(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 741 303**

(51) Int. CI.:

A61F 11/04 (2006.01)
H04R 25/00 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **21.11.2011 E 17162598 (1)**

(97) Fecha y número de publicación de la concesión europea: **15.05.2019 EP 3245990**

(54) Título: **Procedimiento y aparato de sustitución sensorial**

(30) Prioridad:

23.11.2010 IE 20100739

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
10.02.2020

(73) Titular/es:

NATIONAL UNIVERSITY OF IRELAND, MAYNOOTH (100.0%)
National University of Ireland
Co. Kildare, Maynooth, IE

(72) Inventor/es:

O'GRADY, PAUL;
O'NEILL, ROSS y
PEARLMUTTER, BARAK A.

(74) Agente/Representante:

CARPINTERO LÓPEZ, Mario

DESCRIPCIÓN

Procedimiento y aparato de sustitución sensorial

Campo técnico

La presente invención se refiere a procedimientos y aparatos para sustitución sensorial, con particular aplicación en el tratamiento de acúfenos.

Antecedentes de la técnica

Los acúfenos son un comportamiento neurológico anormal que surge de la pérdida de señal a través del oído. Aunque la causa precisa de los acúfenos no se entiende totalmente, ciertas analogías se emplean para describir las causas probables. Por ejemplo, se cree que los acúfenos se provocan a menudo por un impedimento auditivo físico tal como daños en los cabellos en la cóclea. En un intento por compensar la pérdida de información auditiva, el cerebro eleva la amplificación y las ganancias en bucles recurrentes hasta tal punto que se generan señales falsas, de manera similar en principio al ruido resonante que puede ocurrir cuando el volumen de un amplificador de audio en un auditorio se eleva demasiado alto. Como alternativa, se puede imaginar una bomba de agua eléctrica cuyo suministro de agua se limita súbitamente. La bomba oscila y vibra en intento desesperado por compensar la pérdida de entrada. Puede pensarse que los acúfenos surgen esencialmente a partir de los mismos tipos de mecanismos: pérdida de señal a través de los oídos que resulta en una actividad oscilatoria y espontánea incrementada en las neuronas asociadas en el cerebro. Esta actividad se percibe como un sonido ilusorio por el que lo sufre.

Los que sufren de acúfenos son significativamente más proclives a percibir un efecto posterior de audio ilusorio conocido como el tono Zwicker. El tono Zwicker se induce mediante la exposición del individuo a un ruido de espectro amplio (20 Hz – 20 KHz) con un hueco espectral (silencio) a una frecuencia arbitraria. Cuando el ruido se retira, el individuo percibe un "zumbido" a la frecuencia del hueco espectral. Esto sugiere que para compensar la sensibilidad coclear desigual por las frecuencias, el cerebro introduce sensibilidad dependiente de la frecuencia o ganancia similar en un "ecualizador gráfico" en un estéreo. A las frecuencias en que nuestra cóclea es menos sensible, el cerebro incrementa la ganancia en esa banda de frecuencia para compensar. En bandas de frecuencia donde la sensibilidad cae por debajo de un umbral mínimo, el cerebro incrementa la ganancia a niveles patológicos. Esto se manifiesta como ruido ilusorio, zumbido o incluso oscilación caótica, los efectos más comúnmente descritos de los acúfenos. Un número muy grande de tratamientos se han propuesto para los acúfenos, incluyendo radiocirugía, estimulación directa de los nervios auditivos, tratamientos farmacológicos, tratamientos psicológicos y tratamiento mediante la reproducción de sonidos externos al paciente. Aunque muchos de tales tratamientos proporcionan un alivio en algunos grupos de pacientes, actualmente no existe un tratamiento fiable para todos los pacientes, y la presente invención pretende proporcionar un enfoque alternativo adicional.

El documento US2009/306741 A se refiere a sistemas y procedimientos para gestión de funciones cerebrales y corporales y percepción sensorial. Por ejemplo, el documento US2009/306741 proporciona sistemas y procedimientos de sustitución sensorial y mejora sensorial (aumento) así como mejora del control motor.

El documento US2009/270673 A1 desvela procedimientos y sistemas para tratamiento de acúfenos donde un dispositivo se acopla a una superficie de un hueso o a un diente o varios dientes. Tal dispositivo puede comprender un aparato oral con un conjunto electrónico y/o de transductor para generar sonidos mediante un elemento transductor de vibración. En general, el transductor puede programarse para ajustar cualquier números de ajustes y opciones de tratamiento para generar una o más frecuencias de sonido mediante el transductor accionable para transmitir una señal de audio modificada mediante conductancia vibratoria a un oído interno del paciente para enmascarar los acúfenos. La señal de audio se modifica además para compensar cualquier pérdida auditiva del paciente así como un umbral de sensibilidad ósea medido desde el paciente y calibrado por el dispositivo de programación.

M. Gilligan, "Mutebutton (Video presentation of tinnitus treatment system)", (20101020), Vimeo, URL: http://vimeo.com/16020279, (20120120), XP002667771 [I] presenta una visión de conjunto de un sistema de tratamiento de acúfenos.

Divulgación de la invención

La invención se define en las reivindicaciones independientes 1 y 15 y en las reivindicaciones dependientes 2-14.

Se proporciona un aparato como se describe en este documento en referencia a la reivindicación 1 para el uso en el tratamiento de acúfenos, que comprende una unidad de procesamiento de sonido, una unidad táctil y una interfaz entremedias, en el que dicha unidad táctil comprende una agrupación de estimuladores (16) electrocutáneos, cada uno de los cuales puede accionarse independientemente para aplicar un estímulo táctil a un sujeto, y una entrada para recibir una pluralidad de señales de accionamiento desde dicha interfaz y que dirige señales de accionamiento individuales a estimuladores individuales; y dicha unidad de procesamiento de sonido comprende: una entrada configurada para recibir una señal de audio; un procesador (14) de señal digital operable para analizar dicha señal de audio y generar dicha pluralidad de señales de accionamiento representativas de dicha señal de audio al dividir dicha señal de audio en una serie de tramas en el dominio de tiempo, realizando una transformada en cada trama para

generar un conjunto de coeficientes que representan dicha trama, y mapear dicho conjunto de coeficientes en un conjunto de señales de accionamiento a aplicar a la agrupación; y una salida configurada para recibir dicha pluralidad de señales de accionamiento desde dicho procesador de señal digital y proporcionar dicha pluralidad de señales de accionamiento a dicha interfaz, el aparato configurado además para enviar la señal de audio, enviándose dicha señal de audio de forma simultánea con la aplicación de las señales de accionamiento a la agrupación.

Preferentemente, dicho procesador de señal digital es operable además para generar dicha pluralidad de señales de accionamiento como una secuencia variable en el tiempo de patrones de agrupación de salida, en el que cada patrón de agrupación de salida comprende un conjunto de señales de accionamiento a aplicarse a la agrupación durante un periodo de tiempo discreto, representativas de una muestra de tiempo discreta de la señal de entrada. De acuerdo con una realización, dicho procesador de señal digital se programa para analizar dicha señal de audio dividiendo dicha señal de audio en una serie de tramas en el dominio de tiempo, realizando una transformada en cada trama para generar un conjunto de coeficientes que representan dicha trama y mapeando dicho conjunto de coeficientes a un conjunto de señales de accionamiento a aplicar en la agrupación. Dicha transformada realizada en cada trama se selecciona preferentemente desde una transformada de fourier, una transformada de fourier de tiempo corto (STFT), una transformada de ondículas, una transformada curvelet, una transformada gammatone y una transformada zak.

