

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 628 521**

51 Int. Cl.:

G10L 21/0364 (2013.01)

G10L 25/03 (2013.01)

G10L 25/18 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **01.02.2010 PCT/AT2010/000032**

87 Fecha y número de publicación internacional: **12.08.2010 WO10088709**

96 Fecha de presentación y número de la solicitud europea: **01.02.2010 E 10708882 (5)**

97 Fecha y número de publicación de la concesión europea: **22.03.2017 EP 2394271**

54 Título: **Método para la separación de recorridos de señal y uso para la mejora del habla con laringe electrónica**

30 Prioridad:

04.02.2009 AT 1932009

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

03.08.2017

73 Titular/es:

**HEIMOMED HEINZE GMBH & CO. KG (100.0%)
Daimlerstrasse 30
50170 Kerpen, DE**

72 Inventor/es:

**HAGMÜLLER, MARTIN y
KUBIN, GERNOT**

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 628 521 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

Método para la separación de recorridos de señal y uso para la mejora del habla con laringe electrónica

5 En el caso de la invención se trata de un procedimiento para la mejora de la calidad del habla de un hablante mediante laringe electrónica (LE), digitalizándose la señal de habla del hablante a través de medios adecuados. Con medios adecuados se entienden en este caso por ejemplo, un micrófono con convertidor analógico/digital correspondiente, un teléfono u otros medios mediante el uso de equipo electrónico.

10 En el caso de una LE se trata de un dispositivo para la conformación de una voz de sustitución artificial, por ejemplo, para pacientes, a los cuales se les ha retirado operativamente la laringe. La laringe electrónica se coloca en este caso en el lado inferior de la mandíbula; un generador de sonidos con una determinada frecuencia pone en vibración el aire de la cavidad bucal a través de las partes blandas en el lado inferior de la mandíbula. Estas vibraciones se modulan entonces mediante los órganos de articulación, de manera que se posibilita el habla. Dado que no
15 obstante, el generador de sonidos habitualmente solo trabaja con una frecuencia, la voz suena monótona y no natural, o "tipo robot".

Es desventajoso además de ello, que la vibración de la LE perturba la percepción del habla o incluso la supera, dado que solo se articula una parte del sonido en la cavidad bucal. Las partes que salen directamente del aparato o en el
20 punto de paso en la garganta solapan las partes articuladas y reducen la inteligibilidad. Este es el caso en particular en el caso de hablantes, los cuales fueron sometidos a una terapia de radiación en la zona de la garganta, debido a lo cual se rigidiza la estructura del tejido. Se han desarrollado por lo tanto diferentes métodos, los cuales han de reforzar la señal útil –es decir, las vibraciones articuladas- frente a la señal perturbadora – es decir, el sonido directo, o la vibración no modulada de la LE.

25 Estos métodos se usan en este caso sobre todo en situaciones, en las cuales el oyente no está expuesto directamente al sonido emitido, sino que se usan medios electrónicos, por ejemplo, al hablar por teléfono, en el caso de grabaciones de sonido o en general al hablar mediante micrófono y amplificador.

30 En el documento US 6,359,988 B1 se somete una señal de voz de LE a un análisis cepstrum y se solapa con el habla de un hablante normal, debido a lo cual, puede configurarse la modificación de la tésitura de manera más natural de aquel que habla con LE; al mismo tiempo se impide de esta manera también la parte del sonido directo emitido en la señal. Es desventajoso en esta solución sobre todo, que en caso de cada enunciado de un hablante de LE se requiere al mismo tiempo el mismo enunciado de un hablante sano (es decir, hablante sin LE), lo cual
35 prácticamente apenas es realizable.

Otra solución la muestra el documento US 6,975,984 B2, en el cual se describe una solución para la mejora de una señal de habla de LE en la telefonía. En este caso se procesa de tal manera en un procesador de señal digital la señal de habla, que el ruido de base zumbante de la LE se reconoce y se elimina de la señal de habla. La señal de
40 habla se distribuye en lugar de ello en un componente con sonido y en uno sin sonido, y se procesan por separado. La parte con sonido se transforma mediante transformación de Fourier por bloques, se filtra en frecuencia (la frecuencia de base y armónica se continúan utilizando), se retransforma y como consecuencia se substraer de la totalidad de la señal original. Como resultado queda la parte sin sonido de la señal original. Alternativamente se propone también, filtrar la parte sonora mediante paso de baja frecuencia, filtrar completamente en caso del
45 reconocimiento de una pausa de habla y solapar posteriormente la parte sin sonido.

