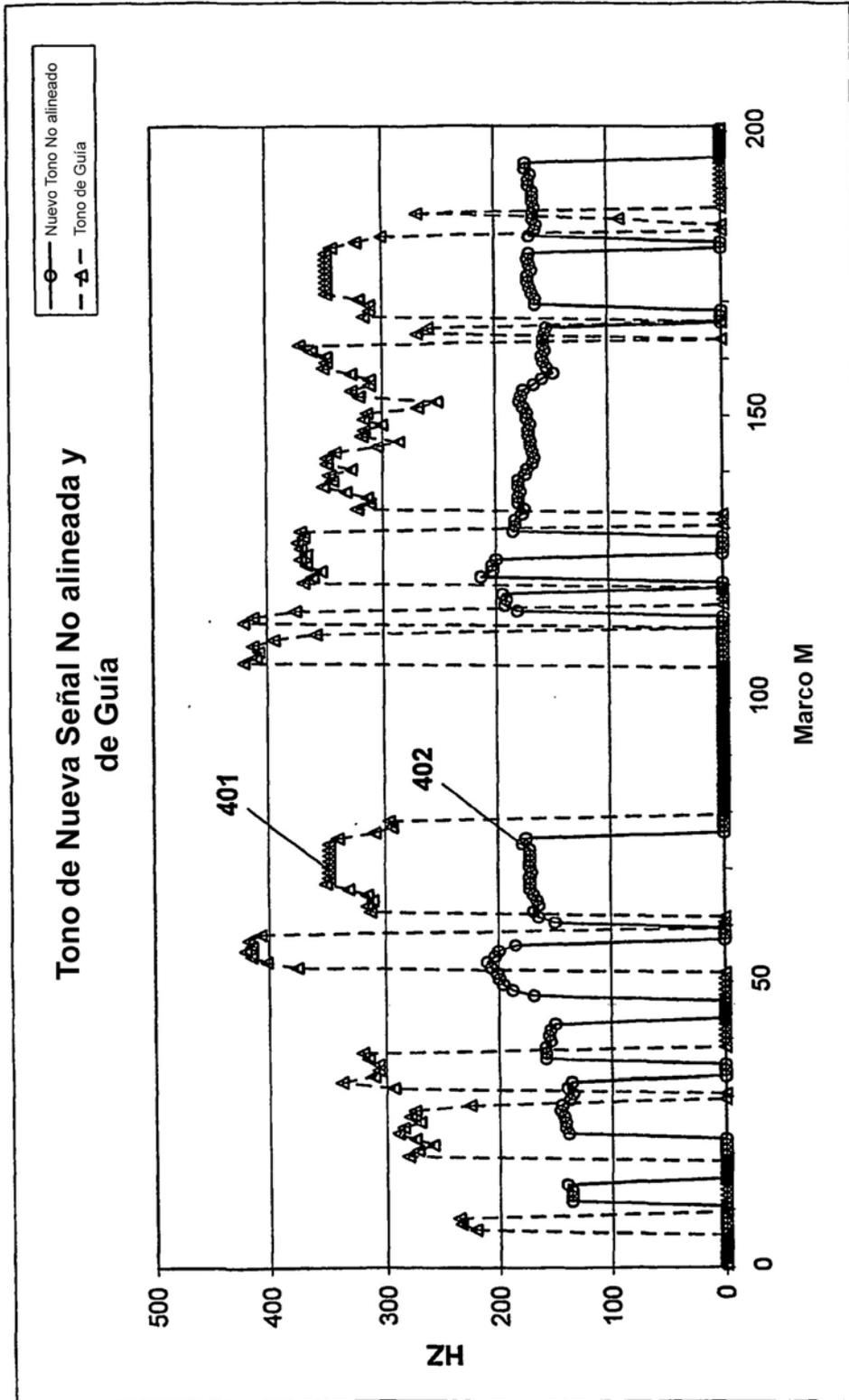


Figura 4

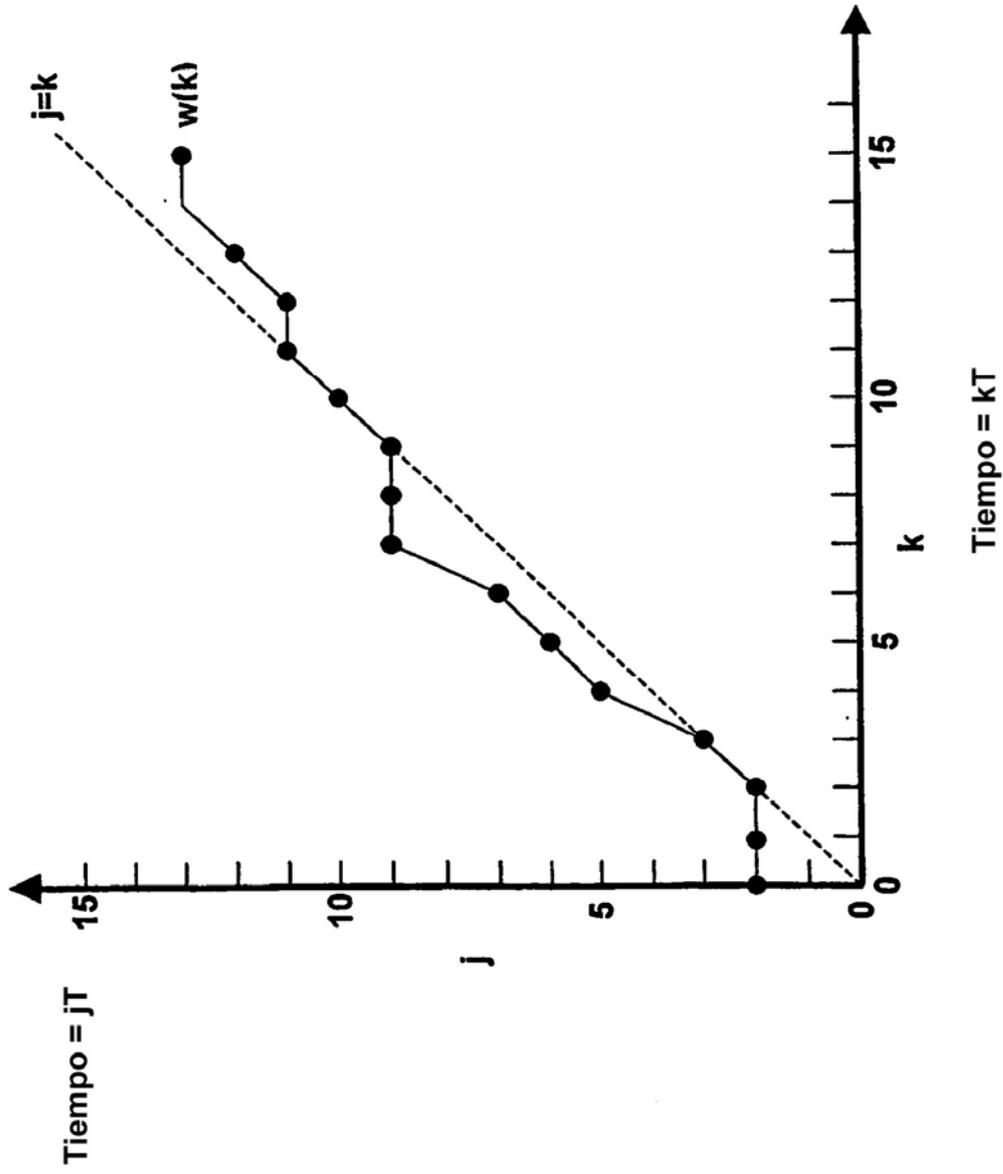


Figura 5

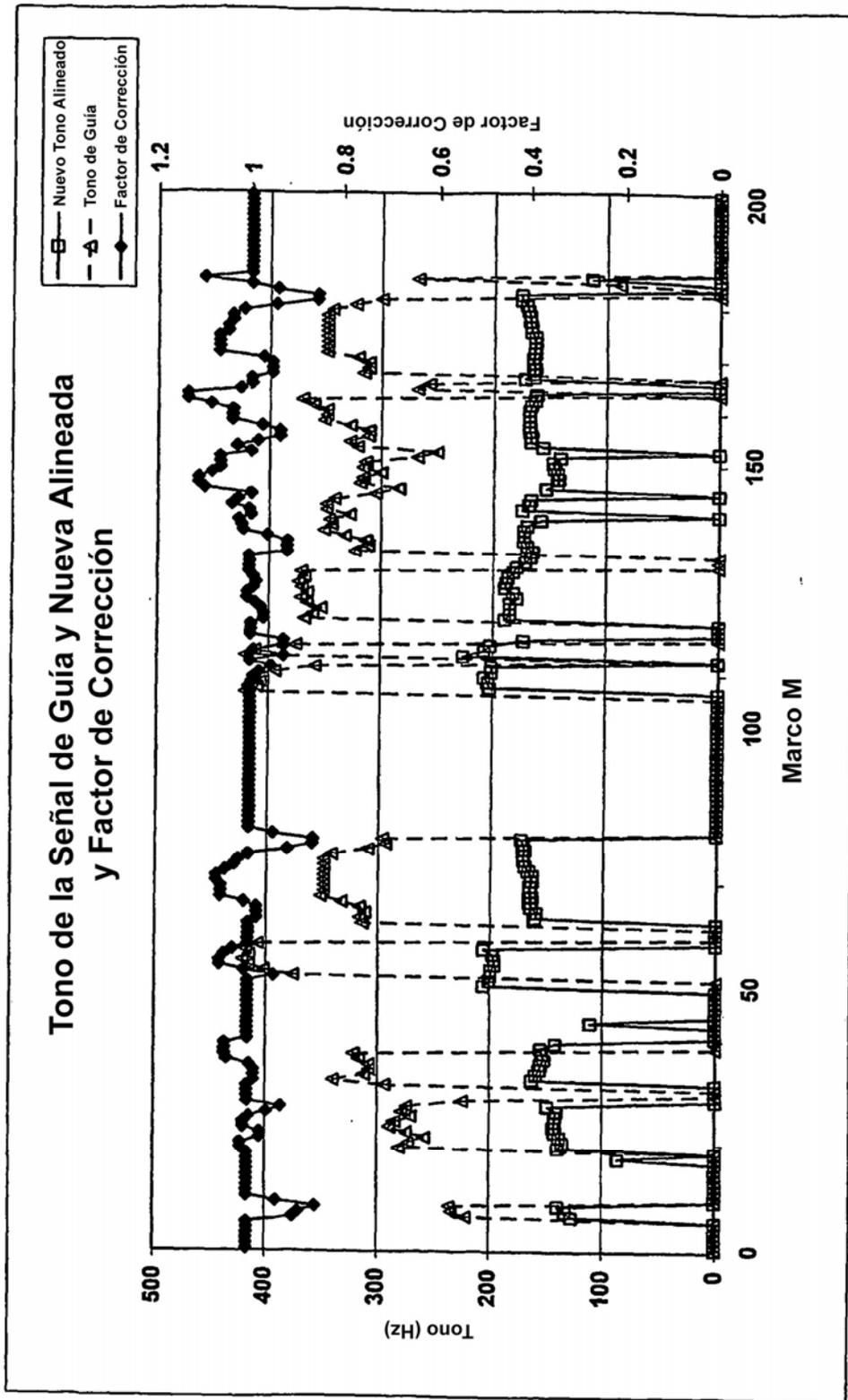


Figura 6

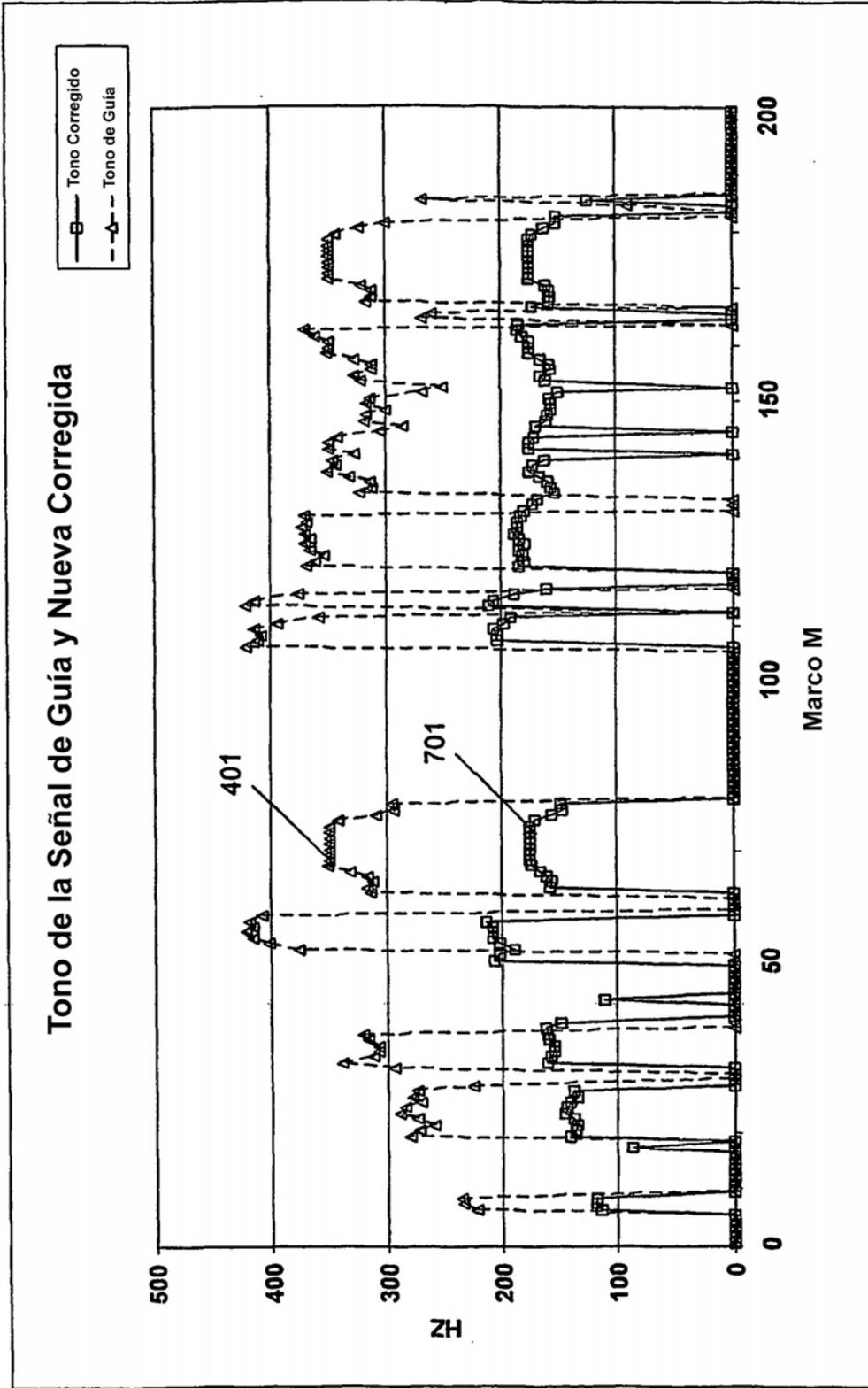


Figura 7

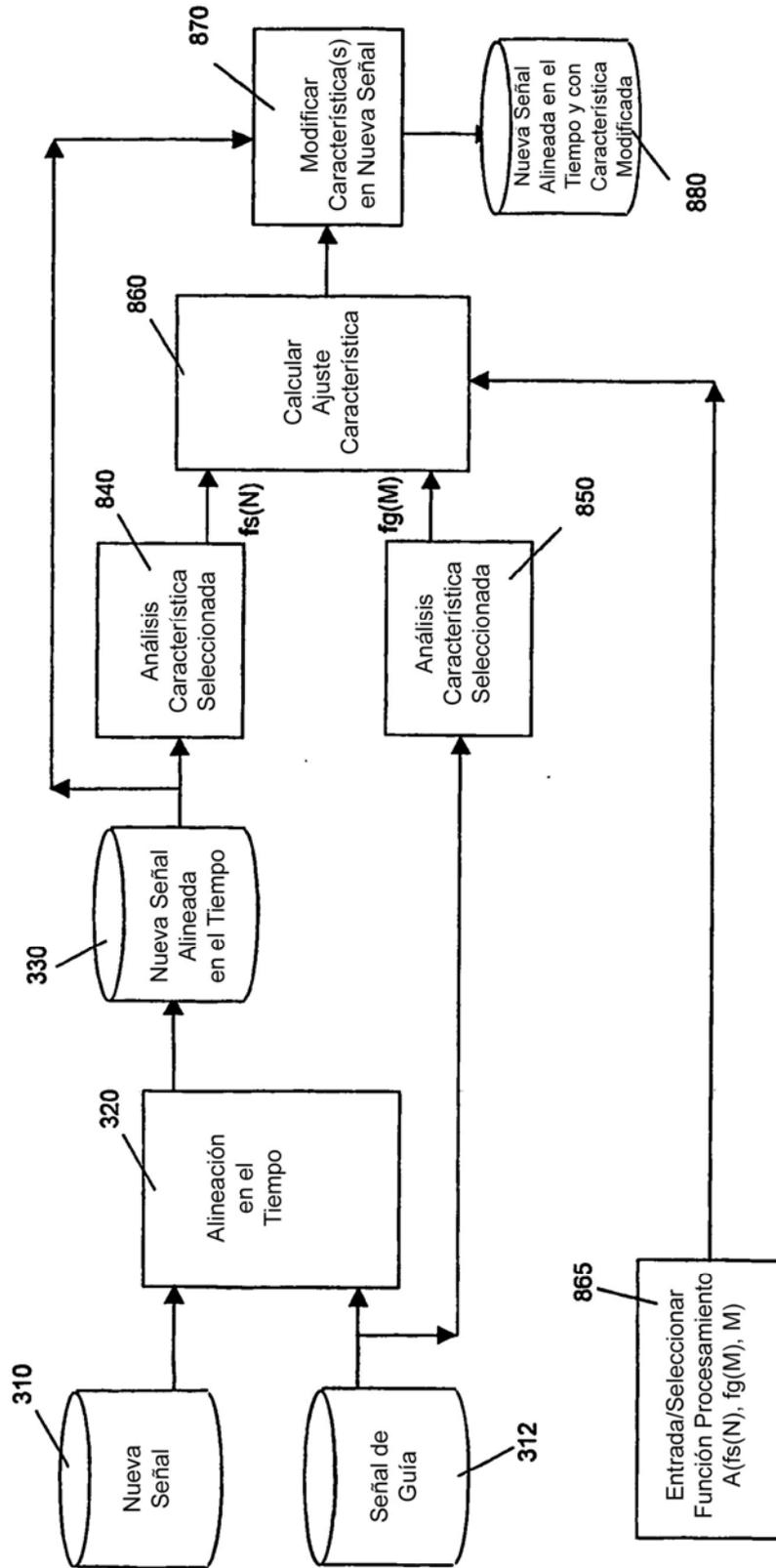


Figura 8

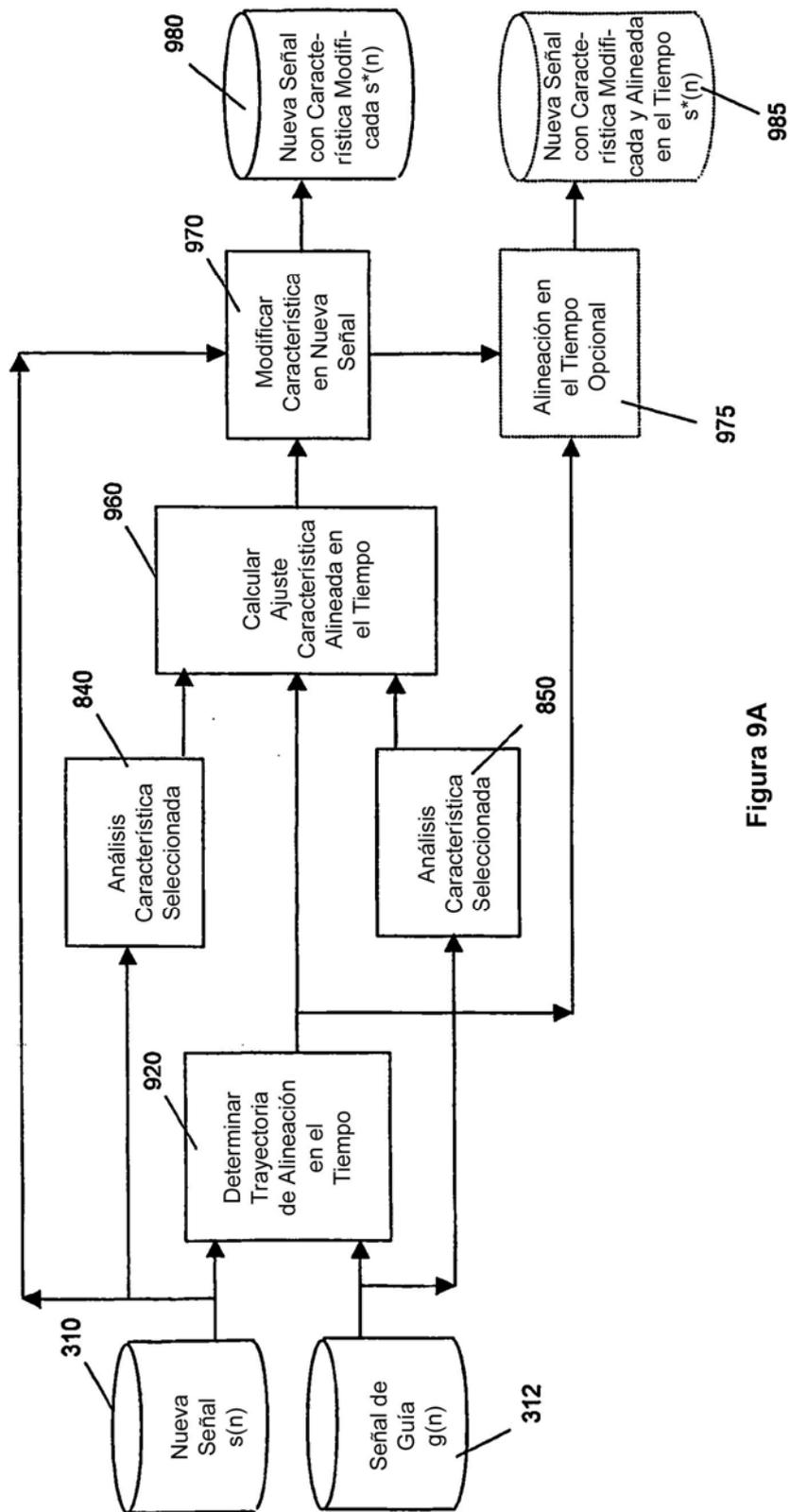


Figura 9A

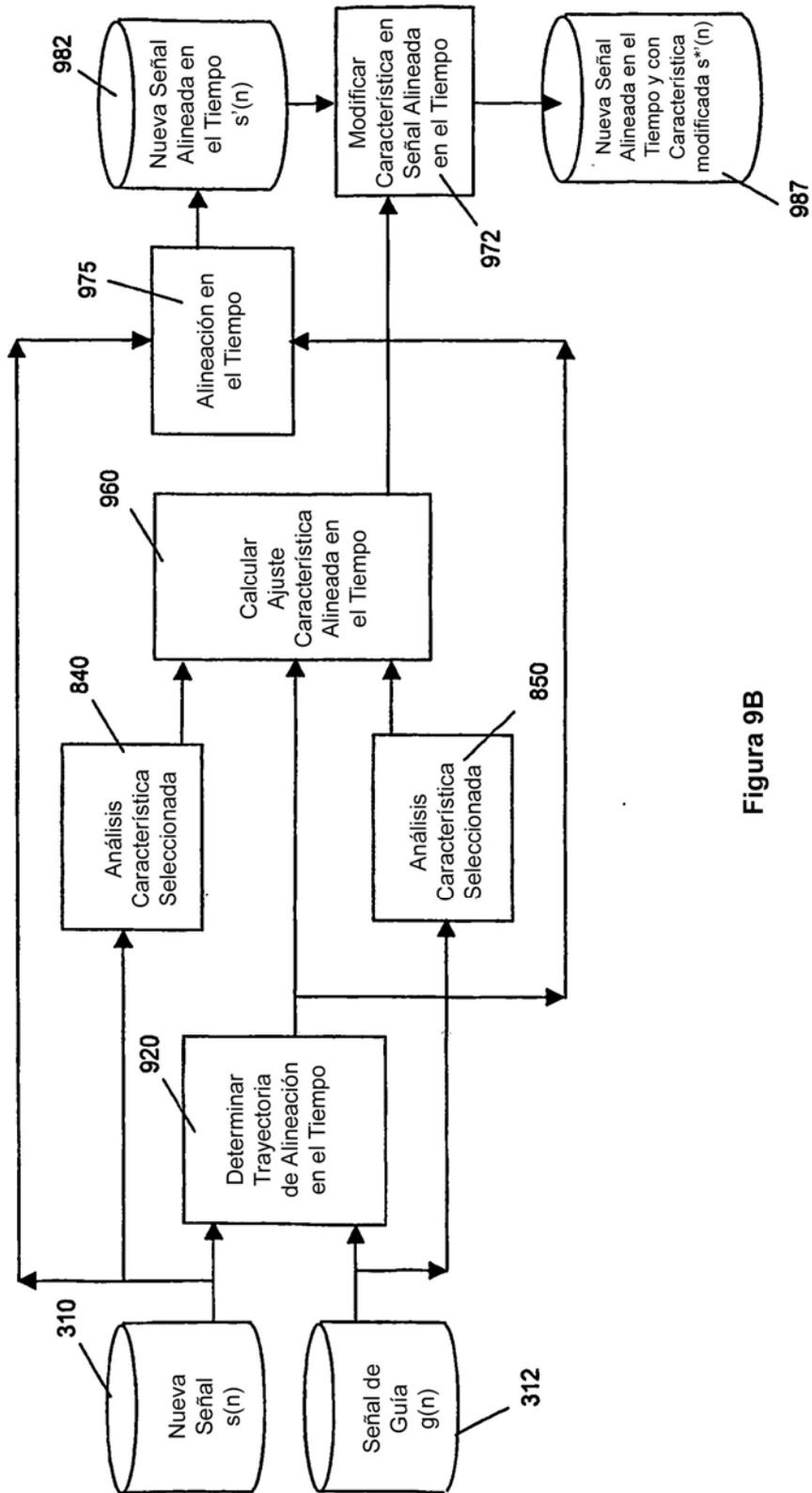


Figura 9B

Ventana de Correlación $ts(v)$ y Ventana $ta(u)$ más cercana en el tiempo