Más preferentemente, dicha transformada es una transformada de fourier o una transformada de fourier de tiempo corto, y en el que dicha señal se muestrea a un índice de muestreo de entre 4 kHz y 12 kHz, más preferentemente entre 6 kHz y 10 kHz, y más preferentemente aproximadamente 8 kHz.

Adecuadamente, dichas series variables en el tiempo de tramas puede superponerse entre sí.

El inicio de cada trama se desvía preferentemente del inicio de la trama precedente entre 10 y 20 ms, más preferentemente 12-18 ms y más preferentemente aproximadamente 16 ms.

El procesador se programa preferentemente para emplear una longitud de trama desde 18 a 164 ms, más preferentemente desde 50 a 150 ms y más preferentemente 64 a 128 ms.

El conjunto de coeficientes representa preferentemente la señal en el dominio de frecuencia y los coeficientes se mapean a las señales de accionamiento de manera que los coeficientes que representan frecuencias similares se mapean a señales de accionamiento dirigidas a estimuladores que están físicamente cerca entre sí en dicha agrupación.

Más preferentemente, los coeficientes que representan frecuencias cercanas se mapean a señales de accionamiento dirigidas a estimuladores que están físicamente adyacentes entre sí.

En realizaciones alternativas, el procesador de señal digital se programa para analizar dicha señal de audio mapeando segmentos sucesivos de dicha señal de audio a un conjunto de características seleccionadas desde un diccionario de dichas características.

La agrupación de estimuladores puede, por ejemplo, ser una disposición rectangular de m x n estimuladores regularmente separados, una disposición hexagonal de subagrupaciones hexagonales concéntricas o una disposición circular de subagrupaciones circulares concéntricas.

Preferentemente, dicho procesador es operable además para normalizar las magnitudes de las señales de accionamiento para que entren dentro de un intervalo predeterminado de intensidades de señal de accionamiento.

En realizaciones preferentes, dicha unidad táctil tiene la forma de un cuerpo dimensionado para colocarse en la lengua de un sujeto humano, y en el que cada estimulador tiene la forma de un electrodo que tiene una superficie redondeada que se proyecta desde dicho cuerpo.

Más preferentemente, la superficie redondeada de cada electrodo es generalmente hemisférica.

La realización preferente usa una agrupación de electrodos basada en lengua como un dispositivo de sustitución sensorial auditivo, por lo que la información de audio se presenta al cerebro mediante la estimulación táctil aplicada a la lengua. El sistema se compone de un dispositivo de representación electrotáctil inalámbrico y un ordenador de procesamiento de audio, que de manera inalámbrica transmite imágenes de estímulos electrotáctiles para representarse usando tecnología Bluetooth en el sistema de representación electrotáctil. Como alternativa, ambos componentes pueden combinarse en una única unidad para una portabilidad añadida. Además, el estímulo táctil generado mediante el sistema puede presentarse en cualquier superficie cutánea en el cuerpo para ese asunto.

Descripción detallada de realizaciones preferentes

En referencia a la Fig. 1 se indica, generalmente en 10, un aparato para tratar acúfenos, que comprende una o más de una pluralidad de fuentes 12 de entrada de audio, un módulo 14 de procesamiento de señal y una agrupación 16 electroestimuladora.

Las fuentes de audio pueden ser de cualquier tipo, y por motivos ilustrativos, la Fig. 1 muestra tres de tales opciones:

una fuente 18 de audio de abordo, tal como un conjunto de archivos MP3 y un descodificador de audio integrado, un micrófono 20 para recibir señales de sonido ambiental o una conexión a una fuente 22 de audio externa, tal como una tarjeta de sonido de un ordenador. El sistema puede tener una pluralidad de tales fuentes, por ejemplo, una fuente de audio incorporada para el uso en sesiones de entrenamiento o de tratamiento activo, un micrófono para procesamiento de sonidos en el entorno externo del usuario y un conector hembra de entrada (p. ej., un conector hembra estándar de 3,5 mm) para la conexión a fuentes de audio externas. Otros procedimientos de entrada tal como un conector de iPod patentado, o un conector para otro reproductor MP3, o una entrada de audio óptica, también pueden proporcionarse (iPod es una marca registrada de Apple Inc. de Cupertino, California).

La Fig. 2 muestra una realización física de tal sistema, que comprende una carcasa 24 que contiene una fuente 18 de audio de abordo y un módulo 14 de procesamiento de señal (Fig. 1, no mostrado en la Fig. 2), una agrupación 16 electroestimuladora que tiene una agrupación de 16 x 16 electrodos en un sustrato dimensionado para colocarse en la lengua humana, una cinta 26 conectora que transporta 256 señales de accionamiento individuales desde la carcasa 24 a los electrodos individuales de la agrupación 16 y un cordón 28 para colgar la carcasa alrededor del cuello del sujeto.

En referencia ahora a la Fig. 3, la entrada de audio se procesa mediante el módulo 14 de procesamiento de señal de la Fig. 1, que produce el conjunto necesario de 256 señales de accionamiento para representar la entrada de audio. El módulo 14 de procesamiento de señal tiene los siguientes módulos funcionales, cada uno de los cuales se describirá a continuación en más detalle: una señal de entrada de audio x(t) recibida como una señal con muestreo de tiempo se somete al muestreo y estructuración 30, después a un procedimiento de descomposición de señal y análisis 32, proporcionando un conjunto de coeficientes o de símbolos representativos del sonido que se someten a una disposición espacial de coeficientes y modelado de receptor inverso 34, seguido por un escalado de valores de coeficientes y conversión de tipo 36 y un procedimiento de corrección y monitorización de recorte repetitivo 38.

Muestreo y estructuración

En referencia a la Fig. 4, el sistema recibe audio con muestreo de tiempo x(t). Los datos con muestreo de tiempo se disponen en pedazos conocidos como tramas en una función 40 de superposición y estructuración que forma parte de la función 30 de muestreo y estructuración de la Fig. 3. Se indica una única trama usando la notación de matriz como x, donde el siguiente procesamiento se aplica a cada trama individualmente, en oposición al procesamiento de todo el flujo de audio de una sola vez.

Como es típico en el análisis de audio, es necesario que el tamaño de trama especificado (es decir, la longitud de ventana de análisis) sea consistente con la duración de los objetos de audio contenidos en el audio en consideración; por ejemplo, fonemas de diálogo, notas musicales. El tamaño de trama, que se mide en las muestras y que se indica con N, depende del índice de muestreo y es típicamente una potencia de dos. Además, es aconsejable mejorar los efectos de límite de trama usando una función de ventana, y se especifica una función de ventana Hamming. Las tramas no se limitan a ser contiguas en el flujo de audio y pueden superponerse. Para esta realización se elige un índice de muestreo, fs, de 8000 Hz, que se corresponde con un ancho de banda de señal de 4000 Hz, que se conoce que captura suficiente información de frecuencia de manera que la señal reconstruida es inteligible como audio. Desde el análisis del corpus de diálogo TIMIT, que incluye tanto frases masculinas como femeninas, se usaron las siguientes estadísticas que pertenecen a una longitud de fonema como una guía para el tamaño de trama: Máxima longitud = 164 ms, longitud mínima = 18ms, longitud promedia = 81 ms, longitud media = 67 ms, y se tiene en mente un tamaño de trama de 512 (64 ms) o 1024 (128 ms) muestras de audio.

Una alternativa posible sería una realización de transmisión por secuencias no basada en tramas, en la que una sucesión de filtros no lineales extraen información deseada desde la forma de onda de audio de entrada con las características que se alinean con muestras o incluso se subalinean con muestras, en lugar de estar limitadas a o alinearse con límites de trama. En este escenario, N puede ser igual a una única muestra.