El documento "*Enhancement of Electrolaryngeal Speech by Adaptive Filtering*" de Carol Y. Espy-Wilson et al. (JSLHR, 41: 1253-1264, 1998) describe un método para la mejora de la calidad del habla de un hablante mediante LE. El ruido de base de la LE se iguala en este caso mediante filtrado adaptativo a la señal de habla perturbada
50 mediante el ruido de base de la LE (o el ruido de base de la LE articulado dando lugar a habla); en otro paso, se separan las señales una de otra. Como resultado queda una señal de error, que se usa para el control y la adaptación de los parámetros de filtro con el objetivo de la minimización de la señal de error. La señal de error es en el presente método la señal de habla liberada del ruido de base de la LE. Se presupone en este caso, que si bien la señal de perturbación en la señal de habla está correlacionada con el ruido de base de la LE, la señal de habla que
55 interesa es sin embargo independiente de las otras señales, que por lo tanto, en cierto modo, el ruido de base perturbador y la señal de habla proceden de diferentes fuentes.

El documento "*Enhancement of Electrolarynx Speech Based on Auditory Masking*" de Hanjun Liu et al. (IEEE Transactions on Biomedical Engineering, 53(5): 865-874, 2006) describe un algoritmo de substracción para la mejora
60 de la señal de un hablante mediante LE, en particular en lo que se refiere al ruido del entorno.

A diferencia de otros métodos, los cuales prevén parámetros de substracción fijos, se adaptan en el caso de este algoritmo los parámetros de substracción en el dominio de frecuencia, en base a enmascaramiento auditivo. En este caso se parte de que el habla y los ruidos de fondo no están correlacionados y debido a ello se calcula el ruido de
65 fondo y puede restarse en el dominio de frecuencia de la señal.

Estas soluciones tienen en común, que se usan métodos los cuales se basan en un modelo, según los cuales el habla y la señal perturbadora (es decir, ruidos del entorno, pero también el ruido de base de la LE) son estadísticamente independientes, o no están correlacionados.

5 Debido a esta suposición, la implementación de los métodos mencionados se produce de manera muy laboriosa. Cuando se intenta suprimir el sonido directo con un filtro notch (adaptativo), se reduce debido a ello también la calidad de la señal de habla, que suena entonces como un susurro; la señal de habla y el ruido perturbador se encuentran sobre el mismo componente armónico.

10 El documento US 2005/0004604 A1 describe una solución para laringe, en la cual, un reproductor de sonidos y un micrófono se disponen directamente delante de la boca de un usuario, emitiendo el reproductor de sonidos un sonido con un volumen más reducido y la señal se graba para su procesamiento a través del micrófono. Durante el procesamiento, la señal se filtra esencialmente con un filtro de peine para reducir o para eliminar los componentes armónicos de la señal. En este caso queda afectada no obstante fuertemente la calidad de la señal de habla.

15 En el documento WO 2006/099670 A1 se describe un dispositivo para la supervisión de las vías respiratorias, introduciéndose sonido de dominio de frecuencia más alto en las vías respiratorias de un objeto y a partir de los sonidos reflejados o procesados, se determina el estado de las vías respiratorias. De esta manera es posible por ejemplo, determinar un desplazamiento de las vías respiratorias. En una variante de la invención se comprueba mediante FFT (del inglés *Fast Fourier Transform*, transformada de Fourier) la superación de determinados valores umbral, de lo cual se sacan conclusiones sobre el tratamiento de la señal medida.

20 Es una tarea de la invención superar las desventajas mencionadas más arriba del estado de la técnica y mejorar la calidad del habla de usuarios de LE al usarse medios electrónicos como por ejemplo, micrófonos.

