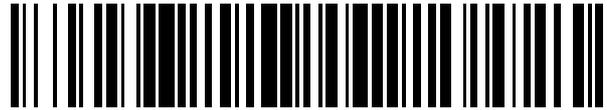


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 477 198**

51 Int. Cl.:

**G10L 25/78** (2013.01)

**G10H 1/36** (2006.01)

**G10L 25/93** (2013.01)

**G10L 25/90** (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **27.05.2010 E 10728822 (7)**

97 Fecha y número de publicación de la concesión europea: **02.04.2014 EP 2436004**

54 Título: **Procedimiento para detectar palabras cantadas y utilización de este procedimiento en un juego de Karaoke**

30 Prioridad:

**29.05.2009 FR 0953603**

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

**16.07.2014**

73 Titular/es:

**VOXLER (100.0%)  
8 passage Brûlon  
75012 Paris, FR**

72 Inventor/es:

**DELORME, NICOLAS;  
HENRY, DAMIEN y  
ZILS, AYMERIC**

74 Agente/Representante:

**DE ELZABURU MÁRQUEZ, Alberto**

ES 2 477 198 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 para detectar palabras cantadas y utilización de este procedimiento en un juego de Karaoke

5 La presente invención concierne a un procedimiento para detectar palabras cantadas en la voz. La invención tiene especialmente por objetivo proponer un procedimiento simple de poner en práctica y poco consumidor de recursos para detectar palabras en la voz.

10 La invención encuentra una aplicación particularmente ventajosa, pero no exclusiva, en aplicaciones de tipo « Karaoke ». Se recuerda que el Karaoke es un juego en el cual un jugador canta una canción conocida sobre un acompañamiento generalmente en lugar del cantante original, siguiendo generalmente las palabras sobre una pantalla. En variante, la invención podría ser utilizada igualmente en aplicaciones interactivas vocales, por ejemplo en cualquier juego de vídeo en cuyo seno se desee detectar si el jugador habla.

15 Juegos de vídeo de Karaoke como « SingStar » (marca registrada) evalúan únicamente la exactitud del canto de un jugador con respecto a una melodía de referencia. En consecuencia, un jugador que tararee en ritmo la melodía (sin cantar las palabras) obtendrá la misma puntuación, incluso una mejor puntuación, que un jugador que cante efectivamente las palabras. En efecto, tarareando, el jugador puede concentrarse únicamente en la exactitud de la melodía y/o la precisión rítmica, lo que es mucho más fácil que si éste tuviera que hacer el esfuerzo de colocar las palabras correctas de la canción en la melodía correcta y/o al ritmo correcto.

20 En particular, en ciertas canciones de rap, no hay melodía y el ritmo es demasiado rápido para ser evaluado de manera fiable. En este caso, la detección de las palabras en la canción es un criterio pertinente para evaluar al jugador. El documento US2007/0059670 describe un procedimiento de detección entre palabras cantadas y palabras habladas.

25 Para tener en cuenta las palabras en la puntuación del jugador, ciertos juegos recientes ensayan integrar el reconocimiento de palabras, con resultados discutibles, siendo estos mecanismos de reconocimiento de palabra muy difíciles de realizar y muy caros algorítmicamente. En efecto, estos necesitan cálculos complejos (utilización de modelos HMM) a fin de reconocer palabras completas, lo que es difícil de poner en práctica y provoca errores frecuentes así como una latencia importante.

30 La presente invención permite verificar si el jugador canta las palabras, de manera mucho más simple que el reconocimiento vocal tradicional, abordando el problema de manera original: no se busca « reconocer » las palabras cantadas por el jugador, lo que verdaderamente no tiene sentido puesto que éstas son ya conocidas (éstas son visualizadas en la pantalla), sino « verificar » si el jugador canta palabras, en lugar de por ejemplo tararear simplemente la melodía.

