

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 571 044**

51 Int. Cl.:

**H04S 7/00**

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **30.12.2010 E 10197378 (2)**

97 Fecha y número de publicación de la concesión europea: **09.03.2016 EP 2357854**

54 Título: **Procedimiento y dispositivo para la generación de señales de audio binaurales ajustables individualmente**

30 Prioridad:

**07.01.2010 DE 102010004171**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

**23.05.2016**

73 Titular/es:

**DEUTSCHE TELEKOM AG (100.0%)  
Friedrich-Ebert-Allee 140  
53113 Bonn, DE**

72 Inventor/es:

**LINDAU, ALEXANDER y  
ESTRELLA CAZURIAGA, ROLANDO JORGOS**

74 Agente/Representante:

**MORGADES MANONELLES, Juan Antonio**

**ES 2 571 044 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

Procedimiento y dispositivo para la generación de señales de audio binaurales ajustables individualmente.

5 La invención se refiere a un procedimiento y dispositivo para generar señales de audio binaurales ajustables individualmente.

10 Durante un largo tiempo, la realización de entornos acústicos virtuales de forma plausible o incluso auténtica ha sido una meta de desarrollo técnico en el sector de los sistemas y procedimientos de audio binaurales. Estos procedimientos están destinados a generar una impresión auditiva para el oyente (usuario, persona de prueba, paciente, etc.) lo más similar posible a la realidad. Los parámetros esenciales para la audición espacial, también llamados "indicaciones" ("cues"), representan diferencias de tiempo de tránsito y de nivel que dependen de la dirección y frecuencia mostradas por las señales acústicas de una fuente cuando chocan en los oídos o más bien en los tímpanos del oyente. En lo que sigue se podrá hacer referencia a la llamada teoría dúplex de audición que se originó en Lord Rayleigh.

15 Un sistema para simular estos entornos acústicos virtuales (espaciales) (también llamados entornos VAE o entornos acústicos casi virtuales) tiene que ser capaz de generar correctamente estos parámetros. En general, se designa como diferencia de tiempo de tránsito (designación completa de medios ITD o retardo de tiempo más bien interaural) o diferencia de nivel interaural (también llamadas ILD o diferencia de nivel más bien interaural). Esta última resulta principalmente del hecho de que la cabeza es un obstáculo para la propagación del sonido, mientras que se pueden atribuir diferencias más sutiles a la forma más específica del oído externo. En contraste, la diferencia de tiempo de tránsito interaural ITD resulta principalmente de la distancia geométrica de ambos oídos, siendo dependiente de la posición a diferentes tiempos de llegada al tímpano auditivo del sonido que se origina de la fuente. Para una determinada localización, el ITD puede ser evaluado solamente de forma significativa dentro de un rango de frecuencia de aproximadamente 1500 Hz, mientras que por encima de este rango las ambigüedades impiden esta asociación y las diferencias de nivel ILD determinan de manera creciente la dirección percibida del sonido entrante. Ambos aspectos pueden ser intercambiados uno con respecto a otro ("tiempo de tránsito - intensidad - compromiso"), del que se aprovecha, por ejemplo, la estereofonía que es realizable como estereofonía de tiempo de tránsito, de nivel o mixta.

20 Una descripción completamente dependiente de la frecuencia de la información temporal y espectral del sonido es proporcionada por la función de transferencia relacionada con la cabeza. Esta función es definida como la proporción de las siguientes funciones de transferencia: por una parte la función de transferencia en el tímpano del oyente (opcionalmente también en el canal auditivo de una cabeza artificial) y, por otra parte, la función de transferencia de un micrófono omnidireccional de una cabeza imaginaria, es decir, no presente. Esta llamada función de transferencia relativa a la cabeza (designación completa de HRIR o más bien HRTF, HRIR, significa la respuesta de impulso relativa a la cabeza y HRTF la función de transferencia relativa a la cabeza) se encuentra libre de reverberación, por definición. Por ejemplo, en el documento GB-A-2369976 se describe un procedimiento de síntesis de HRTF de campo difuso. Un procedimiento y dispositivo para procesar diferencias de tiempo interaural ("Procedimiento y aparato para el proceso de retraso de tiempo interaural en audio digital 3D") se describe en el documento US 7.174.299 B1. En el documento US 6.795.556 se describe una función de transferencia relativa a la cabeza (HRTF). Se describe un procedimiento para la medición de las HRTF en el documento EP 0912077 o DE 69523643 T2.

25 No obstante, si se miden funciones de transferencia de una cabeza artificial procedentes de una fuente electroacústica dentro de un campo ambiente cargado de reverberación, se distingue entre las BRIR o las BRTF (respuestas a impulsos de recinto binaural o, más bien funciones y transferencias) como funciones de transferencias relativas a la cabeza que contienen también un recinto de reverberación. Con ayuda de esta función de transferencia relativa a la cabeza es posible generar una impresión acústica de las llamadas fuentes virtuales con elevada plausibilidad. Si una señal de audio libre de reverberación arbitraria se filtra a través de HRTF del oído izquierdo y del oído derecho se puede generar un evento acústico con impresión más o menos correcta de dirección y distancia cuando se reproduce, por ejemplo, mediante auriculares. En una forma de expresarlo, la teoría de la técnica binaural significa que de esta manera un proceso temporal perfectamente reconstruido de presión acústica en el tímpano conduce a una audición real del incidente auditivo idéntico.

30 En principio, se conocen procedimientos y sistemas basados en ordenador para la simulación numérica de acústicas de recinto que muestran bancos de datos con conjuntos de datos HRTF de oyentes medidos individualmente o cabezas artificiales con elevada resolución de ángulo espacial y proporciona una forma de calcular la respuesta a espacio - impulso para posiciones seleccionadas del oyente. No obstante, esto se hace idealmente solo para dichos oyentes individuales o para un oyente con una fisionomía o antropometría que digiere solo ligeramente de la cabeza artificial utilizada. A efectos de conseguir que posibles trazas acústicas procedentes de la fuente hacia el oyente se reproduzcan a través del modelo de rayos procedente de la óptica, en una forma de expresarlo. Toda traza acústica se correlaciona con una reflexión espacial; al ponderar las trazas acústicas con la dirección de llegada de dichas trazas acústicas, de acuerdo con la HRTF del oído izquierdo y del oído derecho y después de superponer todas las

reflexiones de la respuesta espacio - impulso evaluadas de esa manera se obtiene una respuesta espacio - impulso binaural del espacio virtual. Este espacio puede ser hecho audible al filtrar el material de audio libre de reverberación con esta BRIR; en este caso se habla también de “auralización”.

5 De manera alternativa, las BRIR pueden ser medidas también directamente in situ. Para ello las BRIR son medidas utilizando fuentes de excitación electrodo - acústica (altavoces) y un simulador o robot llamado de tipo cabeza - torso artificial (también llamado HEADS; simulador de cabeza y torso). Este robot (HEADS) permite una medición automatizada y espacialmente fina de las BRIR para los movimientos de la cabeza del oyente, debido a la emulación de los múltiples grados de libertad de movimiento (por ejemplo, para rotación tres grados transitorios), y puede  
10 conseguir una calidad de sonido significativamente mejor con respecto al original. En el artículo de Lindau, A.; Weinzierl, S. (2007): “Fabian: detección rápida de respuestas binaurales espacio - impulso en múltiples grados de libertad” escrito por A. Lindau y S. Weinzierl en “Avances de la acústica”, número 33. de la DAGA Stuttgart, Alemania, esta tecnología es presentada bajo la designación FABIAN. Se realiza FABIAN HEADS por un robot automatizable activado por software que consiste en una cabeza artificial con oídos externos y un cuerpo humano  
15 genérico. El robot sirve para conseguir una influencia natural del campo acústico (desviación, sombras, reflexión) como resultado en la posición de la medición real del campo acústico, siendo el micrófono membranas en el canal auditivo bloqueado. Utilizando este HEADS se puede conseguir una auralización mejorada.

Otra mejora debido al incremento de plausibilidad de simulaciones acústicas espaciales binaurales se puede tener, si se toma en consideración la interactividad del oyente con el significado de que se toma en consideración, la  
20 reacción de la simulación en los movimientos de la cabeza del oyente. Una característica deseada consistiría en que cualquier movimiento de la cabeza deseado o inconsciente fuera compensado a efectos de contribuir a una audición espacial plausible y libre de errores. Para conseguirlo, la función de transferencia relativa a la cabeza para cada posición de la cabeza del oyente (y eventualmente para los diferentes grados de libertad transitorios y de rotación)  
25 dado que se tienen que conocer las HRTF y BRIR. Se conoce ya la implementación de algoritmos de convolución de funcionamiento rápido invariantes en el tiempo, con capacidad de tiempo real, que funcionan en hardware de ordenador moderno. Con esta implementación se pueden simular movimientos relativos entre la fuente y el receptor en el caso de HRTF o un campo de ambiente en el caso de BRIR que no consisten en la rotación con los movimientos de la cabeza. Además, se pueden utilizar de manera efectiva o se pueden reducir artefactos típicos de mapeo o técnicas de cabeza artificial tales como “conmutación frontal - posterior” y “localización en la cabeza”.  
30

En el documento US 2006/045294 se describe un procedimiento para generar señales de audio ajustadas individualmente. Mediante un auricular, el oyente (sujeto que escucha) puede experimentar el sonido de altavoces  
35 virtuales que producen una impresión auditiva próxima a la realidad que difícilmente se pueda diferenciar de la impresión auditiva de altavoces reales. Para conseguir estas respuestas espacio - impulso de tipo múltiple personalizadas (PRIR) para las fuentes de audio de altavoz se detectan fuentes de audio a través de un número finito de posiciones de la cabeza del oyente. Después de ellos las PRIR son utilizadas para alterar o transformar respectivamente la señal de audio para los altavoces en la salida virtual para los auriculares. Dado que la transformación se basa en las posiciones de la cabeza del oyente (sujeto que efectúa la audición) el sistema no es  
40 capaz de ajustar la transformación, de manera que los altavoces virtuales no parezcan que se desplazan cuando el oyente mueve su cabeza.

El artículo “Encontrando el inicio de una respuesta de impulsos del recinto: es sencillo?” escrito por G. Defrance y otros, publicado en el “The Journal of the acoustical Society of America (JASA)”, número 124 (4), páginas 248 - 254,  
45 versión octubre 2008, publicada online el 24 de septiembre de 2008, se refiere a la determinación exacta del inicio (comienzo de las señales de audio/ datos) en relación con la dimensión temporal (temporización) de las respuestas de impulso espacial (respuesta de impulso de recinto: RIR).

En el documento US 7174229 B1 se describe un procedimiento y un dispositivo para procesar retrasos de tiempo interaurales en señales / datos de audio digitales 3D. Para reincluir las ITD extraídas o modeladas en juego de datos  
50 HRFT (sin recinto) se utiliza un retardo complicado en dos etapas que procede de tampones FIFO seleccionados de manera precisa y filtros de paso total fraccionales. Por lo tanto, este enfoque contiene un número desconocido de filtros generados explícitamente para cada unidad fraccional de tiempo que pueden ser realizados, por ejemplo, por paralelización y deben ser hechos seleccionables por empuje de un botón.