En el contexto del dispositivo de representación de agrupación electrotáctil, cada trama producirá una única “imagen” de agrupación electrotáctil (o patrón de agrupación de salida), que va a representarse en el dispositivo. Las tramas consecutivas crean una secuencia de tales imágenes y es análogo a representar una película en una pantalla de televisión. Para este sistema, se usó una agrupación que contiene 16 x 16 electrodos, donde cada electrodo tiene un intervalo dinámico entre 0 a 255 niveles de tensión, que se corresponde con un tipo de datos char sin firma en el lenguaje de programación C. Como se ha analizado antes, el índice de muestreo depende en gran medida del ancho de banda de las señales en consideración. Sin embargo, el índice de actualización de la representación electrotáctil también tiene que considerarse: tomando una pantalla de televisión como un ejemplo, que explota la persistencia de la visión, si el índice de actualización es demasiado bajo existen saltos notables entre escenas. De manera similar, con la representación electrotáctil, si el índice de actualización es demasiado bajo puede haber una falta de continuidad entre imágenes consecutivas y la información presentada exhibirá saltos similares entre tramas. Al contrario, si el índice es demasiado alto, las tramas individuales consecutivas pueden percibirse como una debido a los límites de ancho de banda sensoriales de la lengua, y no obtendrán ningún beneficio. En resumen, existe una compensación entre la frecuencia de muestreo y el índice de actualización del dispositivo, que depende del tamaño de agrupación/imagen.

Para el sistema, se especifica un índice de trama de imagen (índice de actualización de visualización) de 62,5 tramas por segundo, que es consistente con la persistencia de visión, y se decide en un tamaño de trama de 512 muestras (aquí el tamaño de trama es dos veces el tamaño de la agrupación; a continuación, se analiza cómo 256 coeficientes se generan para la agrupación usando el espectrograma de magnitud).

Además, para lograr patrones de estímulo consistentes entre imágenes electrotáctiles consecutivas, la superposición de tramas es necesaria; donde las tramas de longitud N se superponen mediante muestras o, donde cada a = N-o muestras, la ventana de análisis se desliza para incluir las nuevas muestras, en cuyo punto una trama se procesa y una nueva imagen se genera y transmite al sistema de representación. Por tanto, el sistema de representación se actualiza cada número de a muestras, donde a representa un avance de trama (en oposición a la superposición).

Se especifica un índice de trama de 62,5 Hz, que en un índice de muestra de 8 kHz corresponde a una duración de 16 ms. Sin embargo, ya que el tamaño de la trama es de 64 ms de duración (por lo que los objetos de audio pueden capturarse), se superponen tramas como se recomienda anteriormente donde N = 512 y a = 128, lo que satisface la duración de índice de trama de 16 ms.

Finalmente, se resume usando la ilustración cuantitativa siguiente: se actualiza la agrupación a 62,5 Hz, y se desea incluir frecuencias de hasta 4000 Hz. Con esto como la frecuencia de Nyquist, esto implica un índice de muestreo de 8 kHz. Se usan tramas que contienen 4 actualizaciones, lo que significa que cada trama contiene:

4 actualizaciones/trama x 8000 muestras = 512 muestras/trama 62,5 actualizaciones

con tramas superpuestas por lo que una nueva trama se crea cada actualización, lo que en este caso sería cada 128 muestras.

Ya que el tamaño de las tramas es 512 y el tamaño de agrupación es 256 (se recuerda que una trama corresponde a una imagen de agrupación), se genera una representación de coeficiente 256 para representarse en la agrupación usando una descomposición de señal apropiada como se analiza a continuación.

Descomposición de Señal y Análisis

Desde la introducción de la teoría de la información a mediados del último siglo (Shannon, 1948), se ha sugerido que la redundancia de entradas sensoriales es importante para entender la percepción (Attneave, 1954; Barlow, 1959). La redundancia contenida en nuestros entornos es lo que permite que el cerebro aumente los modelos cognitivos del entorno a su alrededor (Barlow, 1961, 1989). Se piensa que las regularidades estadísticas en entradas sensoriales deben de alguna manera separarse de las redundancias y codificarse de una manera eficiente, y esto ha conducido a la hipótesis de Barlow, que afirma que el fin de cada procesamiento perceptual es transformar la entrada sensorial altamente redundante en un código factorial más eficiente, lo que puede mencionarse como la siguiente transformación de matriz,

Ecuación 1

$$S = Wx$$

donde W = [w1|...|wN] es un operador de matriz lineal NxN, x = [x1,...,xN] T es el vector de entrada sensorial y s = [s1,...,sN] T es la entrada codificada, donde los valores de W, s, x son números reales, es decir W, s, x son elementos de R.

Tal transformación se realiza mediante una operación de multiplicación de matriz-vector, y puede usarse en la sustitución sensorial auditiva, donde la x anterior corresponde a la trama, y s es la salida codificada, que va a representarse como tensiones en un sistema de representación electrotáctil.

En referencia a la Fig. 5, se descompone el flujo de audio (después del muestreo y estructuración) en la superposición de ondas básicas mediante Análisis de Armónicos. El tipo más común de Análisis de Armónicos es el análisis de Fourier mediante la transformada de Fourier de tiempo corto discreta (STFT), es decir, la descomposición de una señal en ondas sinusoidales, donde W corresponde a una base de Fourier en el dominio complejo, C, lo que produce un valor s elemento de C que se denomina s_c. La función STFT se representa simbólicamente mediante la función 42 en la Fig. 5.

Para transformar los coeficientes, s_c, a una forma adecuada para la representación en la agrupación electrotáctil, se emplea un espectrograma 44 de magnitud de los coeficientes STFT resultantes, y se representan esos valores. Una ventaja adicional de usar el espectrograma de magnitud es que los coeficientes resultantes, s_n, no son negativos (es decir, se excluyen los valores negativos, $s \geq 0$);

Ecuación 2

$$S_n = mag(s_c)$$

Los valores negativos no pueden representarse en la agrupación electrotáctil, ya que las tensiones de electrodos

representan intensidad, y por tanto los coeficientes a representar en el dispositivo deben tener la forma s_n.

Ya que la STFT tiene como resultado valores simétricos, solo la primera mitad de la trama STFT se necesita para generar el espectrograma de magnitud, teniendo como resultado 256 coeficientes a representar en la agrupación, uno para cada electrodo.

Juntas, la función 42 STFT y la función 44 de espectrograma de magnitud proporcionan la función 32 de descomposición de señal generalizada mostrada en la Fig. 3.

Otras alternativas posibles incluyen las nociones generalizadas de análisis de Fourier tales como transformadas de Ondículas, transformadas Gammatone, transformadas Zak, transformadas Curvelet, etc., que también pueden usarse para representar audio en la agrupación electrotáctil sustituyendo W por las bases que definen estas transformadas.

Tal como se evidencia mediante los ensayos piloto con 20 pacientes de acúfenos (tal como se describe a continuación) usando el sistema de las Figs. 1 y 2, es evidente que con el paso del tiempo un sujeto aprende a asociar el estímulo (coeficientes de espectrograma de magnitud) presentado por la representación electrotáctil en la lengua con sonidos individuales. La premisa del tratamiento es que el córtex somatosensorial, que recibe la información táctil desde la lengua, tiene suficientes trayectorias al córtex auditivo para realizar esta correlación.

En el contexto del tratamiento de acúfenos, ya que el cerebro no tiene referencia externa para el entorno auditivo, se demuestra mediante los ensayos que proporcionar una referencia adicional para el entorno de audio por medio de sustitución sensorial auditiva permite que el cerebro recalibre los mecanismos compensativos a través de la plasticidad del cerebro, mejorando así el efecto de los acúfenos.