35 La invención parte así de la constatación de que cualquier lenguaje hablado, y con mayor motivo cantado, está caracterizado por una alternancia de sonidos variados (diferentes fonemas) denominada en este documento « alternancia fonémica ». Se entiende por fonémica lo que se refiere a los fonemas, es decir a cada uno de los sonidos que componen un lenguaje. Esta alternancia fonémica puede ser definida por ejemplo por una alternancia entre vocales y consonantes, o entre sonidos sonoros y sonidos sordos, o entre diversas vocales, o entre diversas consonantes etc...

40 Se entiende por tarareo la ausencia de alternancia fonémica. Por ejemplo, cuando se tararea, se emiten únicamente sonidos sonoros de tipo « la la la », « mmmm », « ah ah ah » caracterizados por una ausencia de alternancia entre sonidos sonoros y sonidos sordos y por tanto una ausencia de alternancia fonémica si se elige definir ésta por una alternancia de sonidos sonoros y de sonidos sordos. Al contrario, una persona que canta las palabras de una canción, alterna, salvo excepción, la emisión de sonidos sonoros y de sonidos sordos.

La invención propone distinguir la alternancia fonémica, es decir la pronunciación de palabras, con respecto a la ausencia de alternancia fonémica (tarareo).

45 Se recuerda que un sonido se dice « sonoro » si su producción va acompañada de una vibración de las cuerdas vocales, y « sordo » si no. Dado que el lenguaje hablado es un ensamblaje de vocales y de consonantes sonoras que hacen vibrar las cuerdas vocales y de consonantes sordas que no hacen vibrar las cuerdas vocales, se observa naturalmente esta alternancia entre sonidos sonoros y sordos. Esta constatación vale para las principales lenguas habladas en el mundo. En cambio, cuando se tararea, el sonido emitido corresponde a una emisión continua de sonidos sonoros de tipo « lalala » o « aaaaaa » o « mmmmmmm ».

50 En la invención se observa si, durante un período de referencia, la voz del jugador presenta variaciones de sonoridad o no. Si éste es el caso, entonces se deduce que el jugador está cantando palabras en este período de referencia; mientras que si este no es el caso, se deduce que el jugador está tarareando en este período de referencia. Se ha visto que un período de referencia de aproximadamente un segundo permitía obtener buenos resultados. Sin embargo, cualquier otro período de referencia es posible.