En el documento US 679- 5556 se describe un procedimiento destinado al escalado de indicaciones HRTF “no estándar” tanto en el dominio de tiempo como de frecuencia. Este procedimiento supera al anteriormente  
mencionado.

Aunque los procedimientos y dispositivos conocidos son capaces de simular acústica espacialmente de manera altamente factible mediante simulaciones binaurales dinámicas basadas en datos, continúan existiendo múltiples  
60 retos para conseguir la perfección perceptiva. En particular, se deben citar los sectores siguientes:

La percepción de latencia en el proceso del movimiento de la cabeza, la granularidad mínima de la trama de las  
65 BRIR con respecto al movimiento de la cabeza en todos los grados de libertad de rotación, la discriminación de

fuentes separadas espacialmente, los procedimientos de medición BRIR acelerados, la simulación de características de fuente natural (altavoz, instrumentos) y/o la rectificación de frecuencia óptima de la reproducción.

5 Entre los inventores se ha llevado a cabo una evaluación del entorno de simulación mediante una comparación directa A-B- (con el significado de una comparación que conmuta directamente de una fuente a otra para recibir una comparación no influenciada y objetiva) de un campo de audición simulado y real, además de una tasa de detección satisfactoriamente reducida de 52,6%, tiene lugar una serie de artefactos perceptivos restantes en los que tiene lugar el problema especialmente concreto de la descoloración del sonido y errores de localización.

10 De acuerdo con esto, el objetivo de la presente invención es dar a conocer un procedimiento que supera los problemas anteriormente mencionados. En particular, el aspecto de artefactos de percepción, tales como errores de localización, errores de latencia - percepción o errores de atenuación ("fading") se deben eliminar o como mínimo disminuir de manera significativa. Además, se sugiere un dispositivo para llevar a cabo dicho procedimiento.

15 El objetivo es conseguido por un procedimiento que presenta las características de la reivindicación 1, así como un dispositivo con las características de las reivindicaciones dependientes.

20 De acuerdo con lo anterior se da a conocer un procedimiento que lleva a cabo las siguientes etapas para generar individualmente señales de audio binaurales individualizables:

funciones de transferencias relacionadas con la cabeza dependientes del tiempo de tránsito y diferencias de tiempo de tránsito correspondientes que son calculadas a partir de funciones y transferencias relacionadas con la cabeza diseñadas para diferentes posiciones de la cabeza mediante extracción de tiempos de tránsitos interaurales;

25 señales de audio independiente del tiempo de tránsito, refiriéndose a trazas de señales binaurales, son generadas mediante síntesis binaural, de manera que se utilizan funciones de transferencia relacionadas con la cabeza independientes del tiempo de tránsito, dependiendo de datos de la posición de la cabeza, para la posición real de la cabeza; y

30 se generan señales de audio binaurales ajustables individualmente a partir de señales de audio independientes del tiempo de tránsito al cargar la señal de audio dependiente del tiempo de tránsito generada para, como mínimo una traza de señal, con un retardo de tiempo que es dependiente de la diferencia de tiempo de tránsito calculada para la posición real de la cabeza y que depende de un factor individual de escalado. De esta manera, el retardo temporal entre el primer valor y el segundo valor, cuando tiene lugar un cambio de una posición previa de la cabeza a una posición real de la misma, se pueden cambiar adaptativamente mediante conversión de escaneado para interpretar la señal de audio dependiente del tiempo de tránsito que es acelerada o retrasada por el factor de reconversión. Se puede determinar el factor de conversión utilizado para la conversión de proporción de escaneado del retardo de tiempo que corresponde a un cambio de la posición anterior de la cabeza a la posición real de la misma.

40 Además, un dispositivo para realizar dicho procedimiento, en el que dicho dispositivo para generar señales de audio binaurales detectables espacialmente comprende los siguientes componentes:

45 un primer circuito que extrae tiempos de tránsito interaurales de funciones de transferencia relacionadas con la cabeza diseñadas para diferentes posiciones de la cabeza y calcula diferencias de tiempo de tránsito;

un segundo circuito que genera señales de audio independiente del tiempo de tránsito relativas a una correspondiente traza de señal binaural mediante síntesis binaural, de manera que este circuito utiliza funciones de transferencia relacionadas con la cabeza independiente del tiempo de tránsito dependiendo de datos de la posición de la cabeza para una posición real de la misma; y

50 como mínimo un tercer circuito que genera señales de audio ajustables individualmente a partir de señales de audio independientes del tiempo de tránsito al cargar la señal de audio independiente de tránsito como mínimo con una traza de señal con un retardo temporal dependiendo de la diferencia de tiempo de tránsito calculada para la posición real de la cabeza y un factor de escalado individual, siendo capaz al mismo tiempo de ejecutar cambios dependientes de la posición de la cabeza de la señal de manera no audible.

55 De acuerdo con ello se sugiere la extracción de tiempos de tránsito interaurales de funciones de transferencia dependiente de la cabeza, tales como BRIR para obtener funciones de transferencia independiente del tiempo de tránsito, y calcular diferencias del tiempo de cálculo de los tiempos de tránsito esperados para utilizarlos en una etapa posterior con un factor individual de escalado, a efectos de cargar la señales de audios generadas por la síntesis binaural dependiente de la posición de la cabeza a través de las funciones y transferencias independientes del tiempo de tránsito. Por lo tanto, mediante la aplicación de un procedimiento de etapas múltiples, la función de transferencia dependiente de la cabeza es relevada de los tiempos de tránsito interaurales que se refieren a una antropometría específica (por ejemplo, la de una cabeza artificial), y se crean señales de audio independientes del tiempo de tránsito mediante síntesis binaural y como consecuencia se utiliza un retardo temporal como factor de ponderación individual (factor de escalado) que carga los tiempos de tránsito para la persona específica o el usuario,

para conseguir una impresión auditiva espacial muy mejorada. En particular el retardo temporal corresponde a un valor que es calculado a partir de la espera de la diferencia de tránsito para una posición real de la cabeza y un factor individual de escalado.

5 El procedimiento está diseñado dinámicamente, de manera que un cambio que tenga lugar de una posición previa de la cabeza a la posición actual de la misma, cambia adaptativamente el retardo temporal entre un primer valor y un segundo valor a través de una tasa de conversión (SRC). En este caso, la conversión de proporción simple utiliza un factor de conversión para interpretar la señal de audio independiente del tiempo de tránsito que es acelerada o retardada por un factor de conversión, de manera que ese factor de conversión utilizado para la conversión de proporción de muestreo se obtiene debido a los cambios de retardo temporales correspondientes a la posición anterior y a la posición actual de la cabeza, respectivamente.

10 Al aplicar un procedimiento SRC que funciona en base a bloques de audio cortos temporalmente percibidos, es posible un cambio suave libre de latencia de las diferencias de tiempo de tránsito para los movimientos de la cabeza.

15 Con este procedimiento es posible calcular funciones de transferencias basadas en la cabeza, independientes del tiempo de tránsito, para múltiples posiciones de la cabeza y múltiples fuentes de señal de audio. Las posiciones múltiples de la cabeza pueden ser, por ejemplo, la resolución de una dirección de visión en pequeños cambios angulares o pasos angulares, por ejemplo, de 1 grado. Estas fuentes múltiples (fuentes de sonido y también fuentes de audio) pueden ser numerosas, con las que se crea una señal de audio para cada fuente y para cada traza de señal. Estas pueden ser superpuestas consiguientemente de acuerdo con una espera individual (retardo del tiempo de tránsito) para cada traza de señal.

20 La extracción de los tiempos de tránsito interaurales de las funciones de transferencia relacionadas con la cabeza que se ha ejecutado al inicio del procedimiento se puede ejecutar, por ejemplo, a través de uno de los siguientes procedimientos: procedimiento de inicio ("onset"), procedimiento de correlación interaural cruzada, creación de diferencia de tiempo de tránsito dependiente de la frecuencia, creación de diferencia de los gradientes de fase aproximados linealmente, dependientes de la frecuencia o determinación de la parte de exceso de fase a partir de la división de una función de transferencia antes y después de una transformación de Hilbert, respectivamente. Con el procedimiento de inicio se pueden obtener resultados particularmente satisfactorios, tal como se explicará en detalle a continuación.

25 En una forma ventajosa la invención soluciona el problema de las funciones HRTF y BRIR de transferencias relacionadas con la cabeza para procedimientos de auralización que usualmente son efectivos solo para la antropometría de un individuo específico o más específicamente una cabeza específica que provoca diferencias individuales, tales como la circunferencia de la cabeza que no esté representada correctamente por la información de latencia de la función de transferencia. Esto conduce a una concepción de localización más o menos alterada y de los movimientos de la cabeza y a la percepción de un movimiento espacial no natural de las señales de audio (estabilidad de la localización) para diferentes oyentes "ajenos".

30 Por medio de la invención se evitan errores de localización basados en errores de la circunferencia de la cabeza (desviación de la cabeza artificial). Como efecto secundario del procedimiento aplicado (liberación de tiempo de tránsito y atenuación de fase casi mínima) se obtienen significativas mejoras con respecto a los aspectos de latencia y errores audibles para la atenuación ("fading") durante los movimientos de la cabeza.

35 Las investigaciones llevadas a cabo por los inventores han demostrado una influencia con diferente intensidad de los respectivos artefactos en la autenticidad de la percepción de la audición. Por ejemplo, los evidentes errores de tono de color se hacen imperceptibles sin ninguna referencia dado que el oyente se acostumbra a ello; en general la memoria de tono de color global no es tan buena. La adaptación a discrepancia en la dimensión de la cabeza y la falsa información de latencia resultante de las señales acústicas parecen ser más severas. De esta manera, una circunferencia grande de la cabeza en el conjunto de datos conduce a un movimiento percibido de fuentes frontales que se desplazan en dirección opuesta a la dirección de movimiento de la cabeza; en el caso inverso las fuentes siguen los movimientos del oyente hasta una magnitud reducida. Este artefacto puede ser muy difícil de manejar y no se ha investigado todavía la capacidad de adaptación. La invención reduce también los errores de atenuación ("tartamudeo") que aparecen habitualmente que son causados por la atenuación respectivamente de los filtros de HRTF o BRIR con terminaciones de tiempo de tránsito en tiempo real. Esta "desalineación" temporal conduce a artefactos de atenuación similares a un filtro típico de "peine" que aparecen particularmente para porciones casi estacionarias (en aplicaciones de lenguaje, por ejemplo, con vocales, en música, en música, por ejemplo, para cuartetos de cuerda), de manera que produce distracciones. Esta invención reduce dichos errores de atenuación por la atenuación de la función de transferencia e incluyendo las diferencias de tiempo de tránsito dentro de la síntesis binaural de manera temporalmente consecutiva y no, tal como se realiza usualmente, al mismo tiempo.