Disposición espacial de coeficientes

Ya que el audio es una señal de una dimensión y se desea representar en una agrupación electrotáctil de dos dimensiones, es necesario realizar una incrustación de los datos de una dimensión en un espacio en dos dimensiones. La salida desde la fase de descomposición de señal es también de una dimensión. Encontrar una topología apropiada es una tarea fuera de línea, donde la disposición resultante se representa mediante una matriz de permutación N x N. Durante el tiempo de ejecución, los coeficientes s_n se recolocan mediante la realización de una multiplicación de matriz con la matriz de permutación,

Ecuación 3

$$S_n = Ps_n$$

En el contexto de la descomposición de señal usando una base de Fourier, se disponen preferentemente los vectores de manera tonotópica, es decir, donde los componentes que están cerca entre sí en términos de frecuencia se colocan en proximidad entre sí en la agrupación. Donde (0,0) en la notación de matriz (izquierda superior) se corresponde con el componente de frecuencia más grande, mientras que ($\sqrt{N}$, ($\sqrt{N}$) (derecha inferior) se corresponde con el componente de frecuencia más bajo. Esencialmente, la trama se divide en $\sqrt{N}$ vectores, que después se usan para construir las filas de la matriz/imagen a representar en la agrupación electrotáctil.

Modelado receptor inverso y moldeo de señal

En referencia a las Figs. 3 y 6, después de la descomposición de señal y la disposición de coeficientes, puede ser necesario realizar un postprocesamiento de las activaciones de señal resultantes para moldearlas hasta una forma adecuada para que el audio pueda ser más fácilmente perceptible a través de la lengua, como se representa en 36. Por ejemplo, en el contexto de un postprocesamiento de audio, es a veces necesario comprimir el intervalo dinámico del audio (usando una función no lineal), que estrecha la diferencia entre niveles de audio altos y bajos, por lo que las señales silenciosas pueden oírse en entornos ruidosos. Puede ser necesario moldear las señales presentadas al sistema de representación electrotáctil de manera similar, para asegurarse de que la lengua es capaz de escuchar todas las señales en la descomposición de audio siguiente. En su nivel más básico, tal moldeo de señal realiza una amplificación del estímulo en un nivel por electrodo basándose en la sensibilidad de la lengua en esa región a tal estímulo.

Una consideración importante cuando se transforma el audio con muestreo de tiempo en otro dominio mediante la transformada de base (Ecuación 1) es que los coeficientes resultantes experimentan una expansión numérica, que es generalmente indeterminada. Por ejemplo, una entrada de audio se normaliza a entre -1 y +1 cuando viene desde la tarjeta de sonido. Sin embargo, una descomposición de señal de la onda puede producir valores que son mayores que este intervalo, y este comportamiento se denomina expansión numérica. Esto se representa en la Fig. 6 como ocurre entre la descomposición 32 de señal y la disposición 34 de coeficientes.

Por último, los coeficientes se representarán en la agrupación, donde cada electrodo tiene el intervalo dinámico de un tipo char sin firma (0 a 255). Para asegurar de que todos los coeficientes encajan dentro de este intervalo, es necesario realizar un escalado de los coeficientes, ya sea antes o después de la descomposición de señal, para que los coeficientes que se van a representar no tengan como resultado un recorte de la agrupación electrotáctil después de

la conversión a un tipo char sin firma, es decir, los coeficientes a representar no superan el intervalo dinámico de los electrodos. Por tanto, el escalado/normalización, por ejemplo, sμ → s donde μ es el factor de escalado, se requiere para asegurar que los coeficientes que se representan en el dispositivo residen en el intervalo dinámico de la intensidad de tensión de electrodo.

El procesamiento de audio se realiza usando un número de punto flotante de doble precisión (8 bytes). Sin embargo, los electrodos en la agrupación pueden representar intensidades de tensión de 0-255 (1 byte), lo que se corresponde con un char sin firma. La monitorización de coeficiente y el escalado aseguran que los coeficientes (de tipo doble) residen en el intervalo 0->1, que después se convierten a char sin firma para la presentación en la agrupación multiplicando por 255. 0 arroja entonces la variable de coeficiente a un char sin firma. Este procedimiento se muestra de manera repetitiva entre los procedimientos 36 y 38, y se denomina "conversión de tipo" en 36.

El procedimiento típico es determinar un valor de escalado apropiado para la señal de una manera ad hoc, después monitorizar los valores de los coeficientes resultantes, s, disminuyendo el valor de μ mediante una pequeña cantidad si ocurre recorte; el valor de escalado se establece rápidamente en un valor adecuado.

Además, los efectos perceptuales también necesitan considerarse. Por ejemplo, el estándar de compresión de audio MP3 es un procedimiento de compresión de audio con pérdidas donde los sonidos que son imperceptibles por el oído humano (debido la máscara perceptual de tales sonidos) se retiran del audio con poca diferencia perceptible desde el punto de vista del oyente. El estándar MP3 emplea un “modelo receptor inverso" donde las señales que el receptor (humano oído) no puede percibir se retiran mediante el códec MP3 con poca/ninguna degradación en la calidad de la señal perceptible. Tales efectos de enmascarado perceptual se exhibirán seguramente mediante el área del cuerpo a la que la agrupación táctil se aplica (p. ej., lengua) y pueden explotarse cuando se represente información en el sistema de representación electrotáctil.

Agrupación Electrotáctil

La agrupación de electrodo empleada en el dispositivo de la Fig. 2 usa estimuladores electrocutáneos hemisféricos que aseguran una densidad de corriente de la interfaz electrodo-piel homogénea. La electrostática dicta que la distribución de la carga sobre la superficie de un cuerpo cargado es mayor en las áreas de mayor curvatura de superficie. Esto se conoce comúnmente como el "efecto de borde". Los cuerpos cargados con bordes afilados experimentan un aumento de carga a lo largo de esos bordes. La mayoría de estimuladores electrocutáneos usan electrodos planos similares a discos. Las inspecciones microscópicas de estos electrodos revelan bordes afilados en el perímetro del electrodo. Se ha mostrado que estos bordes experimentan una distribución de corriente desigual en la interfaz de electrodo-piel durante la estimulación electrocutánea (Krasteva y Papazov, 2002). Esto afectará a la percepción cualitativa del estímulo y puede incluso causar dolor o quemazón de piel.

La Ley de Gauss para la resistencia del campo fuera de una esfera de radio R es

Ecuación 4

$$E = Q/_{4\pi\epsilon R^2}$$

y la carga de densidad D (carga por unidad de área superficial) en una esfera de radio de R para una carga Q se escala de manera similar:

Ecuación 5

$$D = Q/_{4\pi R^2}$$

En esta configuración, estas ecuaciones significan que la resistencia de campo y la densidad de carga D son inversamente proporcionales al radio de electrodo R. Asumiendo la carga constante Q, esto implica que la resistencia de campo y la densidad de carga serán mayores en el punto de un alfiler que sobre la superficie de una esfera grande.

Esto implica, que para un tamaño de electrodo determinado, si se desea minimizar la resistencia de campo máxima, el electrodo debería ser hemisférico.

La densidad de corriente se proporciona por la ecuación:

Ecuación 6

$$J(r,t) = qn(r,t)v_d(r,t)$$

donde J(r, t) es el vector de densidad de corriente en la ubicación r en el momento t (amperios de unidad SI por metro cuadrado). n (r,t) es la densidad de partículas en un recuento por volumen en la ubicación r en el momento t (unidad SI m-3) es la carga de las partículas individuales con densidad n (unidad SI: Culombios). Se emplea un estimulador electrocutáneo hemisférico con un radio uniforme y una curvatura de superficie que asegurarán una densidad de corriente homogénea en la interfaz de electrodo-piel, reduciendo así el riesgo de concentraciones de corriente dañinas.