- 5 En una puesta en práctica, se mide la alternancia fonémica asociada al carácter sonoro y sordo de la voz. A tal efecto, se calcula un coeficiente de sonoridad de la voz que presenta valores elevados cuando el sonido de la voz es sonoro y valores bajos cuando el sonido de la voz es sordo. En un ejemplo, este coeficiente de sonoridad corresponde a la medición de la calidad de la extracción de la frecuencia fundamental de la señal de voz. Cuando este coeficiente de sonoridad es superior a un valor umbral durante todo el período de referencia entonces se deduce que el jugador está tarareando; en cambio, cuando el coeficiente de sonoridad no es superior al valor umbral durante todo el período de referencia, se deduce que el jugador está cantando.
- 10 La invención consiste así en verificar únicamente si el jugador pronuncia verdaderas palabras y no está tarareando, sin asegurarse de que las palabras correspondan efectivamente a las palabras de la canción. No es útil verificar si « las » palabras cantadas son las verdaderas palabras de la canción, sino únicamente si se « cantan » palabras. En efecto, si el tarareo es una ayuda importante en este tipo de juegos, el hecho de cantar otras palabras sobre una canción es más bien una dificultad suplementaria para el jugador.
- 15 De modo más general, la medición de sonoridad-no sonoridad es solamente un modo de medir la alternancia fonémica. Cualquier otro método que permita medir una variación, por ejemplo una variación de las consonantes pronunciadas (medición de la presencia de ciertas consonantes por otros métodos que la medición de la tasa de sonoridad) o variación de la vocales pronunciadas (en el triángulo vocálico), produciría el mismo tipo de resultado.
- 20 Así, alternativamente, si se elige caracterizar la alternancia fonémica por la alternancia de vocales diferentes, se mide una variación de timbre en el triángulo vocálico, porque un jugador que tararea no hace variar el timbre de su voz mientras que el jugador que canta palabras hace variar naturalmente el timbre de su voz. En el caso en que no se detecte variación del timbre de la voz en el triángulo vocálico en el período de referencia, se deduce que el jugador está tarareando; mientras que en el caso en que se detecte una variación del timbre de la voz en el triángulo vocálico en el período de referencia, se deduce que el jugador está cantando palabras.
- 25 Alternativamente, se separan las consonantes y o las vocales en varios grupos, por ejemplo cuatro grupos de consonantes y de vocales. Si todas las consonantes y todas las vocales pertenecen al mismo grupo, entonces se puede considerar que la persona tararea. Por el contrario, si el grupo al cual pertenecen las consonantes y o las vocales varía, la persona está diciendo palabras, es decir un texto cuyo contenido varía en términos de consonantes y o de vocales.
- 30 Así pues, la invención concierne a un procedimiento como el definido por la reivindicación independiente 1. En las reivindicaciones dependientes asociadas se enuncian puestas en práctica alternativas.
- 35 La invención se comprenderá mejor con la lectura de la descripción que sigue y del examen de las figuras que la acompañan. Estas figuras se dan solamente a título ilustrativo pero en modo alguno limitativo de la invención. Éstas muestran:
- Figura 1: una representación gráfica en función del tiempo de la amplitud de una señal de voz y de la frecuencia fundamental que ha sido extraída de ella con la ayuda de un algoritmo de detección de la frecuencia fundamental así como la señal de calidad de la extracción de la frecuencia fundamental,
- Figura 2: una representación esquemática de las etapas del procedimiento de acuerdo con la invención que permiten calcular estados instantáneos de la señal de voz;
- Figura 3: una representación esquemática de las etapas del procedimiento de acuerdo con la invención que permiten detectar si el jugador canta palabras o tararea a partir de los estados instantáneos de la señal de voz;
- 40 Figura 4: una representación gráfica de la amplitud de la señal de voz correspondiente a palabras cantadas, así como el estado activado o desactivado de la función de detección de palabras de acuerdo con la invención en el transcurso de la canción.
- Los elementos idénticos conservan la misma referencia de una figura a otra.
- 45 La Figura 1 muestra una representación esquemática de la amplitud de una señal S de voz en función del tiempo t.
- En una primera etapa 10 del procedimiento de acuerdo con la invención mostrada en la Figura 2, se mide la energía  $E_i$  instantánea y el coeficiente  $V_i$  de sonoridad representativo de la sonoridad de la voz en todos los puntos  $P_i$  de la señal S de voz analizados en los instantes de análisis  $t_i$  espaciados entre sí en el tiempo por un período  $T_A$  de análisis. Cuanto más elevado es el coeficiente  $V_i$ , más sonoro es el sonido de la voz en el instante  $t_i$ ; mientras que cuanto más bajo es el coeficiente  $V_i$ , menos sonoro es el sonido de la voz en el instante  $t_i$ .
- 50 A partir de estas mediciones, se deduce el estado instantáneo « Estado\_  $P_i$  » de la señal S de voz en cada punto  $P_i$ , pudiendo ser este estado « Estado\_  $P_i$  » el estado « silencio » correspondiente a la ausencia de una señal de voz de potencia suficiente, el estado « sonoro » correspondiente a la emisión de un sonido de naturaleza sonora, y el estado « sordo » correspondiente a la emisión de un sonido de naturaleza sorda.