40 A continuación se describirá la invención y las ventajas resultantes de la misma para realizaciones que se describen en los dibujos adjuntos, que muestran lo siguiente:

45 La figura 1 muestra el esquema de proceso para el procedimiento de acuerdo con la invención;

La figura 2 muestra la extracción de tiempos de tránsito interaurales y el cálculo de diferencias de tiempo de tránsito individuales;

5 La figura 3 muestra el diagrama de bloques de la parte de la invención que provoca el retraso de señal de audio escalable individualmente en base a las diferencias de tiempo de tránsito calculadas;

La figura 4a muestra el diagrama de bloques de un circuito para llevar a cabo la síntesis binaural dinámica;

10 La figura 4b muestra el diagrama de bloques según la figura 4a con descripciones explicativas de bloques y funciones específicas del circuito;

La figura 5a muestra el diagrama de bloques del circuito de acuerdo con la invención que es necesario para llevar a cabo la síntesis binaural dinámica individualizada y para aplicar la invención; y

15 La figura 5b muestra el diagrama de bloques de acuerdo con la figura 5a con descripciones explicativas de bloques de circuito y funciones específicas.

20 La figura 1 muestra el diagrama de proceso de un procedimiento 100, de acuerdo con la invención, que muestra múltiples etapas que contienen esencialmente lo siguiente:

25 En una primera etapa 110 se produce funciones de transferencia independiente del tiempo a partir de funciones de transferencias normales basadas en la cabeza, tales como BRIR, y se extraen tiempos de tránsito interaurales que se convierten en diferencias. Esto se explicará adicionalmente en la figura 2, que contiene un diagrama de bloques esquemático del primer circuito A.

30 En una segunda etapa se lleva a cabo una síntesis binaural mediante las funciones de transferencias independientes de tránsito a efectos de generar señales de audio L' y R' dependientes del tiempo de tránsito. Esto se explicará de manera más precisa en la figura 5 a/b. La síntesis binaural conocida es explicada también por las figuras 4 a/b.

35 En una tercera etapa 130 se lleva a cabo una síntesis en tiempo real a efectos de cargar individualmente las primeras señales de audio con un retardo de tiempo escalable. Esto se explicará adicionalmente en la figura 3 y también en la figura 5 a/b.

En primer lugar, se explica a continuación la primera etapa 110:

40 Tal como se ha mostrado en la figura 2, el circuito A comprende múltiples bloques 111 a 117, de funciones múltiples que representan respectivamente una de las etapas de la secuencia de etapas 110 (ver figura 1). La figura 2 desplaza, por lo tanto, el proceso previo de los conjuntos de datos impulso - respuesta, de manera que se puede conseguir una extracción casi inaudible del ITD de los conjuntos de datos empíricos HRTF - BRIR.

45 En primer lugar, se leen de un banco de datos en la primera etapa 111 funciones BRIR normales de transferencias (alternativamente también HRIR). A continuación, en el bloque 112, se lleva a cabo un sobremuestreo alrededor de un valor que permite una extracción más que audible de manera precisa de los tiempos de tránsito a partir de los impulsos - respuestas (por ejemplo 10 veces, de acuerdo con el muestreo de audio habitual de 44,1 o 48 kHz). Después de ello se pueden encontrar los inicios (comienzo de las señales de audio o de datos) en un bloque 113. A continuación, en el bloque 114, se determina la longitud del impuso - respuesta libre de inicio (casi fase mínima) y se crea como vector en el bloque 115. Debido a una sub-selección siguiente, por ejemplo, nuevamente 10 veces, se calculan en el bloque 116 funciones de transferencia independientes de tiempo de tránsito y también tiempos de tránsito extraídos como diferencias de tránsito ITD y se almacenan los resultados en el bloque 117. Además, se pueden añadir conjuntos de metadatos descritos DSI. El procedimiento, que se muestra en la figura 2, aplica el procedimiento de inicio a efectos de determinar impulsos - respuestas independientes del tiempo de tránsito. Esto se explicará más adelante.

55 Se observará en este punto que existen las siguientes conexiones entre los aspectos teóricos de filtro conocidos de funciones de transferencias, tales como HRTF o BRIR y la invención:

60 Los sistemas lineales invariantes en el tiempo y no recursivos incluyendo la función de transferencia espacial del oído externo y binaural se explican debido a su impulso - respuesta finito dentro del contexto de la teoría elemental de filtros. En el caso digital estos son filtros FIR (impulso - respuesta finito). Estos se realizan por acumulación de líneas de retardo ponderadas con los coeficientes del filtro; por lo tanto, representan implementaciones de la suma de convolución discreta. La resolución temporal se correlaciona con la tasa de selección de las señal de audio (por ejemplo, 44, 1 kHz). En el caso de un HRTF o BRIR empíricos se encuentran presente principalmente filtros de fase mixta  $X_{mixta}(n)$ . Estos se pueden separar en tres partes o porciones:

$$X_{\text{mixta}}(n) = X_{\text{lineal}}(n) + X_{\text{pasogeneral}} + X_{\text{mínimo}}(n) \text{ (Fórmula 1)}$$

Las diferencias de tiempos de tránsito interaural resultan como diferencias de los tiempos de tránsito de las HRTF y de las BRIR del oído izquierdo y del oído derecho. Estos tiempos de tránsito resultan como suma de la parte de fase lineal (parte de tiempo de tránsito puro) y la parte de paso general (distorsiones dependientes de la frecuencia sin distorsiones espectrales) con respecto a la llamada parte de fase en exceso.

$$X_{\text{exceso}}(n) = X_{\text{lineal}}(n) + X_{\text{pasogeneral}}(n) \text{ (Fórmula 2)}$$

La parte de fase mínima restante  $X_{\text{mínimo}}(n)$  describe los cambios espectrales (ILD) solamente y no contiene ya información de tiempo de tránsito absoluto o relativo.

La diferencia de tiempo de tránsito interaural ITD es dependiente de la frecuencia. La parte que es esencial para la localización correcta (<1500 Hz) es, no obstante, relativamente constante y se puede extraer más o menos sin artefactos. Como procedimiento la detección por inicio es particularmente útil, tal como se muestra en la figura 2. Como alternativa es útil la determinación de la parte de fase de exceso mediante una transformación de Hilbert, equilibrado de gradiente dependiente de la frecuencia, máximo de la correlación interaural cruzada o determinación específica del dominio de frecuencia de la diferencia de tiempo de tránsito de grupo interaural. Los procedimientos mencionados se pueden diferenciar por el hecho de que pueden ser extraídos también por la parte de paso general (procedimiento de Hilbert) o no (y todos los demás). Errores audibles como resultado de despreciar la parte de paso general aparecen, no obstante, solamente para algunas dimensiones entrantes del sonido. La proporción principal de las HRTF empíricas está formada a partir de una combinación de partes de fase mínima y de fase lineal. Errores demostrablemente audibles que tienen lugar durante la combinación de tiempos de tránsito de “no fase general” con espectros de fase mínima analíticamente perfectos tienen lugar solamente en el oído contralateral para sonidos que entran de forma extremadamente lateral. La extensión en la que este procedimiento puede ser aplicado a las BRIR que se ha medido en recintos (y no solamente en HRTF libres de reverberación) no ha sido investigada por los inventores. Como resultado, solamente el método de inicio era suficientemente exacto (comparar con el procedimiento de correlación cruzado), suficientemente estable (comparar con respecto al procedimiento de gradiente de fases), aplicable a la totalidad (comparar con respecto al procedimiento de diferencia de tiempo de tránsito de grupo) y de verdadero tono de color (comparar con respecto al procedimiento Hilbert) en el caso de las BRIR. Además, el procedimiento de inicio conserva todas las partes de paso general posibles de la BRIR de manera ventajosa; no se pierden sino que permanecen dentro de los espectros libres de tiempo de tránsito que se llaman, por lo tanto, en este caso de fase casi mínima. En pruebas auditivas formales y libres de criterios se puede verificar que la resintonización de los tiempos de tránsito extraídos y los espectros de fase casi mínima no provocan errores de localización incluso para sonidos que chocan de forma contralateral, tal como era de esperar. Otras pruebas auditivas formales han mostrado que los cambios de estructuras de reverberación provocados por la transformación de Hilbert son audibles en cualquier caso.

Además, se conocen adiciones basadas en modelos geométricos que calculan ITD dependiente de ángulo a partir de datos antropométricos de la cabeza. No obstante, estos procedimientos muestran una debilidad subjetiva y tienen el déficit esencial de que para el ITD basado en un modelo que tiene que ser generado, la dirección de incidencia del sonido, que es la posición de la fuente virtual, sobre la que se basan los juego de datos HRTF o BRIR empíricos, debe ser conocida. Para las HRTF este es un problema trivial dado que una única HRTF describe la incidencia de sonido para una dirección específica o respectivamente para un ángulo específico. Para la medición de campo de sonido en recintos con cabezas artificiales móviles son posibles posiciones fuente arbitrarias (por ejemplo, situaciones de concierto simulados). Estas no pueden ser extraídas de los juego de datos de una forma basada necesariamente en funcionalidad. El modelo conocido basado en recuperación ITD es, por lo tanto, más útil para adiciones de modelo numérico, paramétrico de acústica virtual y no para BRIR medidas in situ.

De acuerdo con la invención, las funciones de transferencias basadas en la cabeza se encuentran libres de sus tiempos de tránsito. Después de la separación de espectro de fase casi mínima y tiempos de tránsito los dos pueden ser introducidos separadamente en una resíntesis, en la que se puede llevar a cabo una resíntesis escalable y ajustada al individuo (ver figura 3 y figuras 5 a/b). En este caso las ventajas consisten en una reducción de la latencia y reducción de las HRTF que se han acercado (si aquellas se utilizan como filtro FIR tal como se supone implícitamente) por el tiempo de tránsito extraído que se han mencionado. En principio las HRTF pueden ser generadas también como filtro IIR (respuesta de impulso infinito) por modelación / estimación de las HRTF medidas que se encuentran necesariamente presentes igual que si el filtro FIR (respuesta de impulso finito) o por modelado parametrizado de características esenciales.

El procedimiento descrito puede ser aplicado principalmente a funciones de transferencias basadas en la cabeza más que para las generadas de forma arbitraria. No obstante, se describe por ejemplo, para juego de datos empíricos HRTF/ BRIR que tiene lugar, por ejemplo, igual que si se hubieran medido filtros FIR con los robots anteriormente mencionados (HEADS FABIAN, ver Lindau y otros. 2007).

En el contexto de la extracción de funciones y transferencias ya existentes BRIR / HRIR, tal como se da a conocer en la figura 2 se debe indicar que estas pueden ser conseguidas con el mencionado robot FABIAN:

El robot FABIAN fue utilizado por los inventores teniendo 7 grados de libertad (4 de rotación, 3 de translación). El robot es capaz de ser automatizado mediante software de control consistiendo en una cabeza artificial con oídos externos y un cuerpo humano reconstruido de manera genérica. El robot sirve para conseguir una fluencia natural del campo sonoro (desviación, sombreado, reflexión) como resultado en la posición de la medición del campo sonoro de las dos membranas del micrófono en el propio canal auditivo bloqueado.