25 Esta tarea se soluciona con un procedimiento del tipo mencionado inicialmente según la invención con los siguientes pasos:

30 a) dividir una señal de habla de un canal $S(w, t)$, la cual consiste en la suma de los componentes de una parte variable en el tiempo $x(w) H(w, t)$ y de una parte no variable en el tiempo $x(w) F(w)$, en una serie de canales de frecuencia mediante la conversión del dominio temporal a un dominio de frecuencia discreto, b) filtrar la parte no variable en el tiempo $x(w) F(w)$ mediante el filtrado de la frecuencia de modulación 0 Hz mediante un filtro de paso alto o de notch en cada canal de frecuencia, y c) retransformar la señal de habla filtrada del dominio de frecuencia al dominio temporal y reunir dando lugar a una señal de salida de un único canal.

40 La invención aprovecha un modelo mejorado del uso de una LE, debido a lo cual, el ruido de base de la LE articulado dando lugar a una señal de habla, así como las partes no modificadas de la LE, que perturban la percepción de la señal de habla, provienen de una misma fuente, en concreto de la LE. Dado que el ruido de base no articulado perturbador de la LE puede reconocerse en el dominio de modulación como señal no variable en el tiempo, puede filtrarse fácilmente mediante un procedimiento adecuado. No se produce por lo tanto una separación de fuentes de señal, sino de recorridos de propagación (de un recorrido de propagación a través de los órganos de articulación de un hablante, otro recorrido de propagación del lugar de uso en la garganta del hablante directamente al oído del oyente, o al micrófono o medio de grabación).

45 El experto conoce una pluralidad de posibilidades de trasladar una señal digitalizada de un canal, al dominio de frecuencia y de dividirla así en una serie de canales de frecuencia. En cada canal de frecuencia se suprime la frecuencia de modulación de la LE mediante filtros adecuados - por ejemplo, filtros notch o de paso alto, aplicados a la cantidad - y de esta manera se mejora la calidad de las partes de señal articuladas.

50 Procedimientos similares del estado de la técnica tienen en consideración las partes articuladas, así como las partes no modificadas, como provenientes de diferentes fuentes y eligen enfoques correspondientes a este modelo, por ejemplo, filtrados mediante filtro de paso banda, que amortiguan entonces no obstante también la señal de habla.

55 El procedimiento según la invención tiene por lo tanto el objetivo de aumentar la inteligibilidad del habla de los usuarios de LE o de hacer más agradable y "humana" la señal. El objetivo es reducir o eliminar el sonido directo de la LE en caso de comunicación a través de medios electrónicos (por ejemplo, el teléfono).

60 La realización del procedimiento según la invención puede producirse por ejemplo, mediante una conexión de software, como solución cableada o también como conexión analógica.

65 De la pluralidad de métodos conocidos para la conversión de una señal al dominio de frecuencia o de vuelta, la conversión se produce en el paso a) del procedimiento según la invención ventajosamente mediante transformación de Fourier y la retransformación en el paso c) mediante transformación de Fourier inversa. La conversión se produce por bloques (por ejemplo, bloques de 20 ms) a intervalos cortos (actualización por ejemplo, cada 10 ms). La división

de la señal en una serie de canales de frecuencia se produce durante la conversión de la señal al dominio de frecuencia.

En una variante de la invención, la conversión de la señal de habla en el paso a) y la retransformación en el paso c) se producen con un correspondiente banco de filtros.

Los resultados del procedimiento según la invención pueden continuar mejorándose cuando antes del filtrado en el paso b) se produce una compresión de señal y tras el paso b) una descompresión. Mediante la compresión puede evitarse que en el caso de amplias amplitudes sus modificaciones sean tan dominantes, que las modificaciones de pequeñas amplitudes no se tengan en cuenta. Mediante la compresión adquieren mejor visibilidad por lo tanto para el filtro las modificaciones relativas.

En otra configuración del procedimiento según la invención se produce tras la retransformación en el paso c) una rectificación de los componentes de señal negativos.