- 5 A tal efecto, en una etapa 13, se compara la energía  $E_i$  instantánea de la señal  $S$  de voz con un umbral  $A$ . En un ejemplo este umbral vale 0,02 para una señal normalizada. Si la energía  $E_i$  de la señal es inferior al umbral  $A$ , entonces se deduce en una etapa 15 que el estado instantáneo « Estado\_Pi » del punto  $P_i$  es « silencio ». Por el contrario, si la energía  $E_i$  de la señal es superior al umbral  $A$ , entonces se deduce que un sonido de potencia suficiente sale efectivamente de la boca del jugador y a continuación se determina si el sonido es sonoro o sordo.
- A tal efecto, en una etapa 17, se compara el coeficiente de sonoridad  $V_i$  con un umbral  $B$ . En un ejemplo,  $B$  vale 0,3 para una señal normalizada. Si el coeficiente  $V_i$  de sonoridad es inferior al umbral  $B$  entonces, en una etapa 18, se deduce que el sonido es sordo (el estado instantáneo « Estado\_Pi » es entonces « sordo »). Esto significa que el jugador está verdaderamente pronunciando un sonido especialmente de tipo  $P, T, K, B, D, G, CH, F, S$ .
- 10 Mientras que si el coeficiente  $V_i$  de sonoridad es superior al umbral  $B$  entonces, en una etapa 19, se deduce que el sonido es sonoro (el estado instantáneo « Estado\_Pi » es entonces « sonoro »). Esto significa que el jugador está verdaderamente pronunciando una vocal o una consonante sonora.
- 15 En un ejemplo, para calcular la energía instantánea  $E_i$  y el coeficiente  $V_i$  de sonoridad, se aplica a la señal  $S$  de voz un algoritmo que permita extraer las frecuencias fundamentales de esta señal  $S$  representadas en función del tiempo por la curva  $S'$  en la Figura 1.
- 20 El coeficiente  $V_i$  de sonoridad corresponde al coeficiente  $Q$  de la medición de la calidad de la detección de la frecuencia fundamental por el algoritmo de detección de la frecuencia representado en función del tiempo por la curva  $S'$ . La calidad de extracción corresponde a la fiabilidad de la detección de la frecuencia fundamental. La calidad  $Q$  de la extracción de la frecuencia fundamental de la señal de voz  $S$ , que está en relación muy estrecha con la sonoridad de la voz, será muy elevada en las partes sonoras de la voz en el transcurso de las cuales las cuerdas vocales vibran, lo que permite extraer fácilmente la frecuencia fundamental de la señal  $S$  de voz. Mientras que la calidad  $Q$  de la extracción de la frecuencia fundamental de la señal  $S$  de voz será poco elevada en las partes sordas en el transcurso de las cuales las cuerdas vocales no vibran o vibran muy poco, lo que hace difícil la extracción de la frecuencia fundamental de la señal  $S$  de voz.
- 25 En un ejemplo, el algoritmo de detección de la frecuencia fundamental es el algoritmo S/N. Este algoritmo, conocido por el especialista en la materia, está descrito de manera precisa en el documento de patente de France Telecom que tiene el número de registro nacional francés 0107284. La calidad de detección de la altura es el valor  $(1-d')$  siendo  $d'$  la función diferencia promediada y normalizada del algoritmo de S/N tal como se describe en el seno del documento de patente de France Telecom que tiene el número de registro nacional francés 0107284, y representada en función del tiempo por la curva  $S''$ .
- 30 En variante, el coeficiente de sonoridad es por ejemplo una medición del ruido no armónico contenido en la señal audio, medida por ejemplo por la tasa de cruce por cero (ZCR), siendo un valor bajo de ZCR característico de un sonido sonoro mientras que un valor elevado de ZCR es característico de un sonido sordo. La utilización del ZCR es particularmente ventajosa en el caso en que se desee reducir al máximo el consumo de la CPU del sistema.
- 35 En un ejemplo, siendo muestreada la señal de voz a 16 kHz, la energía instantánea  $E_i$  y la calidad  $Q_i$  se calculan cada  $TA = 20$  ms aplicando el algoritmo de la frecuencia fundamental en los últimos 1024 puntos muestreados de la señal  $S$  de manera que se efectúe un corte entre los diferentes trozos de la señal  $S$  analizada (correspondiendo los últimos 1024 puntos aproximadamente a 3 periodos  $TA$  de 20 ms). En variante, no hay recorte entre los diferentes trozos de la señal analizada.
- 40 A continuación, como representa la Figura 3, en una etapa 25, se efectúa un análisis de la señal  $S$  de voz en un período de referencia  $TR_i$  de duración de referencia  $TR$  (aproximadamente un segundo) antes del instante  $t_i$ , lo que equivale a conservar los 50 últimos estados instantáneos Estado\_Pj para  $TA = 20$  ms. En variante, el número de estados instantáneos Estado\_Pj conservados podría ser diferente para efectuar un análisis en un período de referencia  $TR_i$  más o menos largo. En variante, el período de referencia  $TR_i$  puede ser reemplazado por un conjunto de puntos alrededor del instante  $t_i$ , sean estos puntos antes o después del instante  $t_i$ .
- 45 En una etapa 27, se analiza si los  $N$  últimos estados (típicamente  $N = 5$ , o sea 100 ms) instantáneos Estado\_Pj de la señal  $S$  son silencios. Si éste es el caso, se deduce que el instante  $t_i$  de análisis es un instante de silencio. Si no, se deduce que  $t_i$  no es un instante de silencio y se determina entonces si se trata de un instante  $t_i$  cantado o tarareado.
- 50 A tal efecto, en una etapa 30, se conservan entre los 50 últimos estados instantáneos de la señal únicamente los estados instantáneos « Estado\_Pj » de tipo « sonoro » o « sordo » con exclusión de los estados de silencio. Después, en una etapa 33, se analiza si todos los instantáneos « Estado\_Pj » conservados son estados « sonoros ». Si éste es el caso, entonces, en la etapa 34, se deduce que la señal  $S$  de voz corresponde a un tarareo en el instante  $t_i$  puesto que a priori es imposible no observar al menos un paso sordo en el transcurso del período de referencia  $TR_i$  en un lenguaje cantado. Por el contrario, si no hay únicamente estados sonoros, entonces, en la etapa 35, se deduce que la señal  $S$  de voz corresponde a un canto de palabras en el instante  $t_i$  puesto que a priori es natural observar al menos un paso sordo en el transcurso del período de referencia  $TR_i$  en un lenguaje cantado con palabras.
- 55