Los algoritmos de la invención descritos a continuación se refieren, por ejemplo, a la utilización de juego de datos BRIR que han sido obtenidos con dicho robot. No obstante, el procedimiento no está restringido a aquel juego de datos sino que es aplicable para cualquier procedimiento de auralización que realiza localización sonora por filtrado de impulsos - respuestas relacionadas con la cabeza con señales de audio. El procedimiento consigue una mejora significativa de la síntesis binaural dinámica que se basa en los juego de datos HRTF/ BRIR y muestra las siguientes ventajas de manera específica:

- i) una extracción exacta casi inaudible y
- ii) una reconstrucción y manipulación en tiempo real individual también inaudible (escalado dependiente de la frecuencia) de la diferencia de tiempo de tránsito interaural.

El procedimiento 100 (ver figura 1) incluye una etapa de proceso 110 (ver figura 1 así como los bloques parciales de la figura 2) y una resíntesis en la etapa 130 en tiempo real (ver figura 1 así como la figura 3) que se realiza después de una síntesis binaural dinámica de algoritmo / etapa 120 (ver figura 1 y también las figuras 4a/b) (ver figuras 5a/b).

Con respecto a la funcionalidad y construcción del dispositivo, de acuerdo con la invención (ver figura 5a/b) y también en cuanto al procedimiento ejecutado (ver figura 1) se puede indicar lo siguiente:

En primer lugar, se lleva a cabo una etapa 110 un preproceso de los juego de datos impulso - respuesta. El método propuesto empieza con una extracción del tiempo de tránsito interaural de los juego de datos genéricos HRTF / BRIR. La producción de los juego de datos BRIR (bloque 111 en la figura 2) se lleva a cabo preferentemente por el robot anteriormente descrito en recintos reales para números arbitrarios de fuentes electroacústicas y mientras se utiliza una trama finamente resuelta del movimiento de la cabeza dentro de rangos fisiológicamente útiles.

Si existen juego de datos empíricos, entonces los tiempos de tránsito son extraídos inicialmente (especialmente bloque 113) y se calcula la diferencia ITD de tiempo de tránsito interaural. En este caso la ITD se define del modo siguiente:

$$ITD = | TL - IR | \text{ (Fórmula 3)}$$

en la que  $TL = t_{pasada}$  (oído izquierdo) y  $TR = t_{pasada}$  (oído derecho)

Tal como se ha mencionado anteriormente, el método de inicio es el más útil para la extracción entre los métodos mencionados. Los otros métodos muestran una estabilidad o robustez menor para el conjunto de datos empíricos de impulsos - respuestas espaciales binaurales. El método de Hilbert parece ser el no utilizable dado que cambia la estructura de reflexión en el medio ambiente a una medida eventualmente audible debido a un compactado de energía en la dirección del inicio del impulso - respuesta. Los inventores han sido capaces de hacer plausible este método con la ayuda de pruebas previas parecidas a modelo auditor y una prueba auditiva anterior. El método de inicio es ajustado, de manera tal que el ruido blanco (típicamente alrededor de -52 a -90 dB con respecto al valor máximo del impulso - respuesta) se determina y se escoge un umbral que se encuentra significativamente por encima (por ejemplo, 15 dB más elevado, significando -35 dB<sub>rel</sub>). El juego de datos impulso - respuesta es escaneado a continuación automáticamente empezando respectivamente desde el inicio de los impulsos - respuestas y se calcula de acuerdo con los casos de tiempo de superación del umbral en el canal de la izquierda y de la derecha del HRTF / BRIR calculado formando la diferencia del ITD (ver figura 3). Los tiempos de tránsito son eliminados entonces y se reservan nuevamente los impulsos - respuestas independientes del tiempo de tránsito obtenidas (bloque 117). El método de inicio es aplicado en una señal de tiempo amortiguado 10 veces para obtener una discretización del ITD en etapas de 2,3 μs. Esta resolución corresponde aproximadamente a una quinta parte del umbral de ITD perceptible. El ITD de diferencia de tiempo de tránsito es extraído de este modo pudiendo ser añadido adicionalmente al juego de datos independientes del tiempo de tránsito en formato de lista legible a máquina (por ejemplo, \*.txt, \*.csv, \*.xml).

El procedimiento de pre-avance, tal como se describe en la figura 2 incluye las siguientes etapas:

111: leer un impulso - respuesta (archivo \*.wav)

112: amortiguar impulso - respuesta 10x

113: hallar inicio de acuerdo con criterio de nivel; índices de inicio de la matriz de la derecha para canales izquierdo y derecho; calcular ITD a partir de la diferencia de índices de inicio; calcular el resultado de cálculo por el factor de amortiguación; encontrar tiempo de tránsito máximo en juego de datos

5 114: determinar nueva longitud de impulso - respuesta,  $new\_len=size\_IR \max\_time\_of\_flight$  (resultados en longitud del IR más corto posible en el rango de amortiguación, servicio según longitud para todas las IR extraídas a continuación)

10 115: extraer del vector impulso - respuesta de longitud  $new\_len$  empezando con los índices de inicio del 116: amortiguación de extracción del impulso - respuesta  $10x$

117: reservar IR como nuevo\*.wav archivo

15 Adicionalmente: reservar las ITD en archivo legible a máquina; escribir también un juego de datos de descripción de \*.xml archivo DSI (con información sobre los puntos inicial y final, resoluciones para azimut y elevación).

Este proceso es llevado a cabo en un primer circuito A que es parte del dispositivo (ver figuras 5a/b).

20 Otro circuito B con bloques de función BA y BB (ver figura 4a/b) es aplicado en el dispositivo ver figuras 5a/b) en una forma modificada y facilita una resíntesis del tiempo real adaptativa individual que se describe con mayor detalle a continuación:

25 Es punto inicial o de modo más preciso condición previa para la reproducción de audio binaural individualizada, la presencia de un algoritmo de convolución en tiempo real conocido ya y descrito en la figura 4a/b. Este algoritmo debe ser capaz de llevar a cabo la convolución de basada en un bloque individual para cada fuente de audio individual que se tiene que auralizar. En este caso, el procedimiento técnico de la convolución rápida en tiempo real no difiere del tiempo de tránsito involucrado y BRIR libres de tiempo de tránsito. No obstante, debido a la reconstrucción de datos aparecen varias ventajas de sonido que se explicarán más adelante.

30 Para modificar el proceso de convolución o más bien para individualizar el tiempo de tránsito se utiliza una arquitectura llamada "servidor de audio de jack" ("jack audio server") y, por lo tanto, se realiza un plugIn autónomo. Se pueden realizar también, por ejemplo, soluciones equivalentes continuamente a partir de disposiciones de hardware de DSP especial mediante procedimientos que corresponden a la arquitectura VST - PlugIn - (es decir basados en el interfaz VST). Después de poner en marcha el dispositivo controlado por ordenador (por ejemplo, un PC) esto consiste en la lectura de un archivo de configuración, un archivo de descripción de juego de datos y una lista basada en texto que contiene las ITD del juego de datos BRIR que está siendo auralizado en el momento por el proceso clásico de convolución. Entonces se inicia en el servidor de audio y crea dos entradas y salidas (para la señal izquierda y derecha binaural del proceso de convolución) para cada caso (es decir, para cada fuente de audio auralizante ajustable individualmente). Las entradas pueden ser entonces conectadas a las salidas del proceso de convolución en forma PlugIn (consultar figuras 5a/b). Si se tienen que auralizar múltiples fuentes entonces se tiene que poner en marcha para cada fuente un correspondiente proceso de convolución y un proceso para individualizar los tiempos de tránsito. El proceso que individualiza los tiempos de tránsito copia el resultado de la salida del proceso de convolución (señales de tiempo binaural sin diferencias de tiempo de tránsito ver L' y R' en las figuras 5a/b) en forma de bloque en sus propios tampones de entrada. A efectos de que la latencia base no se encuentre por encima de la carga dentro de este bloque encascado serie se han demostrado que las muestras de tamaño 128-512 son válidas para ambos procesos.

50 Las diferencias de tiempo de tránsito faltantes pueden ser introducidas ahora en uno de los dos canales de audio como retardo de tiempo VDL específico de la posición de la cabeza sin artefactos audibles. Esto se realiza, por ejemplo, para L' en el bloque VDL / SCR, de manera que el bloque VDL / SCR representa el ajuste dinámico del tiempo de tránsito durante el movimiento de la cabeza y el bloque S representa un ajuste estático que es obtenido / ajustado para una posición rígida de la cabeza. Para el ajuste, el proceso de individualización lee los datos de movimiento de la cabeza como flujo de datos del seguidor de cabeza HTDAT (por ejemplo, a través de un protocolo de transporte basado en IP), que se envían con los mismo al proceso de convolución. Durante la inicialización del proceso de individualización el HDAT tiene que realizar la primera diferencia de tiempo de tránsito interaural efectiva determinada por la posición inicial de la cabeza mediante una reproducción temporalmente retrasada o adelantada en una magnitud fija. A efectos de simplicidad la primera diferencia temporal se puede suponer ajustada a 0 o respectivamente se puede suponer el primer factor de conversión ajustado a 1. Como consecuencia, los cambios de posición de la cabeza en la diferencia de tiempo de tránsito interaural se pueden conseguir de la manera siguiente: para cada bloque de audio procesado se comprueba la posición real de la cabeza y, por lo tanto, el cambio en las diferencias de tiempo de tránsito. Si se han detectado cambios de las diferencias de tiempo de tránsito su magnitud se determina de acuerdo con la formación de diferencias ( $\Delta ITD = \text{nuevo ITD} - \text{anterior ITD}$ ). Al crear la proporción fraccional de longitud de bloque de audio que incluye el impacto de diferencias de tiempo real ( $\Delta ITD$ ) utilizadas como factor de conversión en un algoritmo para convertir las proporciones de selección en tiempo real en la máxima cavidad de audio. Este algoritmo consigue el ajuste de los valores de retraso cambiados a través de una reproducción acelerada o retrasada en el factor de conversión por la interpolación de ordenadores de

selección en la señal actual que corresponde a una proporción de selección más alta o más baja y emitiendo estos valores como sustitución.