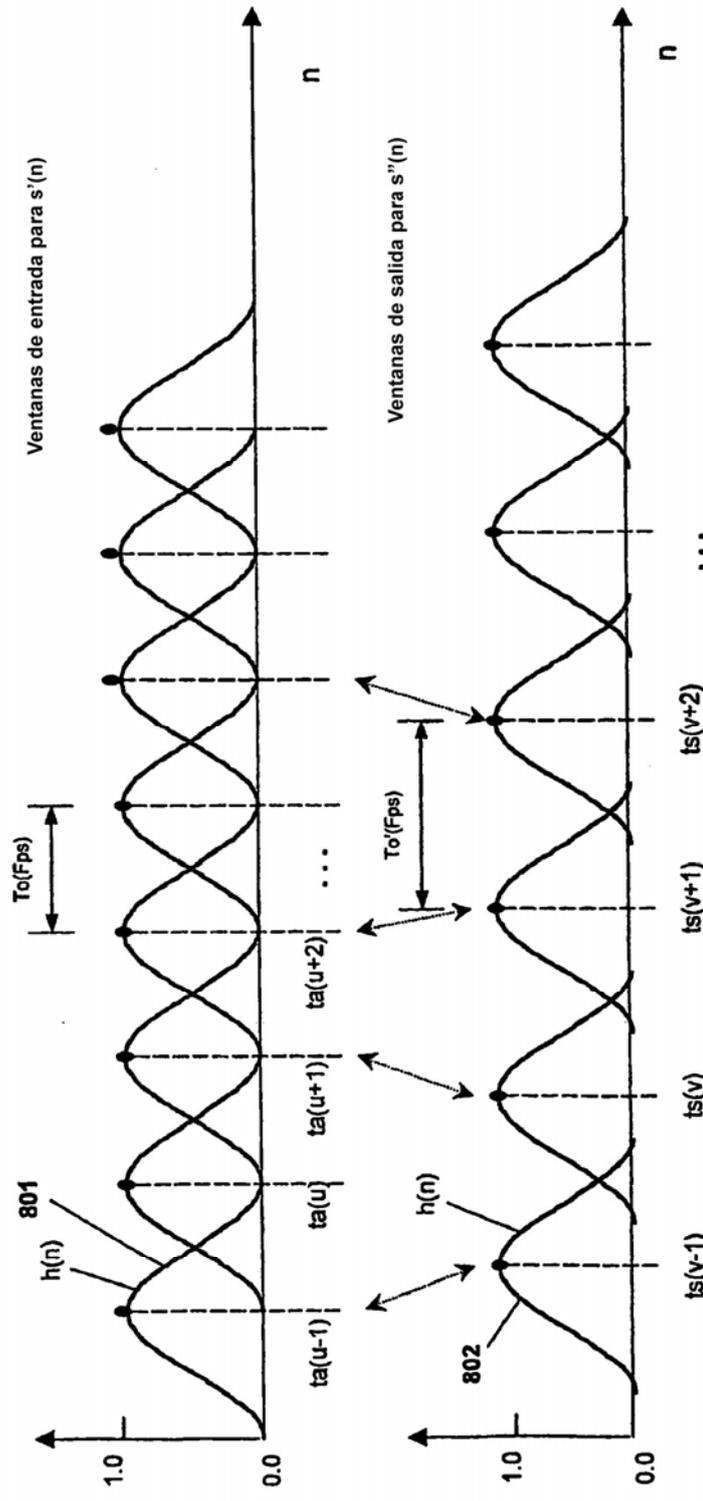


Figura 10

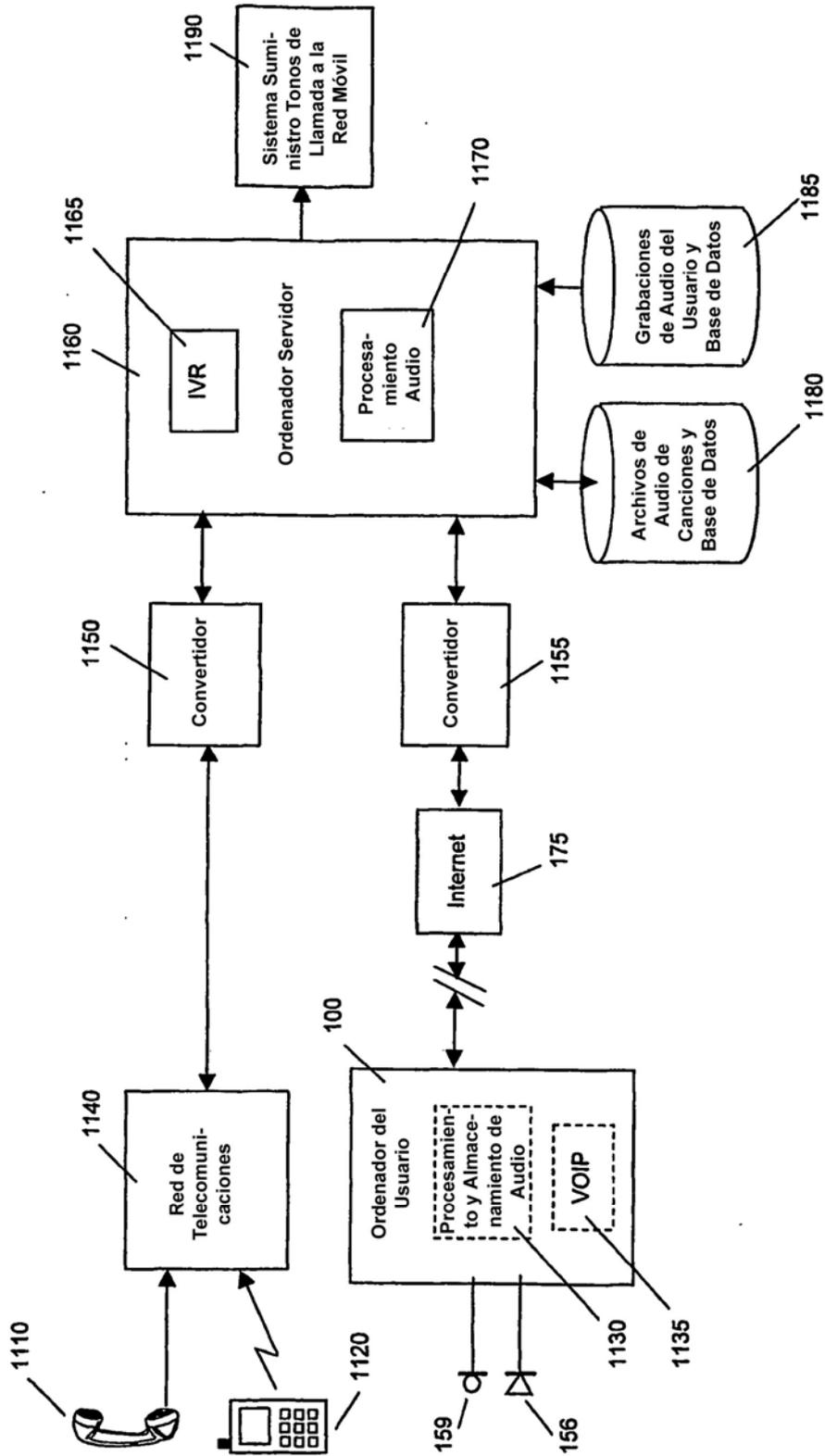


Figura 11

REFERENCIAS CITADAS EN LA DESCRIPCIÓN

5 Esta lista de referencias citadas por el solicitante está prevista únicamente para ayudar al lector y no forma parte del documento de patente europea. Aunque se ha puesto el máximo cuidado en su realización, no se pueden excluir errores u omisiones y la OEP declina cualquier responsabilidad al respecto.