Durante la utilización de la invención en un Karaoke, el jugador podrá ser penalizado por cada instante  $t_i$  durante el cual éste ha tarareado en lugar de cantar las palabras de la canción que hay que interpretar, o por el contrario recompensado por cada instante  $t_i$  en que éste ha cantado con las palabras.

5 Es posible que ciertas canciones presenten pasos sonoros que tengan una duración superior a la duración TR del período de referencia. Así, la Figura 4 muestra la amplitud 41 de la señal S de voz correspondiente a las palabras 42 de una canción en la cual el paso 42.1 completamente sonoro « la luna mi amigo » (en gris) tiene una duración TD superior a la duración TR del período de referencia.

10 A fin de evitar falsas detecciones de tarareo en estas palabras particulares, puede ser útil inhibir la función de detección de las palabras en toda la duración TD del paso sonoro 42.1. Así, como muestra la banda 43 de la Figura 4, la función de detección de palabras de acuerdo con la invención está inhibida en el período TD (puesta en OFF) pero activada en el resto de la canción (puesta en ON).

Se puede también activar esta función de detección de las palabras durante una parte solamente de la canción (por ejemplo el estribillo) en la cual hay que conocer las palabras y no durante otras (por ejemplo las estrofas) durante las cuales el conocimiento de estas palabras pasa a ser facultativo.