Con ayuda de la conversión de proporción de selección (SRC) se pueden hallar los valores de amplitud correctos a partir de puntos de selección de señales de tiempo digitales en momentos de tiempo arbitrario con el significado de tiempos continuos. La llamada conversión de proporción de selección fraccional es conocida también (ver Smith III, J. O. (2002): página reselección audio digital. página publicada en <http://www.ccrma.stanford.edu/~jos/resample/>, Stanford: Center for computer research in music and acoustics (CCRMA), Stanford University). De acuerdo con el teorema de Shannons de la reconstrucción perfecta del proceso de tiempo de señales seleccionadas por filtros de paso bajo perfectos filtro "anti - aliasing" (anti - desviación) el procedimiento de interpolación utiliza una tabla de coeficientes filtrados casi continuos interpolados en este punto para proporcionar efectivamente el impulso - respuesta de paso bajo específico en forma de una señal de tiempo continuo. Un tabla de filtro precalculada única es suficiente para todos los tiempos de interpolación y proporciones SRC. De acuerdo con la proporción de conversión de la tasa de selección  $f_{s_{nuevo}} / f_{s_{anterior}}$  (es decir, el factor de conversión es  $<1$  o  $>1$ ) se puede obtener de la tabla un filtro con frecuencia de corte más alta o más baja. Esto es posible dado que el impulso - respuesta del filtro es tabulado de manera casi continua con el significado de que tiene lugar con precisión de selección muy elevada y con una gran duración. Para cada bloque de selecciones de salida a tabular, para bloques con duración muy corta la conversión variable es, por lo tanto, casi posible en tiempo real, se calcula el resultado de interpolación en los puntos de selección formando la convolución suma de las selecciones de señal de tiempo con el nuevo impulso - respuesta de paso bajo. Dependiendo de la calidad necesaria (es decir, anchura de banda y atenuación del interpolador) se puede estimar la dimensión de la tabla de paso bajo de la interpolación por fórmulas adecuadas. En este caso se utiliza un interpolador flotante de banda limitada, de alta calidad basado en una función de sincronización analítica con una proporción señal a ruido que en el peor caso es de 97 dB con una amplitud de banda de 97%.

Tal como se ha mencionado anteriormente el factor de conversión perteneciente al bloque de audio procesado en el momento se puede determinar a partir de la diferencia de la ITD correspondiente a la posición de la cabeza anterior y la actual. Una biblioteca de software disponible en todo momento permite un cambio inaudible y continuo del factor de conversión, a efectos de resintetizar correctamente la diferencia de tiempo de tránsito de la correspondiente posición de la cabeza. Para la realización práctica de la diferencia de tiempo de tránsito interaural, esto significa que se crea un número de nuevas selecciones de salida que es siempre el mismo y que corresponde a la longitud del bloque de audio subyacente, a partir de un número variable de selecciones de entrada a través de la biblioteca anteriormente mencionada que permite interpolación limitada de banda. Dependiendo de la proporción de conversión - relación de selección ( $f_{s_{nuevo}} / f_{s_{anterior}}$ ) ( $<1$  o  $>1$ ) la conversión de tasa de selección necesita más o menos selecciones que el tamaño del bloque actual subyacente. A causa de esta diferencia se obtiene como nuevo resultado excepto para el bloque de las nuevas selecciones de audio interpoladas, un desplazamiento temporal fijo de la tasa de lectura de la conversión de la proporción de selección que es el número de selecciones que ya no se utilizan o que se utilizan adicionalmente. En las figuras 5a/b este desplazamiento es mostrado por el bloque S como retardo fijo y retardo preciso de selección con un valor corriente que se puede obtener a partir de todo los cambios previos de diferencia de tiempo de tránsito.

Si no se ha detectado cambio alguno de diferencia de tiempo de tránsito interaural dentro de dos bloques de audio este valor de retraso fijo (retraso estático S) sigue siendo el mismo y, por lo tanto, realiza adicionalmente la diferencia de tiempo de tránsito interaural concreta.

En anticipación del escalado individual del factor de conversión que se explicará a continuación se debe observar que este retraso fijo es escalado individualmente también de manera correcta dado que resulta solamente de la acumulación de valores de salida de una conversión de tasa de selección ya existente.

Una ventaja significativa de la invención es que la diferencia de tiempo de tránsito se puede ajustar individualmente escalando el factor de conversión corriente respectivo. Para proceder de este modo la ITD, por ejemplo, antes de calcular el factor de conversión, es simplemente escalada con un valor fijo (valores útiles entre 50 - 150%). De este modo se obtiene una ITD \* individualmente ajustada entre las señales de salida  $L^*$  y  $R^*$ . Con este escalado dependiente de la frecuencia de la ITD se puede conseguir un ajuste que ya es significativamente incrementado en cuanto a percepción con respecto al individuo específico. Además, este escalado de la ITD que corresponde aproximadamente a un escalado subsiguiente de la circunferencia de la cabeza es controlable en tiempo real. A continuación se explica la forma óptima en la que se puede encontrar un factor de escalado individual opcional con la ayuda de dos ejemplos:

El añadido para el primer ejemplo se basa en la construcción del factor por una persona de prueba: dado que el valor de corrección es construible en tiempo real, puede ser construido al inicio de la aplicación acústica virtual con ayuda de una escena acústica adecuada (por ejemplo, reproducción de una fuente acústica frontal única que emite impulsos sonoros cortos pero bien localizables). Para conseguirlo la persona de prueba modifica el valor de corrección empezando por un valor inicial relativamente alto, (por ejemplo, 150% de la ITD) en dirección negativa hasta que se consigue la percepción de localización correcta que se distingue por una localización de la fuente de percepción estable que, es decir, particularmente sin los movimientos de interacción que siguen a los movimientos

de la cabeza. El valor inicial escogido como relativamente alto al inicio hace el artefacto, que tiene que ser minimizado, extremadamente bien audible.

5 Otro añadido que sigue como segundo ejemplo se basa en una predicción del valor de corrección individual a base de una medición antropométrica. Se han llevado a cabo pre exámenes con múltiples personas de prueba. En este caso ha sido llevada a cabo la prueba de audición de acuerdo con el escenario acústico mencionado anteriormente; además de la fuente virtual las personas de prueba fueron también capaces de escuchar la fuente acústica real. La meta consistía en ajustar el factor de conversión de la simulación para obtener una buena correspondencia cuando se conmuta entre simulación y realidad. Esta prueba ha sido repetida varias veces por cada persona de prueba. 10 Además, se llevaron a cabo cuatro mediciones adecuadas de la cabeza de cada persona respectiva. A continuación, la predicción del valor medio de los valores de escalado individuales a partir de las mediciones de la cabeza fue comprobada por múltiples regresiones lineales. Resulto que solamente la distancia entre tragos, medida con un medidor de longitudes, en milímetros, desde la incisure anterior del oído izquierdo a la del oído derecho, es suficiente para la predicción del factor de conversión. La representación gráfica de las pruebas mostró una línea de regresión y un intervalo de confianza de 95% del modelo lineal, así como los datos de valor medio individuales de las muestras. El KI de 95% mostro que el factor de conversión puede ser objeto de predicción con una precisión de  $\pm 1,25\%$  (rango medio del modelo) y de  $\pm 2,5\%$  (rango externo del modelo). 15

La variación del factor escalado con la sonda puede ser  $\pm 4\%$  lo que significa que en el caso no individualizado podría haber sido posible un error en el caso peor del ITD de hasta 8%. La predicción individual reduce este error por un factor de dos o de un factor cuatro (en el mejor de los casos). Un error ITD de 1,25% corresponde a un error de localización ligeramente más de un 1º y, por lo tanto, se encuentra (nuevamente, en el mejor de los casos) en el rango del cambio de localización que es justamente perceptible. La fórmula de regresión utilizada para la predicción del factor de corrección individual es la siguiente: 20

$$25 \text{ Escala} = 0,00304 * \text{distancia entre tragos [mm]} + 0,5792$$

Esta ecuación de regresión proporciona una predicción del factor de corrección individual. No obstante, esta forma es solamente aplicable a juego de datos que han sido medidos con el robot FABIAN HEADS. Una posibilidad de generalización a juego de datos arbitrarios HRTF / BRIR consiste en añadir una corrección de los resultados de la fórmula previamente mencionada por multiplicación con la proporción de medición de la circunferencia de la cabeza distancia entre tragos<sub>ajeno</sub>/ distancia entre tragos<sub>FABIAN</sub>. Esta corrección implica evidentemente que la distancia entre tragos de los datos ajenos es conocida. 30

35 El principio del procedimiento de resintetización, de acuerdo con la invención se muestra en la figura 3:

a) Inicialización

- 40 i) lectura \* .xml archivo de configuración
- ii) leer definición del conjunto de datos del archivo de descripción \*.xml
- iii) disponer un conjunto de 1 o 2 dimensiones por lectura de los ITD de 1 archivo legible a máquina
- 45 iv) factor de escalado inicial fijo de ITD=1
- v) inicializar tampón de anillo del proceso SCR
- 50 vi) iniciar servidor OSC para recibir datos del movimiento de la cabeza
- vii) inicializar "jack" cliente
- viii) inicializar modos de control en tiempo real (OSC, GUI, teclado)

55 b) Proceso en tiempo real

- i) copiar muestras de entrada para la izquierda y para la derecha en el tampón anular
- 60 a. leer ITD con respecto a la posición de la cabeza de acuerdo con el seguimiento de la cabeza
- b. escalar ITD en tiempo real si se desea  
( $ITD_{nuevo} = ITD * \text{factor de escalado}$ )
- 65 c. calcular ITD fraccional en muestras alrededor de  
(nuevo retardo =  $ITD_{nuevo} / \text{tamaño muestra}$ )

d. determinar cambio de ITD desde el último calculo  
(retardo delta = nuevo retardo - retardo viejo)

e. determinar factor SRC fraccional  
(relación=muestras salida/ (muestras entrada-retardo delta))

ii) facilitar al proceso SCR: factor, tampón de entrada y de salida para el canal de la izquierda.

iii) reservar las muestras no utilizadas en el tampón anular.

iv) calcular la longitud del tampón de entrada del nuevo ciclo (longitud tampón entrada = muestras salida+  
muestras no utilizadas)

v) resultado salida: acción del proceso SCR; canal derecho del tampón anular.

Debido a la invención se consiguen las siguientes ventajas, entre otras, en comparación con el estado de la técnica:

- 1) Al conseguir una localización significativamente más precisa para todas las posiciones de la cabeza por la utilización de juego de datos genéricos, el ITD escalado individualmente incrementa esencialmente la plausibilidad de la simulación binaural.
- 2) La amortiguación de fase casi mínima debido a la extracción del ITD tiene ventajas auditivas significativas porque no ocurren prácticamente fenómenos de amortiguación durante la amortiguación de espectros de fase casi mínima.
- 3) Otra ventaja significativa adicional resulta de la situación de fase casi mínima de los espectros. Forma la base para todos los métodos eficientes espectrales y temporales de compresión e interpolación para HRTF/BRIR cuya prueba y posibilidad pueden funcionar, por ejemplo, en plataformas limitadas, tales como por ejemplo dispositivos móviles.
- 4) La degradación de calidad de la ITD permite una interpolación separada de características ILD - ITD -. De esta manera se puede resintetizar una ITD que es perceptivamente más importante con una precisión más fina, mientras que espectros de fase mínima son procesados con una precisión relativamente grosera (por ejemplo, 5°), por ejemplo. Esto simplifica la utilización en teléfonos móviles.
- 5) La distancia del tiempo de tránsito acústico de todos los juegos de datos BRIR conduce a una minimización de la latencia y a una reducción de las necesidades de memoria que hacen atractiva la aplicación en dispositivos móviles.
- 6) La implementación del ITD mediante remuestreo conduce a una modulación "más física" del campo acústico. La razón de ello es que cuanto más lenta es la reproducción dentro del proceso SCR proporciona (como mínimo para la corrección directa de sonido) limitación del efecto Doppler.
- 7) Debido a la división de la convolución en procesos individuales para cada fuente se crea la condición previa para un cálculo paralelo (incremento de eficiencia de cálculo).