En lo sucesivo se explica la invención con mayor detalle mediante un ejemplo de realización no limitativo, que se representa en el dibujo. En este muestra esquemáticamente:

La Figura 1, una representación simplificada del uso de una LE y los recorridos de señal que aparecen,
La Figura 2, una representación simplificada de la situación en la cual se usa el método según la invención y
La Figura 3, un diagrama de bloques del método según la invención.

En la Figura 1 se muestran esquemáticamente los diferentes recorridos de transmisión de la señal de una LE 1. En este caso, hay dispuesta en la garganta de un hablante 2 una EL 1. El sonido producido por la EL 1 se expande por un lado por los canales de habla normales (boca y nariz) 5 del primer hablante 2 y se articula allí dando lugar a habla; esta primera señal 3 es claramente modificable o variable en el tiempo. Al oído de un oyente llega además de esta señal 3 variable en el tiempo, también una segunda señal 6 (representada en la Figura 1 mediante puntos y líneas) en forma de sonido directo de la EL 1, siendo esta señal 4 en su mayor medida estacionaria y asumiéndose por lo tanto, como temporalmente no variable. La segunda parte 6 de la totalidad de la señal, es decir, el ruido de base de la EL 1, es percibida por el oyente 4 como señal de perturbación y reduce la inteligibilidad del habla del oyente 2. El estímulo original mediante la EL 1 se transmite por lo tanto, a través de dos recorridos diferentes.

Bien es cierto que la invención se refiere a la mejora de la calidad del habla de un hablante mediante EL al usarse medios electrónicos – en lugar de un oyente, se grabarían por ejemplo las señales con un micrófono. Para la ilustración de la situación de partida se ha elegido no obstante por motivos de comprensión, este modelo general.

La Figura 2 muestra una representación de modelo simplificada de la situación, en la cual se usa el método según la invención para la supresión de una segunda señal 6 perturbadora (véase la Figura 1). Puede verse bien, que en el caso del método según la invención, no se da una separación de fuentes de señal, sino de recorridos de propagación.

Una señal de fuente $x(w)$ de una fuente de señal 7 se propaga por dos diferentes recorridos de señal. En el primer recorrido de señal, la señal de partida se modula mediante un filtro variable en el tiempo $H(w, t)$ dando lugar a una señal variable en el tiempo $x(w) H(w, t)$. En un segundo recorrido de señal, la señal de partida se modifica solo mediante un filtro no variante en el tiempo $F(w)$ dando lugar a una señal $x(w) F(w)$.

Las señales de los dos recorridos se suman entonces en un receptor 8 – por ejemplo, el oído de un oyente, un micrófono o similar – dando lugar a una señal $S(w, t)$ a disposición para una medición. La señal consiste entonces en la suma de los componentes $S(w, t) = x(w) H(w, t) + x(w) F(w)$.

Pueden separarse ahora las partes de señal del recorrido de señal no variable en el tiempo y del variable en el tiempo, en cuanto que o bien se amortiguan todas las partes de señal, que se modifican temporalmente, o que son temporalmente constantes. Se obtiene por lo tanto por ejemplo, como resultado solo la parte variable en el tiempo $S1(w, t) \square x(w) H(w, t)$.

Al usarse para el habla con LE, la parte de señal no articulada $x(w) F(w)$ (es decir, el ruido de base de la LE), solapa la señal de habla variable en el tiempo $x(w) H(w, t)$ y provoca debido a ello una pérdida de inteligibilidad para la señal de habla. La inteligibilidad del habla mejora en cuanto que se separa la parte de señal variable en el tiempo de la parte de señal no variable en el tiempo.

La Figura 3 muestra una posible realización del método según la invención. En este caso, puede llegar al inicio una señal de habla 9 digital cualquiera de un hablante con LE. En un primer paso 10 se transforma mediante el uso de la transformación de Fourier a corto plazo, la señal de habla 9 por bloques en el dominio de frecuencia y se divide de esta manera en una serie de canales de frecuencia. El experto puede elegir en este caso a partir de diferentes métodos establecidos para la transformación de una señal del dominio temporal al dominio de frecuencia; además de la transformación de Fourier se usa por ejemplo también, la transformada de coseno discreta. El requisito para un

uso según la invención es no obstante, que la transformación sea reversible. La señal se divide con una determinada frecuencia de actualización (por ejemplo, 10 ms) en bloques de por ejemplo, 20 ms de longitud, los cuales se alinean respectivamente en una hilera de canales de frecuencia 11. La señal de habla 9 originalmente de un canal se divide por lo tanto, en una pluralidad de dominios de frecuencia, los cuales se modifican como consecuencia del tiempo. La señal de frecuencia es compleja, pero en lo sucesivo solo se modifica el valor total, la fase 15 permanece sin cambios.