15 Se observa que la detección de los silencios en la señal S de voz optimiza el funcionamiento del procedimiento de acuerdo con la invención porque éste evita que ciertos ruidos blancos parásitos no sean considerados arbitrariamente como sonidos de tipo sonoro o sordo. Sin embargo, en variante, en un funcionamiento degradado, se suprimen las etapas 13, 15, 27 y 29 de detección de los silencios y se analiza simplemente si el estado instantáneo « Estado\_Pi » de la señal S es « sonoro » o « sordo », después se analizan los estados instantáneos de la señal S de voz en el período de referencia TR<sub>i</sub>. Se deduce que el jugador tararea si todos los estados instantáneos son de tipo sonoros y que éste canta en el caso contrario.

20

**REIVINDICACIONES**

1. Procedimiento para detectar palabras cantadas con respecto al tarareo en una señal (S) de voz de un usuario caracterizado por que comprende las etapas siguientes:
- medir un coeficiente de sonoridad ( $V_i$ ) en diferentes instantes de un período de referencia ( $TR_i$ ),
- 5 - comparar los coeficientes de sonoridad ( $V_i$ ) así medidos en el período de referencia ( $TR_i$ ) con un valor umbral (B), y
- en función de los resultados de estas comparaciones (Estado\_Pj) en el período de referencia ( $TR_i$ ), deducir si, en un instante de análisis (ti), el usuario está pronunciando palabras cantadas si no hay únicamente estados sonoros en el transcurso del citado período de referencia ( $TR_i$ ), o está tarareando si todos los estados instantáneos conservados en el transcurso del citado período de referencia ( $TR_i$ ) son estados sonoros.
- 10 2. Procedimiento de acuerdo con la reivindicación 1, caracterizado por que el período de referencia ( $TR_i$ ) precede al instante de análisis (ti).
3. Procedimiento de acuerdo con las reivindicaciones 1 o 2, caracterizado por que:
- si el coeficiente ( $V_i$ ) de sonoridad es superior al valor umbral (B) durante el período ( $TR_i$ ) de referencia, entonces:
- 15 - se deduce que no hay ningún instante sordo en la voz durante esta duración umbral y que el usuario tararea en el instante de análisis (ti),
- si no, se deduce que el usuario pronuncia palabras en el instante de análisis (ti).
4. Procedimiento de acuerdo con una de las reivindicaciones 1 a 3, caracterizado por que el coeficiente ( $V_i$ ) de sonoridad es el valor ( $1-d'$ ), siendo  $d'$  la función diferencia promediada y normalizada del algoritmo de S/N.
- 20 5. Procedimiento de acuerdo con una de las reivindicaciones 1 a 4, caracterizado por que el período ( $TR_i$ ) de referencia es del orden de 1 segundo.
6. Procedimiento de acuerdo con una de las reivindicaciones 3 a 5, caracterizado por que la etapa de comparación del parámetro ( $V_i$ ) de sonoridad con el valor umbral (B) se efectúa únicamente si la energía ( $E_i$ ) de la señal (S) de voz es superior a un valor umbral (A).
- 25 7. Procedimiento de acuerdo con las reivindicaciones 1 o 2, caracterizado por que, siendo muestreada la señal de voz (S), éste comprende las etapas siguientes:
- calcular una intensidad instantánea ( $E_i$ ) y un coeficiente ( $V_i$ ) de sonoridad instantánea para puntos ( $P_i$ ) de la señal de voz en instantes (ti) de análisis espaciados entre sí por un período de análisis ( $TA$ ) en el período de referencia ( $TR_i$ ),
- 30 - determinar los estados instantáneos « Estado\_Pi » de la señal (S) de voz en cada instante ti a partir de las mediciones de la energía  $E_i$  instantánea y de la sonoridad ( $V_i$ ) de la señal (S) de voz, pudiendo ser estos estados instantáneos el estado « sonoro » correspondiente a la emisión de un sonido de naturaleza sonora, o el estado « sordo » correspondiente a la emisión de un sonido de naturaleza sorda,
- si todos los estados instantáneos « Estado\_Pj » son de tipo « sonoros » en el período ( $TR_i$ ) de referencia entonces se deduce que no hay pronunciación de palabras en la señal (S) de voz en el instante de análisis (ti),
- 35 - si no, se deduce que hay pronunciación de palabras en la señal (S) de voz en el instante de análisis (ti).
8. Procedimiento de acuerdo con la reivindicación 7, caracterizado por que para determinar el estado instantáneo « Estado\_Pi » de la señal S de voz en el instante de análisis (ti),
- se compara el coeficiente de sonoridad  $V_i$  con un umbral (B),
- 40 - si el coeficiente ( $V_i$ ) de sonoridad es inferior al umbral (B) entonces el estado instantáneo « Estado\_Pi » es « sordo »,
- si no, se deduce que el estado instantáneo Estado\_Pi es « sonoro ».
9. Procedimiento de acuerdo con la reivindicación 7, caracterizado por que el estado instantáneo « Estado\_Pi » puede tomar además el estado « silencio » correspondiente a la ausencia de un sonido de potencia suficiente,
- 45 - si los N últimos estados instantáneos « Estado\_Pj » en el período ( $TR_i$ ) de referencia son de tipo « silencio » entonces se deduce que la señal no contiene voz en el instante (ti), si no