La invención hace posible múltiples mejoras incluyendo las ventajas de amortiguación libre de errores, minimización de la latencia y efecto Doppler. Además, se debe poner énfasis en la añadidura de la elección de interpolación limitada en banda para la conversión de tasa de muestreo cuando se genera la ITD variable, la capacidad en tiempo real de manipulación de ITD, el método de extracción inaudible y la arquitectura PlugIn. Además, se debe poner énfasis en importantes aspectos, tales como la reducción efectiva y separada y la interpolación de ITD e ILD (es decir, espectros HRTF / BRIR).

El ajuste individual de las ITD independientes del tiempo de tránsito a los individuos específicos se puede realizar por escalado del factor de conversión corriente en el contexto de una conversión de tasa de muestreo que permite un ajuste dinámico para el cambio de la posición de la cabeza. Si se alcanza la condición estática, es decir, no cambia la posición de la cabeza, se dispone un nuevo ajuste estático de las ITD (automático). O bien el ajuste individual es realizado por un escalado de las ITD que corresponden a la posición de la cabeza previa y actual antes de calcular el factor de conversión (no escalable).

En otras palabras: la individualización y ajuste de la diferencia del tiempo de tránsito ITD \*entre las señales de salida L\* y R\* se puede conseguir por escalado del factor de conversión corriente específico (aplicación del factor ISF en el bloque VDL / SCR; ver figuras 5a/b). De manera alternativa, la individualización y ajuste de la diferencia de tiempo de tránsito ITD entre las señales de salida L\* y R\* se puede conseguir por escalado de las diferencias de tiempo de tránsito ITD correspondientes a las funciones de transferencias basadas en la cabeza (xBRIR\*).

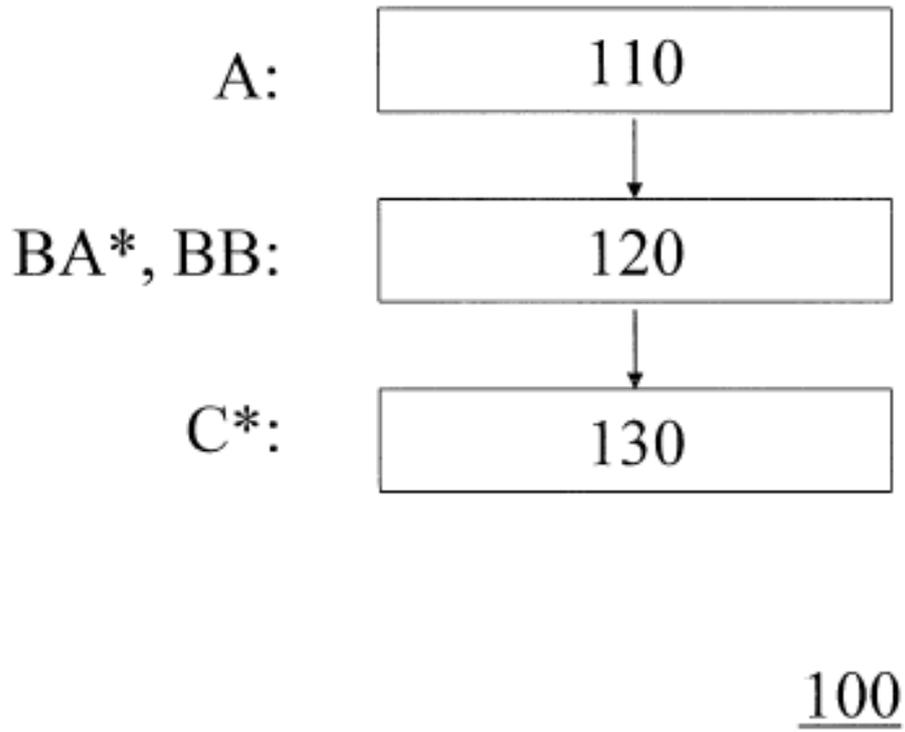
5 La invención es aplicable en muchas aplicaciones. El procedimiento propuesto constituye una mejora de todas las aplicaciones existentes en tiempo real para simulaciones acústicas binaurales (ambiente) (demostración en general de auditorio virtual 3D, acústica ambiental en juegos de ordenador, reuniones virtuales de conversación ("chat"), sistemas de guía y alerta binaural, circulaciones ("walk-throughs") binaurales, por ejemplo, a través de una exhibición de medios multimodales o de arquitectura virtual). Una posible aplicación es la añadidura subsiguiente de API de audio 3D comercial. También se puede mejorar la calidad de futuras aplicaciones ricas en medios que requieren un PC multimedia del lado del consumidor, tal como teleconferencia binaural vía VoIP o emisión binaural en continuo.

## REIVINDICACIONES

1. Procedimiento (100) para generar señales de audio binaurales individualmente ajustables ( $L^*$ ,  $R^*$ ), que comprende las siguientes etapas:
- 5 a partir de funciones de transferencia relacionadas con la cabeza ( $xRIR$ ) que son asignadas a varias posiciones de la cabeza ( $1...k$ ) se calculan funciones y transferencias relacionadas con la cabeza ( $xRIR$ ) libres de tiempo de tránsito por la extracción de tiempos de tránsito interaurales ( $TL$ ,  $TR$ ), y se calculan correspondientes diferencias de tiempo de tránsito ( $ITD$ ) (etapa 110);
- 10 por medio de síntesis binaural se generan señales de audio ( $L'$ ,  $R'$ ) libres de tiempo de tránsito, cada una de las cuales se refiere a una trayectoria de señal binaural, de manera que dependiendo de los datos de posición de la cabeza ( $HTDAT$ ) se utilizan las funciones de transferencia relacionadas con la cabeza ( $xRIR^*$ ) libres de tiempo de tránsito para una posición actual de la cabeza ( $1...k$ ) (etapa 120);
- 15 caracterizado porque
- a partir de las señales de audio ( $L'$ ,  $R'$ ) libres de tiempo de tránsito se generan señales de audio binaurales ajustables individualmente ( $L^*$ ,  $R^*$ ) cargando como mínimo una trayectoria de señal con la señal de audio generada ( $L'$ ) libre de tiempo de tránsito con un retardo de tiempo ( $VDL$ ) dependiendo de la diferencia de tiempo de tránsito ( $ITD$ ) calculada para la verdadera posición de la cabeza ( $1...k$ ) y dependiendo del factor de escalado individual ( $ISF$ ) (etapa 130);
- 20 porque cuando tiene lugar un cambio de la posición anterior de la cabeza ( $k=1$ ) pasando a la posición actual de la cabeza ( $k=2$ ), el retardo de tiempo ( $VDL$ ) entre un primer valor y un segundo valor es cambiado adaptativamente por una conversión de tasa de muestreo ( $SRC$ ), de manera que la conversión de tasa de muestreo aplica un factor de conversión para reproducir de modo adelantado o retrasado la señal de audio liberada de tiempo de tránsito (etapa 130); y
- 25 porque el factor de conversión utilizado para la conversión de tasa de muestreo ( $SRC$ ) se determina de acuerdo con el cambio de retardo de tiempo correspondiente al cambio de la posición anterior de la cabeza ( $k=1$ ) a la posición actual de la cabeza ( $k=2$ ) (etapa 130).
- 30
2. Procedimiento (100), según la reivindicación 1, caracterizado porque el retardo de tiempo ( $VDL$ ) corresponde a un valor que es calculado a partir de la ponderación de la diferencia de tiempo de tránsito ( $ITD$ ) calculada a partir de la posición actual de la cabeza ( $1...k$ ) y el factor de escalado individual ( $ISF$ ) (etapa 130).
- 35
3. Procedimiento (100), según la reivindicación 1, caracterizado porque la diferencia de tiempo de tránsito ( $ITD^*$ ) que se produce entre señales de audio binaurales ajustables individualmente ( $L^*$ ,  $R^*$ ) se ajusta individualmente por el escalado del factor actual (etapa 130).
- 40
4. Procedimiento (100), según la reivindicación 3, caracterizado porque la diferencia de tiempo de tránsito que se produce ( $ITD^*$ ) es ajustada individualmente por escalado de las diferencias de tiempo de tránsito ( $ITD$ ) correspondientes a las funciones de transferencia relacionadas con la cabeza ( $xRIR^*$ ) liberadas de tiempo de tránsito.
- 45
5. Procedimiento (100), según una de las reivindicaciones anteriores, caracterizado porque las funciones de transferencia relacionadas con la cabeza ( $xRIR^*$ ) libres de tiempo de tránsito son calculadas para una pluralidad de posiciones de la cabeza ( $1...k$ ) y para una pluralidad de fuentes y señal de audio ( $1...n$ ) (etapa 110).
- 50
6. Procedimiento (100), según una de las reivindicaciones anteriores, caracterizado porque los tiempos de tránsito interaurales ( $TL$ ,  $TR$ ) son extraídos de las funciones de transferencia relacionadas con la cabeza ( $xRIR$ ) por medio de uno de los procedimientos siguientes: procedimiento de inicio, procedimiento de correlación cruzada interaural, constitución de diferencia de retraso de grupo específico de frecuencia (tiempo de tránsito), ajuste del gradiente de fase específico del rango de frecuencia o de terminación de la fase en exceso por la transformación de Hilbert (etapa 110).
- 55
7. Procedimiento (100), según la reivindicación 6, caracterizado porque el método de inicio se aplica de manera que se determina un suelo de ruido de medición natural y a continuación se determina un umbral situado por encima que es menor que el máximo absoluto del impulso - respuesta resultante de la función de transferencia correspondiente relacionada con la cabeza (etapa 110).
- 60
8. Procedimiento (100), según la reivindicación 7, caracterizado porque el umbral se determina con un valor mínimo de 10dB por encima del suelo de ruido de medio natural y/o porque el umbral es determinado con
- 65

un valor mínimo de 10dB por debajo del máximo absoluto del impulso - respuesta resultante de la función de transferencia correspondiente relacionada con la cabeza (etapa 110).