En el paso 10 puede usarse también un banco de filtros, reduciéndose la frecuencia de muestreo de la señal tras el banco de filtros. La reducción de la frecuencia de muestreo se corresponde en este caso con la configuración de bloqueo al usarse transformación de Fourier.

En otro bloque funcional 12 se filtra ahora cada canal de frecuencia 11, por ejemplo con un filtro de paso alto o notch. Este filtrado permite el filtro de determinadas frecuencias – en la técnica de sonido se eliminan perturbaciones de banda estrecha con filtros notch. Dado que la LE oscila en una determinada frecuencia – por ejemplo, 100 Hz – la señal de perturbación, la cual no está modificada por los órganos de articulación de un hablante, da como resultado en el dominio de frecuencias, amplitudes en el canal de 100 Hz con la frecuencia de modulación 0 Hz – es decir, que la amplitud de la señal de LE no cambia. La señal de perturbación se caracteriza por que es perfectamente no variable temporalmente. Para el filtro del ruido de base de la LE se usan un filtro de notch o uno de paso alto. Como frecuencia límite para el filtro de paso alto sirve en este caso la frecuencia de modulación de la LE; el filtro de notch se elige de tal manera, que bloquea exactamente en la frecuencia de modulación de la LE.

En una realización real, no podrá lograrse naturalmente una no variación temporal perfecta debido a reflexiones, refracciones, ruidos de entorno y requerimientos constructivos de la LE. Dado que tampoco el filtro está limitado a solo una frecuencia, sino que cubre un determinado dominio de frecuencias – en este caso un dominio de frecuencias de modulación-, se asegura la función del método según la invención.

En un bloque de funcionamiento 13 final se produce el retorno de las señales al dominio temporal, por ejemplo, mediante transformación de Fourier inversa y la reunión de los canales de frecuencia 11 de vuelta a un canal mediante método de solapamiento y suma. El método de solapamiento y suma es en este caso un procedimiento de procesamiento de datos digital conocido por el experto. El resultado es una señal de salida de un canal, en la cual, la señal de perturbación de la LE se ha filtrado o está al menos amortiguada. La señal de salida puede entonces continuar procesándose.

Al usarse un banco de filtros en el paso 10 vuelve a aumentarse la frecuencia de muestreo de la señal tras el filtrado en el paso 12 y se continúa tratando como se ha explicado.

Básicamente, estas realizaciones solo representan los componentes más importantes del método según la invención; antes del filtrado en el bloque 12 la señal puede ser comprimida, tras el filtrado puede estar prevista una descompresión. Puede ser ventajosa también una rectificación antes de la retransformación al dominio temporal, dado que durante el procesamiento pueden resultar valores negativos no permitidos.

La invención puede usarse por ejemplo, como dispositivo adicional para hablar por teléfono. En el caso de un teléfono analógico convencional, el dispositivo se integra fácilmente en el auricular. En el caso de un teléfono con procesador de señal digital integrado, la integración de la invención es posible mediante una conexión de software. Es posible también la realización en el marco de una solución cableada, por ejemplo, también en una conexión analógica.

El método según la invención puede utilizarse también al usarse una LE, en la cual, puede conmutarse en una y otra dirección entre dos o más frecuencias para darle al habla un sonido más realista. Esto es válido tanto para saltos de frecuencia discretos, como también para modificaciones continuas de la frecuencia de base, suponiéndose que las frecuencias, entre las cuales se cambia, se encuentran dentro de una banda de frecuencia, en la cual se divide la señal de base.