- se conservan, en el período de referencia ( $TR_i$ ), únicamente los estados instantáneos de tipo « sonoros » o « sordos » con la exclusión de los estados instantáneos « Estado\_Pj » de tipo « silencio ».

10. Procedimiento de acuerdo con la reivindicación 9, caracterizado por que para determinar el estado instantáneo « Estado\_Pi » de la señal (S) de voz,

- 5 - se compara la energía ( $E_i$ ) instantánea de la señal S de voz con un primer umbral (A),
- si la energía ( $E_i$ ) de la señal es inferior al umbral (A), entonces se deduce que el estado instantáneo « Estado\_Pi » es « silencio »,
- si no, se compara el coeficiente de sonoridad ( $V_i$ ) con un segundo umbral (B),

10 si el coeficiente ( $V_i$ ) de sonoridad es inferior al segundo umbral (B) entonces el estado instantáneo « Estado\_Pi » es « sordo »,

si no, se deduce que el estado instantáneo « Estado\_Pi » es « sonoro ».

11. Procedimiento de acuerdo con una de las reivindicaciones 7 a 10, caracterizado por que el período (TA) de análisis vale 20 ms y la duración (TR) del período de referencia 1 s.

15 12. Procedimiento de acuerdo con las reivindicaciones 7 a 11, caracterizado por que la señal de voz (S) es muestreada a 16 kHz.

13. Utilización del procedimiento de acuerdo con una de las reivindicaciones 1 a 12 en una aplicación de tipo juego de Karaoke.

20 14. Utilización de acuerdo con la reivindicación 13, caracterizada por que se inhibe la puesta en práctica del procedimiento de acuerdo con una de las reivindicaciones 1 a 12 en los pasos sonoros (42.1) de canción que tengan una duración (TD) superior a la duración (TR) del período de referencia o en pasos de canciones arbitrariamente elegidos.