Una agrupación de electrodo hexagonal distribuida de manera uniforme consiste en agrupaciones hexagonales concéntricas con electrodos distribuidos uniformemente. El número de electrodos e se proporciona mediante la siguiente ecuación:

$$e = 1 + \sum_{n=1}^{k} 6n = 1 + 3k(k+1) = 3k^n + 3k + 1$$

donde k es el número de agrupaciones hexagonales concéntricas en la agrupación alrededor del electrodo central. La ventaja de esta agrupación es que la separación entre electrodos es uniforme por toda agrupación.

Ensayos piloto

Se inscribió a 20 participantes para tomar parte en un ensayo de cuatro semanas de un dispositivo de tratamiento que presentó simultáneamente sonido en las modalidades de escucha y tacto. La música que se reprodujo al usuario a través de auriculares se descompuso simultáneamente en ondas constituyentes (usando el procedimiento de transformada STFT antes descrito) que se codificaron en patrones táctiles y se presentaron al usuario a través de un sensor electrotáctil intraoral colocado en la lengua.

El tratamiento se proporcionó a 12 varones y 8 mujeres con una edad promedia de 48 ± 22 años con acúfenos permanentes (síntomas persistentes >6 meses) debido a pérdida de escucha relacionada con ruido y/o la edad. Los participantes no recibieron ningún otro tratamiento para su pérdida de escucha o acúfenos.

El régimen del tratamiento consistía en el uso del dispositivo durante 30 minutos por la mañana y de nuevo por la tarde. En cada sesión de tratamiento, los participantes escucharon 30 minutos de música prescrita en los auriculares, mientras simultáneamente sentían las representaciones táctiles de la música en la lengua.

Los participantes fueron evaluados antes y después de la intervención usando el *Tinnitus Handicap Inventory* (THI) [véase C. W. Newman *et al.*, *Arch Otolaryngol Head Neck Surg*., volumen 122, páginas 143-148, 1996; y A. McCombe *et al.*, *Clin Otolaryngol*, volumen 26, páginas 388-393, 1999], y usando el *Tinnitus Reaction Questionnaire* (TRQ) [véase P. H. Wilson *el al.*, *Journal of Speech and Hearing Research*, volumen 34, páginas 197-201, 1991.]. El *Tinnitus Handicap Inventory* es una medida de discapacidad de acúfenos de autoinforme que cuantifica el impacto de 25 aspectos de acúfenos en la vida diaria. El THI categoriza las puntuaciones de inventario en 5 grados de severidad: Grado 1: Ligero, 2: Medio, 3: Moderado, 4: Severo y 5: Catastrófico. La medida THI de preintervención se usó para evaluar el impacto de acúfenos en el participante en el periodo de cuatro semanas antes del comienzo del estudio. La medida después de la intervención se usó para evaluar el impacto de los acúfenos en el participante en el periodo de cuatro semanas desde que el participante recibió el tratamiento.

El *Tinnitus Reaction Questionnaire* es una medida de reacción de acúfenos de autoinforme que evalúa 26 aspectos de acúfenos en la calidad de vida. Una puntuación TRQ de 16 o más se considera clínicamente significativa. La medida TRQ de antes de la intervención se usó para evaluar el impacto de acúfenos en el participante en el periodo de una semana antes del comienzo del estudio. La medida después de la intervención se usó para evaluar el impacto de acúfenos en el participante en el periodo de una semana después de completar del estudio.

Además del THI y el TRQ, a los participantes se les pidió que describieran cualquier cambio sintomático y que dijeran si los síntomas habían Desaparecido/ Mejorado mucho/ Mejorado/ No habían cambiado/ Estaban peor/ Mucho peor.

Conformidad del participante: de los 20 participantes reclutados para el estudio, 17 completaron con éxito las cuatro semanas de tratamiento.

Más del 60 % de los participantes que completaron el tratamiento de cuatro semanas informaron de que sus síntomas habían “Mejorado” o “Mejorado mucho”.

Casi el 60 % de los participantes registraron una reducción de un grado o más en sus puntuaciones THI.

Casi el 90 % de los participantes registraron una mejora en la puntuación TRQ, con el 65 % registrando mejoras mayores del 20 %. El 30 % de los participantes pasó de puntuaciones TRQ clínicamente significativas entre (>16) a puntuaciones TRQ clínicamente no significativas (<16).

Posibles descomposiciones de señal alternativas

Ya que la premisa para esta etapa de descomposición de señal se basa en la noción general de explotar la redundancia en entradas sensoriales, la descomposición de señal no se limita al análisis de armónicos debido al hecho de que muchas suposiciones diferentes pueden usarse para lograr la Ecuación 1. En las siguientes secciones se describen las posibles alternativas para esta etapa.

Aunque el procedimiento de análisis de armónicos es extremadamente útil, es posible construir representaciones más parsimoniosas usando características (que se llaman diccionario de señales o vectores de base) que se aprenden a

partir de un corpus de datos. Tales procedimientos parsimoniosos o escasos crean representaciones con menos coeficientes activos y producen mejores imágenes que pueden representarse en la agrupación. Las representaciones parsimoniosas son comunes en el procesamiento/análisis de señal y se usan típicamente para hacer que una señal sea compresible, por ejemplo, las Transformaciones de Coseno Discretas se usan en compresión de vídeo MPEG.

La etapa de aprendizaje se realiza fuera de línea, es decir a priori, mientras que la descomposición del flujo de audio de entrada (encaje) con el diccionario descubierto se realiza en tiempo real. A continuación, se señala un número de tales enfoques que pueden usarse para construir un diccionario de señales que se sintoniza con un corpus de sonido, por ejemplo, diálogo.

Cuando las señales de diccionario se aprenden, es necesario aplicar el procedimiento elegido en algunos datos de entrenamiento (p. ej., diálogo, música, audio en entornos diferentes, etc.) fuera de línea, donde los datos del entrenamiento se muestrean y se pasan por tramas usando el mismo esquema descrito anteriormente, sin embargo, una secuencia de K tramas se consideraron a la vez y se usa X = [x1|...|xK] para indicar la matriz de datos de entrenamiento. El diccionario de señal resultante, W, produce codificaciones, s, que se optimizan para los datos de entrenamiento, que pueden usarse para producir codificaciones óptimas para estos tipos de datos, siempre que K sea suficientemente grande. Por ejemplo, cuando se escucha música en un evento exterior, es útil usar un diccionario de señales que se aprende a partir de registros de eventos de música exterior (en oposición a por ejemplo música de ballena) para lograr una codificación parsimoniosa.

En la Ecuación 1, W se construye a partir de este diccionario de señal precalculado y cuando el sonido se presenta en el sistema, se codifica usando esta matriz de diccionario. Los coeficientes de las codificaciones resultantes, s1, ..., sN, se organizan topológicamente dependiendo del mismo criterio (p. ej., dependencias mutuas entre s1, ... , sN) y se representan en el sistema de representación electrotáctil. Tal organización es similar a la organización tonotópica de receptores auditivos en el córtex auditivo. De esta manera, el sistema realiza la codificación perceptual de modalidad disfuncional.

Los siguientes procedimientos aprenden W a partir de un corpus de datos, X, de una manera fuera de línea, x se proyectan posteriormente sobre el diccionario de señales fijo W mediante una transformada lineal (Ecuación 1) en tiempo real cuando el sistema es operativo.