Documentos de patente citados en la descripción

- US 5966687 A [0008]
- JP 2003044066 B [0008]
- US 5750912 A [0008]
- US 6836761 B [0010]
- GB 2117168 A [0015]
- US 4591928 A, Bloom [0015] [0034] [0059] [0063] [0064] [0065]
- WO 2004040576 A [0019]

10

Documentos no procedentes de patentes citados en la descripción

- D.T. Chappell ; J.H.L. Hansen. Speaker – specific pitch contour modelling and modification. Acoustics, Speech and Signal Processing, 1998, Proceedings of the 1998 IEEE International Conference in Seattle, WA, USA, May 1998 [0011]
- J Holmes ; W Holmes. Speech synthesis and recognition. Taylor and Francis, 2001 [0015]
- K. Lent. An efficient method for pitch shifting digitally sampled sounds. Computer Music Journal, 1989, vol. 13 (4), 65-71 [0016]
- L.R. Rabiner ; R.W. Schafer. Digital Processing of Speech Signals. Prentice Hall, 1978 [0034] [0046]
- Wolfgang Hess. Pitch Determination of Speech Signals. Algorithms and Devices. Springer-Verlag, 1983 [0066]
- R.J. McAulay ; T.F. Quatieri. Pitch estimation and voicing detection based on a sinusoidal model. Proc. Int Conf. on Acoustics, Speech and Signal Processing, Albuquerque, NM, 1990, 249-252 [0066]
- T.F. Quatieri. Discrete-Time Speech Signal Processing: Principles and Practice. Prentice Hall, 2002 [0066]