- 5 9. Procedimiento (100), según la reivindicación 8, caracterizado porque el umbral se encuentra en un rango promedio de la dinámica de medición (etapa 110).
- 10 10. Procedimiento (100), según la reivindicación 6, caracterizado porque el método de inicio se aplica de manera que se detecta como inicio un valor de porcentaje del máximo absoluto del impulso - respuesta resultante de la función de transferencia correspondiente relativa a la cabeza, en particular como valor comprendido entre 10% y 90% (etapa 110).
- 15 11. Dispositivo para llevar a cabo el procedimiento de una de las reivindicaciones anteriores en el que el dispositivo para generar señales de audio binaurales ( $L^*$ ,  $R^*$ ) para percepción espacial, comprende:
- 20 un primer circuito (A) para la extracción de funciones ( $xRIR$ ) de transferencia relacionadas con la cabeza, que son asignadas a diferentes posiciones de la cabeza ( $1...k$ ) funciones de transferencia relacionadas con la cabeza ( $xRIR^*$ ) liberadas de tiempo de tránsito y para el cálculo de las diferencias de tiempo de tránsito (ITD) de las mismas;
- 25 un segundo circuito (BB) para la generación por medio de señales de audio de síntesis binaural ( $L'$ ,  $R'$ ) liberadas de tiempo de tránsito, cada una de las cuales se refiere a una trayectoria señal binaural, de manera que este circuito utiliza dependiendo de datos de la posición de la cabeza (HTDAT) las funciones de transferencia relacionadas con la cabeza ( $xRIR^*$ ) liberadas de tiempo de tránsito para una posición real de la cabeza ( $1...k$ );
- 30 caracterizado por
- como mínimo un tercer circuito ( $BA^*$ ;  $C^*$ ) para generar a partir de señales de audio ( $L'$ ,  $R'$ ) liberadas de tiempo de tránsito las señales de audio binaurales ajustables individualmente ( $L^*$ ,  $R^*$ ) cargando como mínimo una trayectoria de señal con la señal de audio generada ( $L'$ ) liberada de tiempo de tránsito con un tiempo de retardo (VDL) con dependencia de la diferencia de tiempo de tránsito (ITD) calculada a partir de la posición actual de la cabeza ( $1...k$ ) y dependiendo del factor de escalado individual (ISF);
- 35 en el que él como mínimo tercer circuito ( $BA^*$ ;  $C^*$ ), cuando tiene lugar un cambio de la posición anterior de la cabeza ( $k=1$ ) a la posición actual de la cabeza ( $k=2$ ), cambia adaptativamente el retardo de tiempo (VDL) entre un primer valor y un segundo valor por medio de una conversión de tasa de muestreo (SRC), en la que la conversión de tasa de muestreo aplica un factor de conversión para reproducir con adelanto o retardo la señal de audio liberada de tiempo de tránsito; y
- 40 determina el factor de conversión utilizado para la conversión de la tasa de muestreo (SRC) de acuerdo con el cambio de retardo de tiempo correspondiente al cambio de la posición previa de la cabeza a la posición actual de la cabeza.



**Fig. 1**

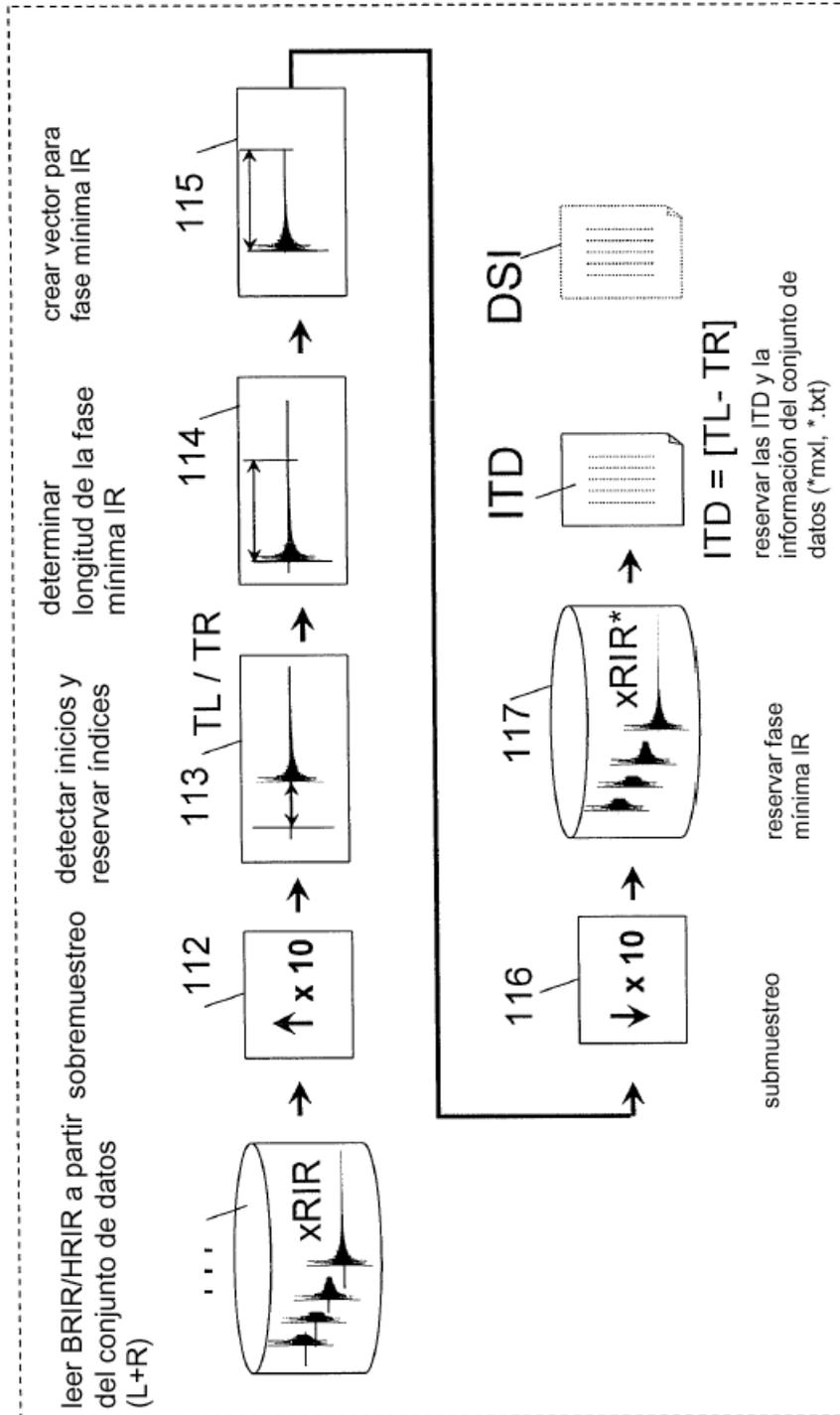


Fig. 2

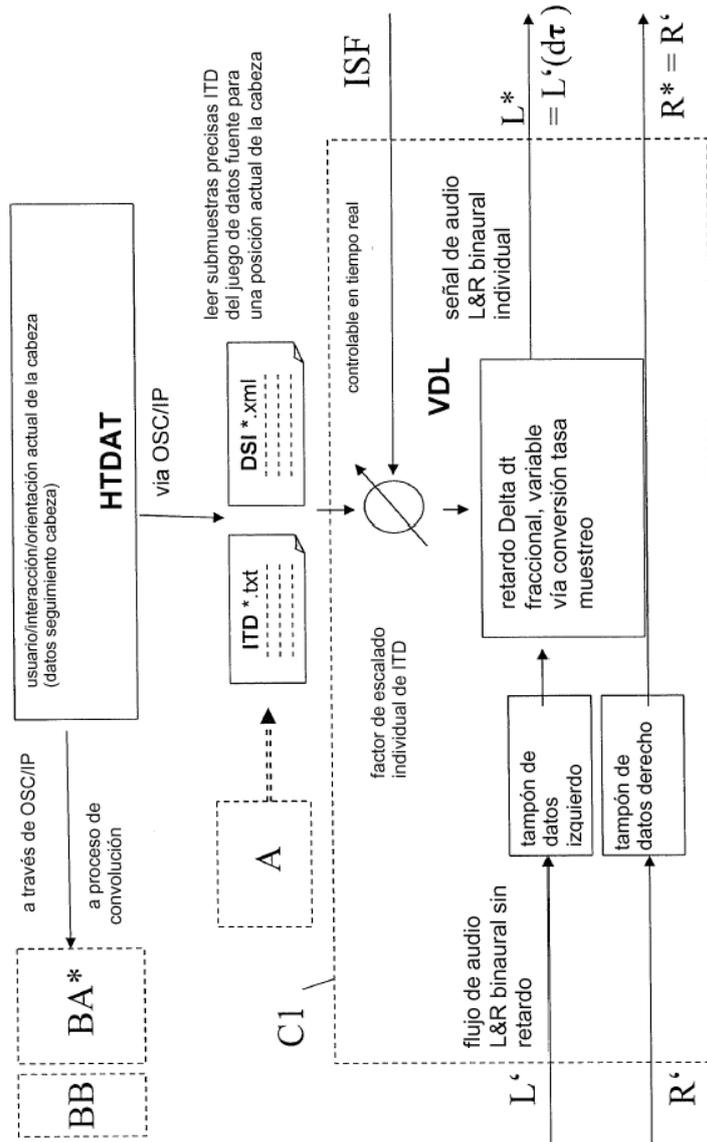
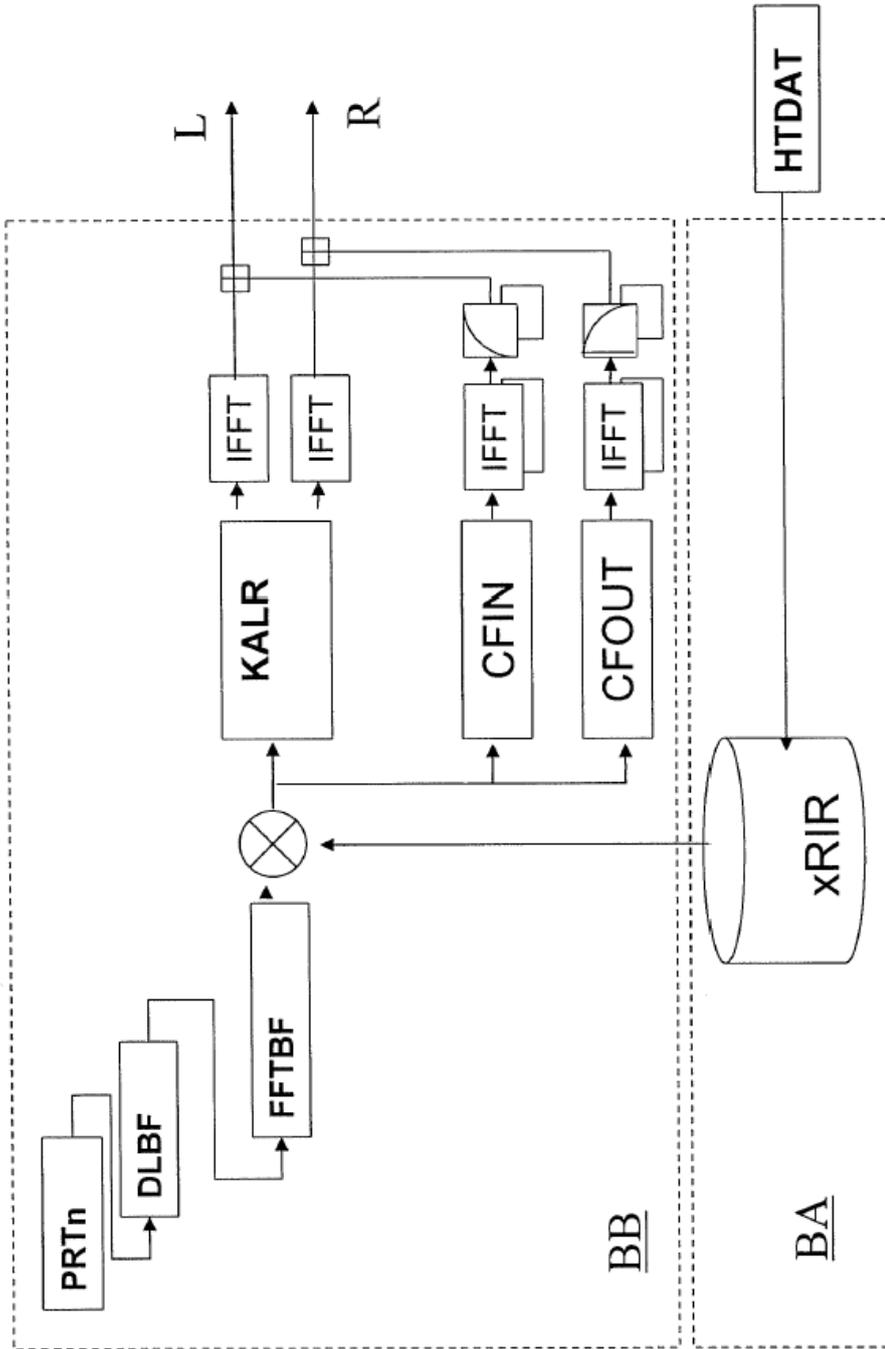