Análisis de componente principal

El Análisis de Componente Principal (PCA) (Pearson, 1901) (también conocido como la transformada de Karhunen-Lo'eve o la transformada de Hotelling) es una técnica para la reducción de dimensionalidad de datos multivariantes, X es una multivariada, que retiene las características de los datos que contribuyen en gran medida a su varianza estadística. PCA es una transformación lineal que no tiene un conjunto fijo de vectores básicos (a diferencia de la Transformada de Fourier, por ejemplo). En su lugar, el PCA transforma los datos de entrenamiento X en un sistema de coordenadas ortogonal que se corresponde con las direcciones de la varianza de datos. Los vectores que definen las direcciones de varianza, [w1|...| wN], se conocen como los componentes principales de los datos:

Ecuación 7

$$\Sigma_x = W\Lambda W^{-1}$$

Donde $\Sigma_X = \langle \mathbf{X}\mathbf{X}^T \rangle$ es la matriz de covarianza de **X** y la entrada a la fase de aprendizaje. Después del aprendizaje, es decir, Σx se ha diagonalizado en la forma de la Ecuación 7, W contiene los eigen vectores (componentes principales) de Σ_X y la matriz diagonal Λ contiene sus eigen valores asociados λi...λN. Durante el tiempo de ejecución, cuando una trama **x**, desde la misma clase de datos de audio, se proyecta en **W** (Ecuación 1), entonces las variantes de s, s1, ...sN, son (aproximadamente) descorrelacionadas, es decir, la matriz de correlación para s es una matriz diagonal. De esta manera, los componentes descorrelacionados del sonido pueden representarse en el dispositivo.

Análisis de componente independiente

El Análisis de Componente Independiente (ICA) (Comon, 1994) abarca un intervalo de procedimientos para la separación de datos multivariantes en componentes independientes estadísticamente. Inspirado por la hipótesis de Barlow, (Atick y Redlich, 1990) se postuló el principio de redundancia mínima, que propone un modelo que utiliza tal procedimiento como el mecanismo para lograr un código eficiente. El ICA proporciona un operador de matriz lineal, W, (que se aprende a partir de datos de entrenamiento X) que factoriza la distribución de probabilidad conjunta de s en componentes independientes,

Ecuación 8

$$P(s) = P(s_1,\ldots,s_N) = \prod_{i=1}^{N} P(s_i)$$

y se soluciona generalmente como un problema de optimización, donde W se descubre maximizando alguna medida

de independencia. Tales medidas incluyen información mutua (Comon, 1994), entropía (Bell y Sejnowski, 1995), no gaussianidad (Hyvarinen y Oja, 1997) y escasez (Zibulevsky y Pearlmutter, 2001). Al usar un diccionario de señales, W, construido por ICA en la Ecuación 1, es posible representar los componentes estadísticamente independientes del sonido, lo que produce codificaciones que son tanto descorrelacionadas como mutuamente independientes.

Factorización de matriz no-negativa

(NMF) es una factorización aproximativa de bajo rango no negativa lineal para la descomposición de datos multivariantes (Lee y Seung, 2001; Paatero y Tapper, 1994). NMF es un enfoque basado en porciones que no realizan ninguna suposición estadística sobre los datos. En su lugar, se asume que para el dominio en cuestión, por ejemplo, imágenes en escala de grises, los números negativos son físicamente insignificantes. Los componentes negativos no tienen representación en el mundo real en un contexto de imagen de escala de grises, lo que conduce a una limitación de que la búsqueda para W debería confinarse a valores no negativos, es decir, valores mayores que e incluyendo cero. Los datos que contienen componentes negativos, por ejemplo, sonido, deben transformarse a un formato no negativo antes de que pueda aplicarse NMF. Normalmente, el espectrograma de magnitud se usa para este fin, donde los datos de entrenamiento X también sufren el procedimiento de espectrograma de magnitud como se ha señalado antes. Normalmente, NMF puede interpretarse como

Ecuación 9

$$\min_{\mathbf{A},\mathbf{S}} \frac{1}{2}\|\mathbf{X}-\mathbf{AS}\|^2, \quad \mathbf{A},\mathbf{X},\mathbf{S} \geq 0$$

donde $A = W^{-1}$ es una matriz N x R con $R \leq N$, de manera que el error de reconstrucción se minimiza. Los factores A y S se aprenden a partir de los datos de entrenamiento X usando reglas de actualización multiplicativas (Lee y Seung, 2001), donde A contiene las características de los datos y S sus activaciones, que se descartan después de la etapa de aprendizaje. Las codificaciones NMF no son negativas y de esta manera están fácilmente disponibles para la representación, mientras que las codificaciones producidas por PCA e ICA pueden incluir componentes negativos, lo que requiere una transformada no lineal adicional (p. ej., valor absoluto) antes de que puedan representarse.

Durante el tiempo de ejecución, cada trama x tiene como resultado s_n al realizar la siguiente optimización,

Ecuación 10

$$\min_{\mathbf{A},\mathbf{s}} \frac{1}{2}\|\mathbf{x}-\mathbf{As}_n\|^2, \quad \mathbf{A},\mathbf{x},\mathbf{s}_n \geq 0$$

Descomposición sobrecompleta escasa

En la Ecuación 1, el diccionario en el que el sonido de entrada x se descompone, es decir, las columnas de W-1, tiene un tamaño igual a (o menor que) la dimensionalidad N de x. Tal como se analiza, puede ser aconsejable que la descomposición de x sea escasa, lo que significa que se exprese en términos de un pequeño número de elementos de diccionario, lo que se corresponde en este caso con el vector de s, que es escaso. Si la distribución desde la que x se extrae es suficientemente rica, esto puede ser imposible con cualquier diccionario solo con N elementos.

Como alternativa, se puede usar un diccionario sobrecompleto, lo que significa que tiene más de N elementos, es decir, el diccionario de señal es una matriz gorda. Si el diccionario se coloca en las columnas de una matriz D, entonces esto se corresponde con encontrar un vector s que es escaso, y que también satisface la ecuación menos limitada $Ds \approx x$.

Hasta este punto se ha descrito, usando la Ecuación 1, la transformación de x usando el diccionario de señales. Ahora se especifica un diccionario de señales sobrecompleto D, donde la transformación usando la Ecuación 1 no es posible ya que la Ecuación está menos limitada para D, y por tanto, se usa la Ecuación de optimización 11 a continuación para lograr la transformación.

Existen una variedad de algoritmos para encontrar tal vector s, el más prominente de los cuales es la descomposición L1, en la que una s se encuentra, cuyos elementos con una suma mínima de valores absolutos se somete a la anterior condición donde ≈ se hace concreta como error al cuadrado, teniendo como resultado:

Ecuación 11

$$\min_{\mathbf{s}} \sum_{i=0}^{M} |s_i| + \lambda \|\mathbf{Ds}-\mathbf{x}\|_2^2$$

donde λ es una constante que compensa la escasez de s contra la fidelidad de la representación de x.

Justo como en el caso para W en la Ecuación 1, el diccionario D puede encontrarse mediante una variedad de medios, incluyendo PCA, ICA, NMF y variantes de los mismos, tal como un ICA sobrecompleto y una NMF convolutiva escasa (O'Grady y Pearlmutter, 2008).

Además, es posible combinar diccionarios de señales que se entrenan sobre diferentes datos y construyen un diccionario sobrecompleto. Por ejemplo, usando ICA se pueden construir dos diccionarios de señales con las mismas dimensiones que W para el diálogo, donde uno se sintoniza con el diálogo varón, que se denomina M, y el otro se sintoniza con el diálogo de mujer, que se denomina F. Se pueden combinar entonces ambos para construir D, donde D = [M|F]. La ventaja de esto es que cuando el diálogo varón se captura en el flujo de audio, la porción M de D es la más activa aportando representaciones escasas y viceversa.