La anchura del filtro de frecuencia de modulación determina en este caso, cómo de rápido puede cambiar la frecuencia. En el caso de modificaciones continuas muy lentas, la frecuencia puede modificarse con supresión en funcionamiento por la totalidad del dominio de la banda de frecuencia – lo importante no es la magnitud, sino la velocidad de la modificación. Al conectarse y desconectarse la LE, lo cual se corresponde a una modificación rápida, la supresión comienza solo tras unos milisegundos – independientemente de lo ancho que se elija el filtro de notch o de donde se encuentre la frecuencia de base del filtro de paso alto.

En este caso, las modificaciones de la frecuencia de base no deben ser sin embargo, demasiado grandes. Para asegurar el funcionamiento según la invención, los canales de frecuencia, en los cuales se divide la señal, deberían por ejemplo, ensancharse, o el filtrado mediante filtro de paso alto debería iniciarse en una frecuencia algo más alta.

REIVINDICACIONES

1. Procedimiento para la mejora de la calidad del habla de un hablante mediante laringe electrónica (LE), cuya señal de habla $S(w, t)$ se digitaliza mediante medios adecuados, **caracterizado por** los siguientes pasos:
5 a) dividir una señal de habla de un canal $S(w, t)$, la cual consiste en la suma de los componentes de una parte variable en el tiempo $x(w) H(w, t)$ y de una parte no variable en el tiempo $x(w) F(w)$, en una serie de canales de frecuencia mediante la conversión del dominio temporal a un dominio de frecuencia discreto,
10 b) filtrar la parte no variable en el tiempo $x(w) F(w)$ mediante el filtrado de la frecuencia de modulación 0 Hz mediante un filtro de paso alto o de notch en cada canal de frecuencia, y
 c) retransformar la señal de habla filtrada del dominio de frecuencia al dominio temporal y reunir dando lugar a una señal de salida de un único canal.

2. Procedimiento según la reivindicación 1, **caracterizado por que** la conversión de la señal de habla en el paso a) se produce mediante transformación de Fourier y la retransformación en el paso c) mediante transformación de Fourier inversa.

3. Procedimiento según la reivindicación 1, **caracterizado por que** la conversión de la señal de habla en el paso a) y la síntesis de los canales de frecuencia en el paso c) se produce con un banco de filtros.
20
4. Procedimiento según una de las reivindicaciones 1 a 3, **caracterizado por que** antes del filtrado en el paso b) se produce una compresión de señal y tras el paso b) se produce una descompresión.

5. Procedimiento según una de las reivindicaciones 1 a 4, **caracterizado por que** antes de la retransformación en el paso c), se produce una rectificación de los componentes de señal negativos.
25