- K. Lent. An efficient method for pitch shifting digitally sampled sounds. Computer Music Journal, 1989, vol. 13 (4), 65-71 [0088]
- J. McAulay ; T. Quatieri. Shape Invariant Time-Scale and Pitch Modification of Speech. IEEE Trans. Sig. Processing, IEEE Trans. Sig.
- N. Schnell ; G. Peeters ; S. Lemouton ; P. Manoury ; X. Rodet. Synthesizing a choir in realtime using Pitch Synchronous Overlap Add (PSOLA). International Computer Music Conference, 2000, 102-108 [0088]
- J. Laroche; M. Dolson. New Phase-Vocoder Techniques for Pitch-Shifting, Harmonizing and other Exotic Effects. Proc. 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 1999, 91-94 [0088]
- G. Peeters. Analyse-Synthese des sons musicaux par la methode PSOLA. Proceedings of the Journees d'Informatique Musicale, Agelonde, France, 1998 [0088]
- V. Goncharoff ; P. Gries. An algorithm for accurately marking pitch pulses in speech signals. Proceedings of the IASTED International Conference Signal and Image Processing (SIP'98), 28 October 1998 [0088]
- D. Malah. Time Domain Algorithms for Harmonic Bandwidth Reduction and Time Scaling of Speech Signals. IEEE Transactions Acoustics, Speech and Signal Processing, 1979, vol. 27 (2), 121-133 [0089]

Processing,

March 1992, vol. 40 (3), 497-510 [0107]

- D. O'Brien; A. Monaghan. Shape Invariant Pitch Modification of Speech Using a Harmonic Model. EuroSpeech,

1999, 1059-1062 [0107]