Como se mencionó antes, los electrodos de la representación electrotáctil representan intensidad y por tanto los coeficientes a representar en el dispositivo, s_n, no deben ser negativos. Para los procedimientos alternativos antes descritos, con la excepción de NMF y el espectrograma de magnitud, es necesario pasar los coeficientes resultantes, s, a través de una no linealidad que produce solo valores no negativos en la salida. Se usa el valor absoluto de s para generar s_n,

Ecuación 12

$$S_n = abs(s)$$

Para representar elementos de diccionario que describen la señal de audio en una agrupación en dos dimensiones, los elementos de diccionario usados para descomponer la señal, que son de longitud N, pueden pensarse como puntos en el espacio N, con diferentes distancias entre cada vector. Es posible incrustar estos vectores multidimensionales en un espacio en dos dimensiones, y por tanto lograr una disposición en dos dimensiones, usando procedimientos tales como la Incrustación Lineal Localmente o el Escalado multidimensional. Estos procedimientos requieren una medida de distancia (o divergencia), que puede determinarse ya sea directamente desde los elementos de diccionario o usando estadísticas de sus propiedades de respuesta. Esto significaría explotar las estadísticas de estos vectores, disponiendo así vectores que tienen estadísticas similares tal como correlación cruzada, entropía relativa, dependencia mutua, etc., en proximidad cercana.

REIVINDICACIONES

1. Un aparato configurado para usarse en el tratamiento de acúfenos, que comprende una unidad de procesamiento de sonido, una unidad táctil y una interfaz entremedias, en el que:

dicha unidad táctil comprende una agrupación de estimuladores (16) electrocutáneos, cada uno de los cuales puede accionarse independientemente para aplicar un estímulo táctil a un sujeto y una entrada para recibir una pluralidad de señales de accionamiento desde dicha interfaz y que dirige señales de accionamiento individuales a estimuladores individuales; y
dicha unidad de procesamiento de sonido comprende:

una entrada configurada para recibir una señal de audio;
un procesador (14) de señal digital operable para analizar dicha señal de audio y generar dicha pluralidad de señales de accionamiento representativas de dicha señal de audio dividiendo dicha señal de audio en una serie de tramas en el dominio de tiempo, realizando una transformada en cada trama para generar un conjunto de coeficientes que representan dicha trama y mapeando dicho conjunto de coeficientes en un conjunto de señales de accionamiento a aplicar a la agrupación; y
una salida configurada para recibir dicha pluralidad de señales de accionamiento desde dicho procesador de señal digital y para proporcionar dicha pluralidad de señales de accionamiento a dicha interfaz, el aparato configurado además para enviar la señal de audio, enviándose dicha señal de audio de forma simultánea con la aplicación de las señales de accionamiento a la agrupación.

2. Un aparato de acuerdo con la reivindicación 1, en el que dicho procesador (14) de señal es operable además para generar dicha pluralidad de señales de accionamiento como una secuencia variable en el tiempo de patrones de agrupación de salida, en el que cada patrón de agrupación de salida comprende un conjunto de señales de accionamiento a aplicar en la agrupación durante un periodo de tiempo discreto, representativo de una muestra de tiempo discreta de la señal de entrada.

3. Un aparato de acuerdo con la reivindicación 1 o 2, en el que dicha transformada realizada en cada trama se selecciona a partir de una transformada de fourier, una transformada (42) de fourier de tiempo corto (STFT), una transformada de ondículas, una transformada curvelet, una transformada gammatone y una transformada zak.

4. Un aparato de acuerdo con la reivindicación 3, en el que dicha transformada es una transformada de fourier o transformada de fourier de tiempo corto, y en el que dicha señal se muestrea en un índice de muestreo de entre 4 kHz y 12 kHz, más preferentemente entre 6 kHz y 10 KHz, y más preferentemente aproximadamente 8 kHz.

5. Un aparato de acuerdo con cualquiera de las reivindicaciones 1 a 4, en el que dichas series variables en el tiempo de tramas se superponen entre sí.

6. Un aparato de acuerdo con cualquiera de las reivindicaciones 1 a 5, en el que el inicio de cada trama se desvía del inicio de la trama precedente por entre 10 y 20 ms, más preferentemente entre 12-18 ms, y más preferentemente aproximadamente 16 ms.

7. Un aparato de acuerdo con cualquiera de las reivindicaciones 1 a 6, en el que dicho procesador se programa para emplear una longitud de trama desde 18 a 164 ms, más preferentemente desde 50 a 150 ms, y más preferentemente 64 o 128 ms.

8. Un aparato de acuerdo con cualquiera de las reivindicaciones 1 a 7, en el que dicho conjunto de coeficientes representa la señal en el dominio de frecuencia, y en el que los coeficientes se mapean a las señales de accionamiento de manera que los coeficientes que representan frecuencias similares se mapean a señales de accionamiento dirigidas a estimuladores que están físicamente cerca entre sí en dicha agrupación.

9. Un aparato de acuerdo con la reivindicación 8, en el que los coeficientes que representan frecuencias cercanas se mapean a señales de accionamiento dirigidas a estimuladores que son físicamente adyacentes entre sí.

10. Un aparato de acuerdo con la reivindicación 1 o la reivindicación 2, en el que dicho procesador (14) de señal se programa para analizar dicha señal de audio mapeando segmentos sucesivos de dicha señal de audio a un conjunto de características seleccionadas desde un diccionario de dichas características.

11. Un aparato de acuerdo con cualquier reivindicación anterior, en el que dicha agrupación de estimuladores es una disposición rectangular de m x n estimuladores separados regularmente o en el que dicha agrupación de estimuladores es una disposición hexagonal de subagrupaciones hexagonales concéntricas o una disposición circular de subagrupaciones circulares concéntricas.

12. Un aparato de acuerdo con cualquier reivindicación anterior, en el que dicho procesador (14) es operable además para normalizar las magnitudes de las señales de accionamiento para que entren dentro de un intervalo predeterminado de intensidades de señal de accionamiento.

13. Un aparato de acuerdo con cualquier reivindicación anterior, en el que dicha unidad táctil tiene la forma de un cuerpo dimensionado para colocarse en la lengua de un sujeto humano, y en el que cada estimulador tiene la forma de un electrodo que tiene una superficie redondeada que se proyecta desde dicho cuerpo.

14. El aparato de acuerdo con cualquier reivindicación anterior en el que el medio de salida de audio comprende auriculares.

15. Un procedimiento para convertir una señal de audio en una pluralidad de señales de accionamiento para la posterior transmisión a una agrupación de estimuladores electrocutáneos, **caracterizado porque** el procedimiento comprende las etapas de

recibir una señal de audio;
analizar dicha señal de audio y generar dicha pluralidad de señales de accionamiento representativas de dicha señal de audio como una secuencia variable en el tiempo de patrones de agrupación de salida dividiendo dicha señal de audio en una serie de tramas en el dominio de tiempo, realizando una transformada en cada trama para generar un conjunto de coeficientes que representan dicha trama, y mapeando dicho conjunto de coeficientes en un conjunto de señales de accionamiento a aplicar a la agrupación de manera que las señales de accionamiento individuales se dirijan a estimuladores individuales; y

enviar dicha pluralidad de señales de accionamiento para la transmisión a dicha agrupación de estimuladores electrocutáneos y enviar la señal de audio, enviándose dicha señal de audio de forma simultánea con la transmisión de las señales de accionamiento a la agrupación.