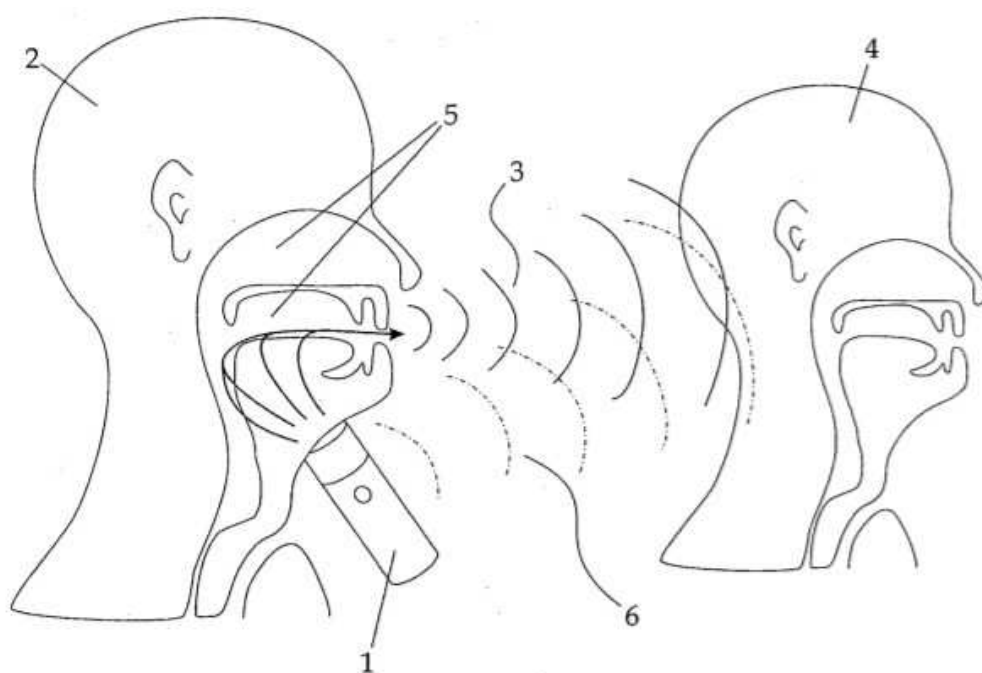


Fig. 1

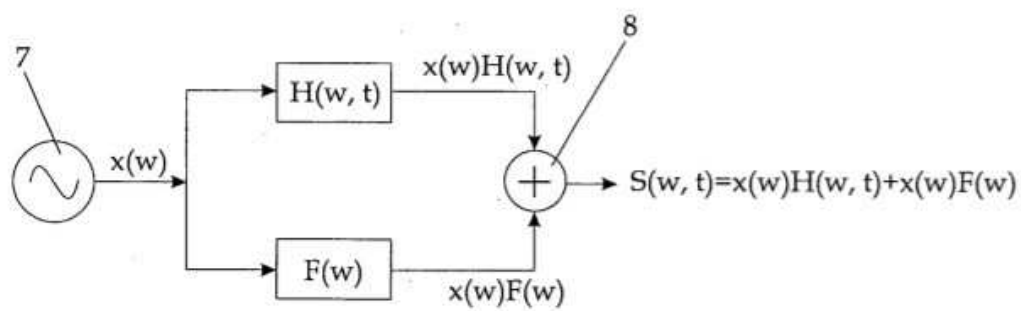


Fig. 2

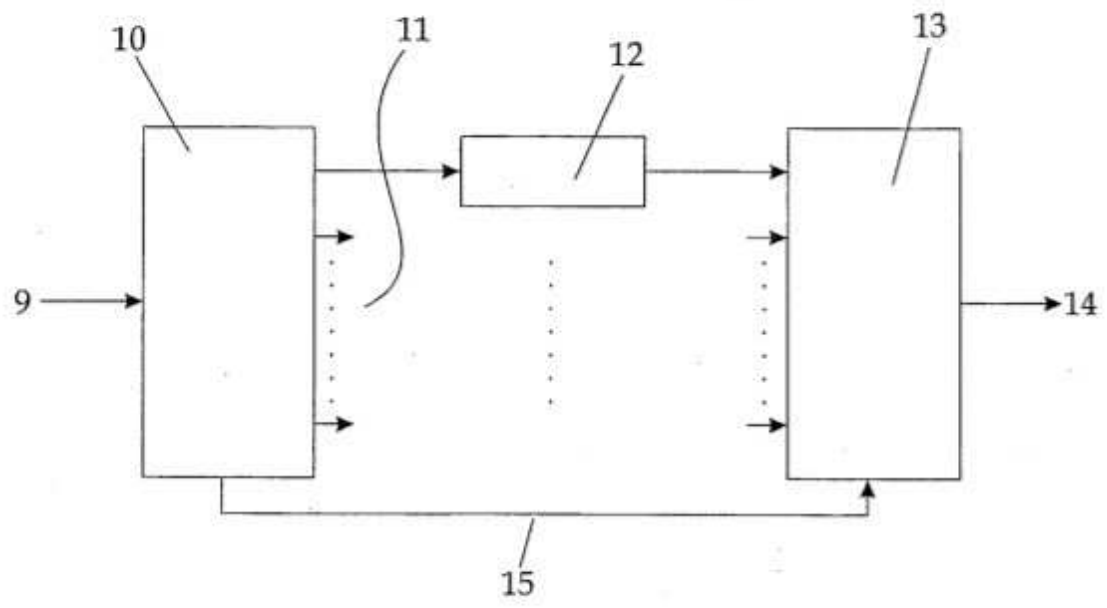


Fig. 3