Fig. 3



**Fig. 4a**

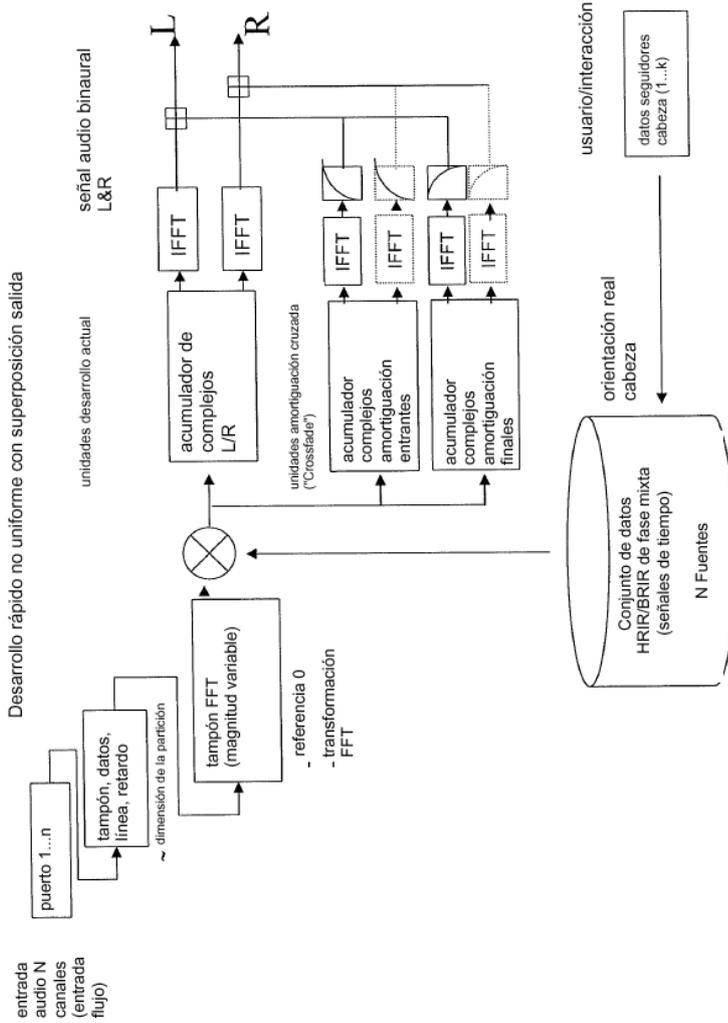


Fig. 4b

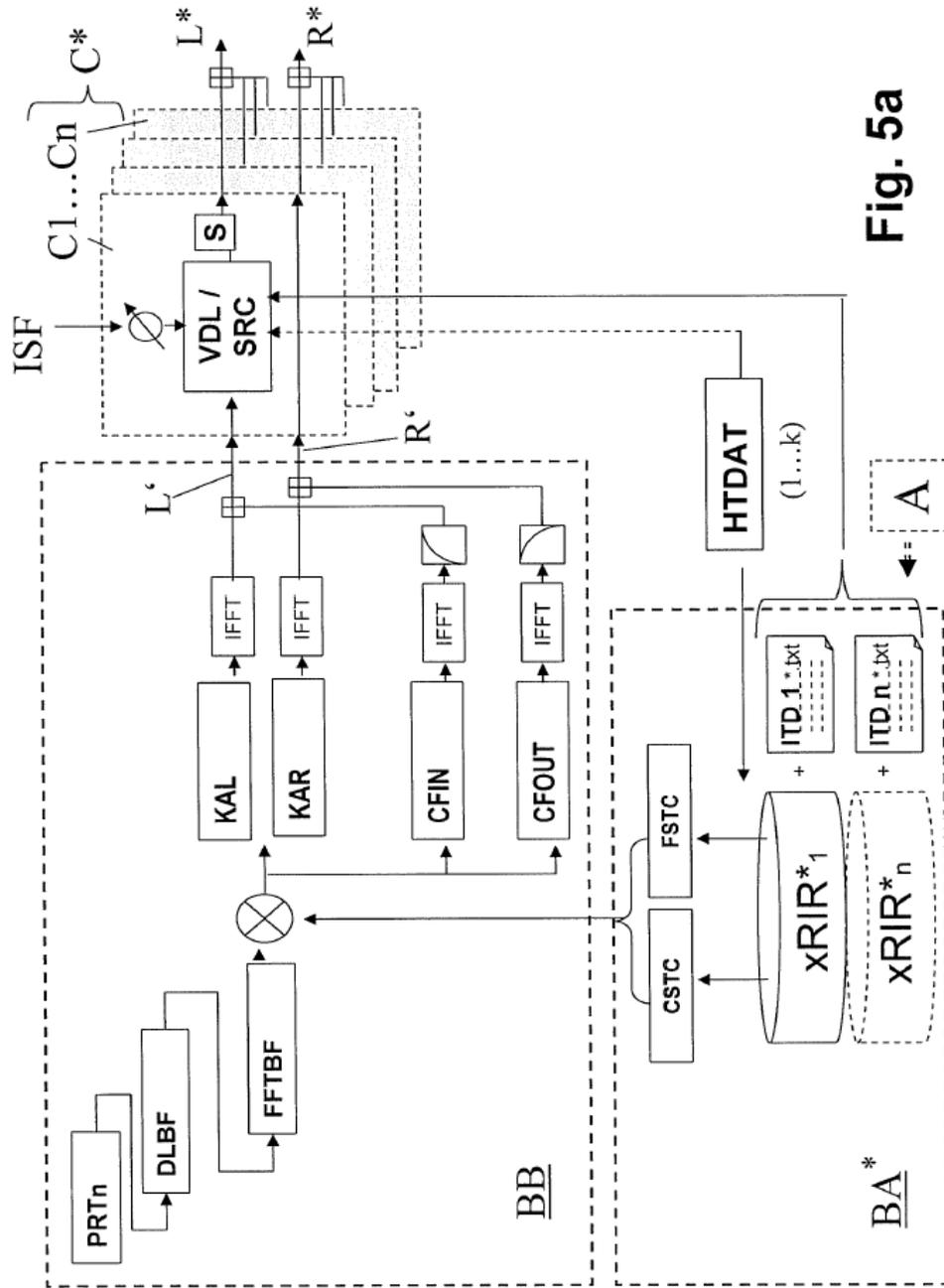


Fig. 5a

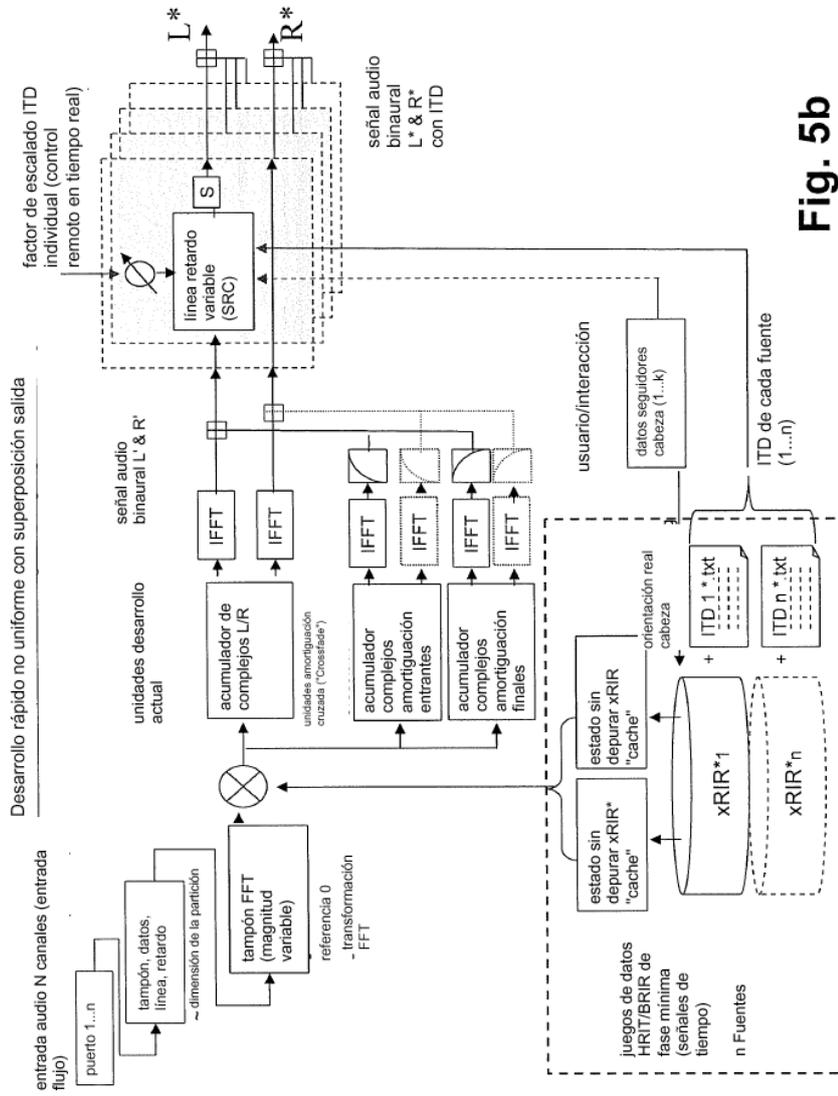


Fig. 5b