Fig. 1

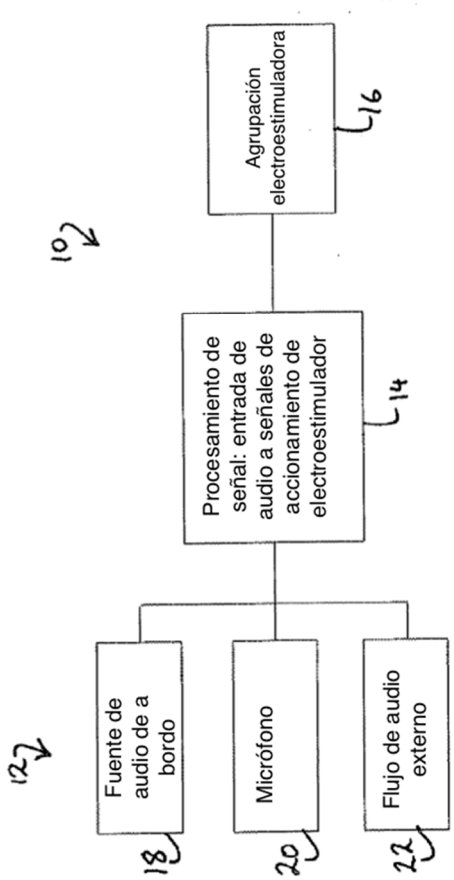
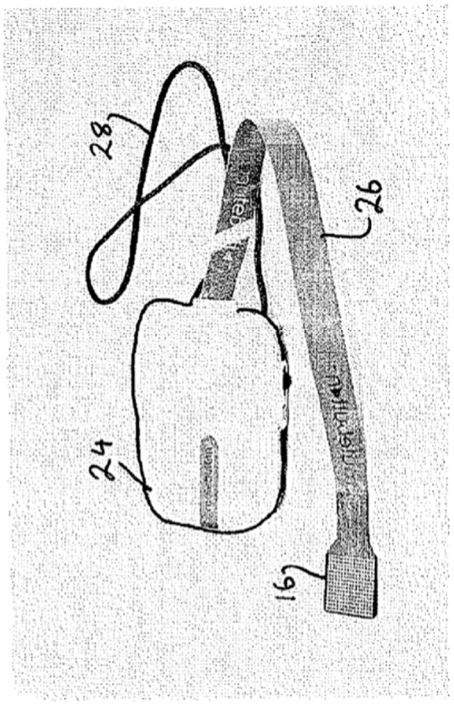

10
12
Agrupación electroestimuladora
16
Procesamiento de señal: entrada de audio a señales de accionamiento de electroestimulador
14
Fuente de audio de a bordo
18
Micrófono
20
Flujo de audio externo
22

Fig. 2


28
26
24
16

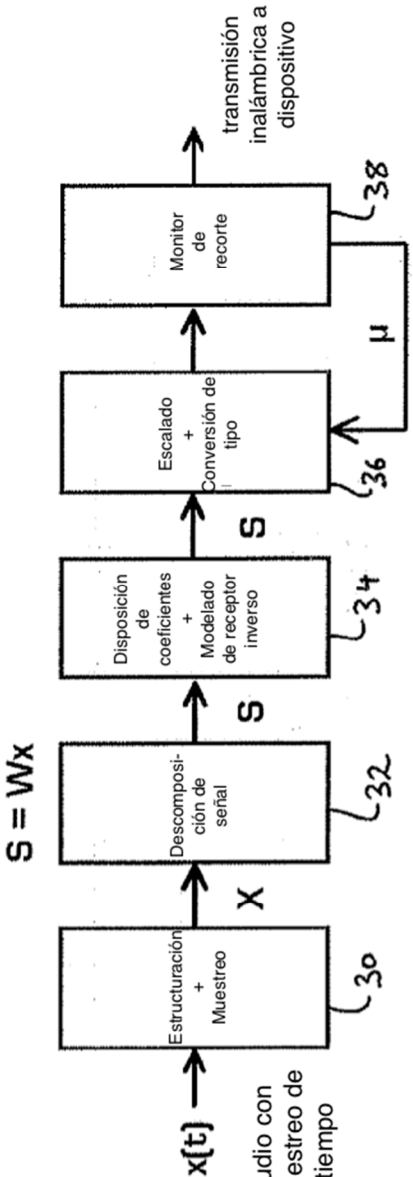

x[t]
audio con muestreo de tiempo
Estructuración + Muestreo
30
X
Descomposición de señal
32
S = Wx
S
Disposición de coeficientes + Modelado de receptor inverso
34
S
Escalado + Conversión de tipo
36
μ
Monitor de recorte
38
transmisión inalámbrica a dispositivo

Fig. 3

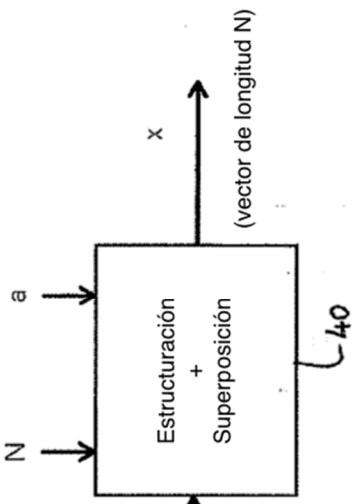

audio con muestreo de tiempo
x[t]
fs = 8kHz
N
a
Estructuración + Superposición
40
x
(vector de longitud N)

Fig. 4

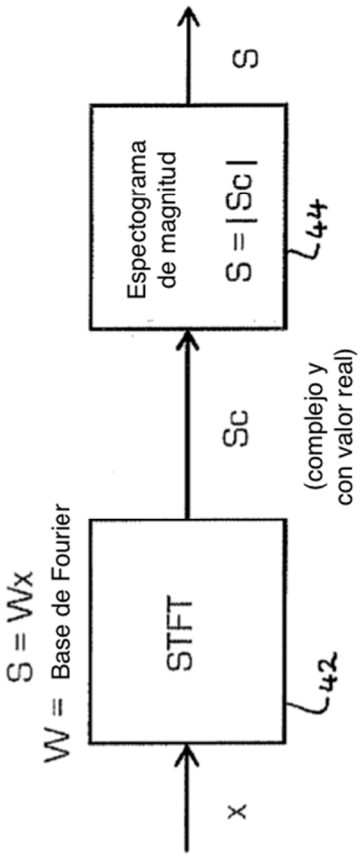

S = Wx
W = Base de Fourier
STFT
42
x
Sc
(complejo y con valor real)
Espectograma de magnitud
S = |Sc|
44
S

Fig. 5

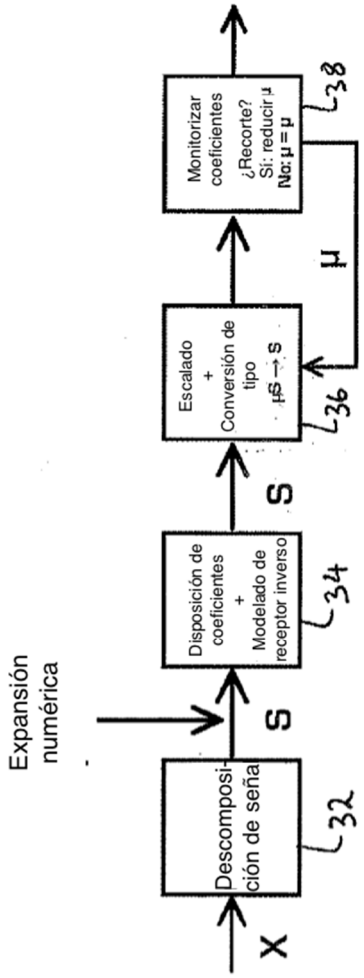

Expansión numérica
Descomposición de señal
32
S
Disposición de coeficientes + Modelado de receptor inverso
34
S
Escalado + Conversión de tipo
36
Monitorizar coeficientes
¿Recorte?
Si: reducir μ
No: μ = μ
38
μ
X

Fig. 6