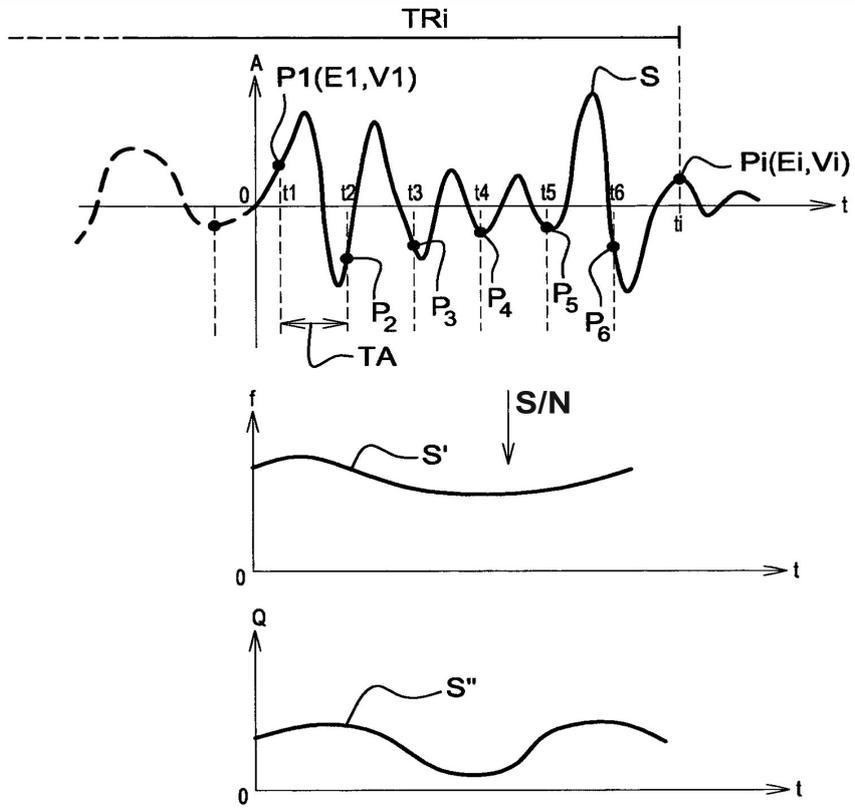


Fig. 1

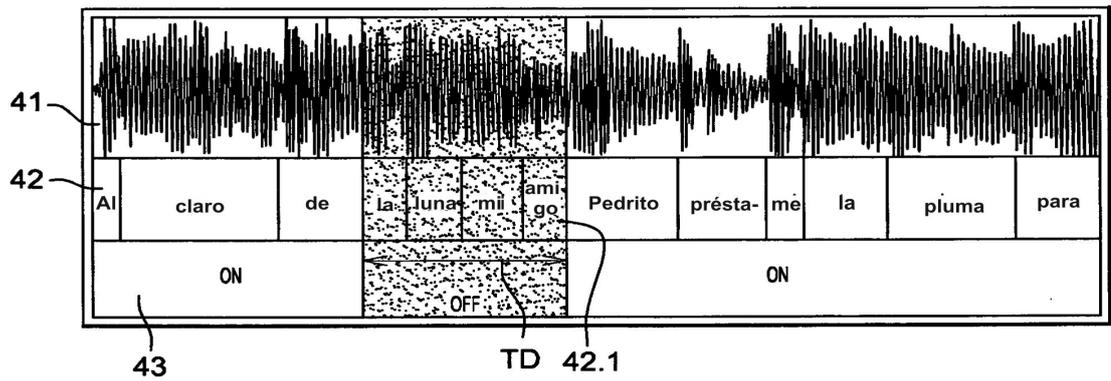


Fig. 4

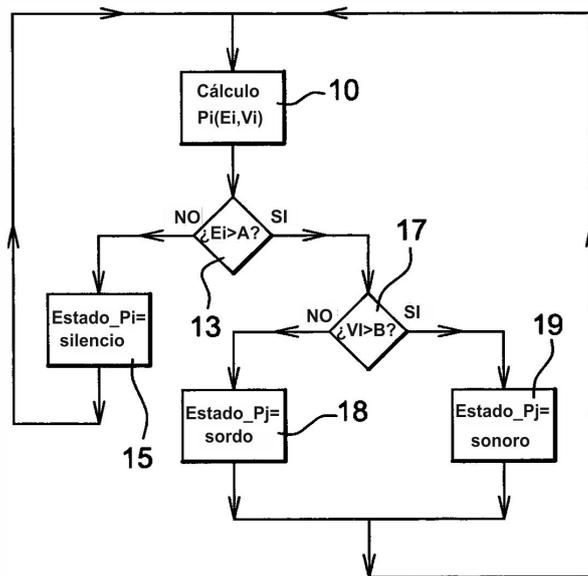


Fig. 2

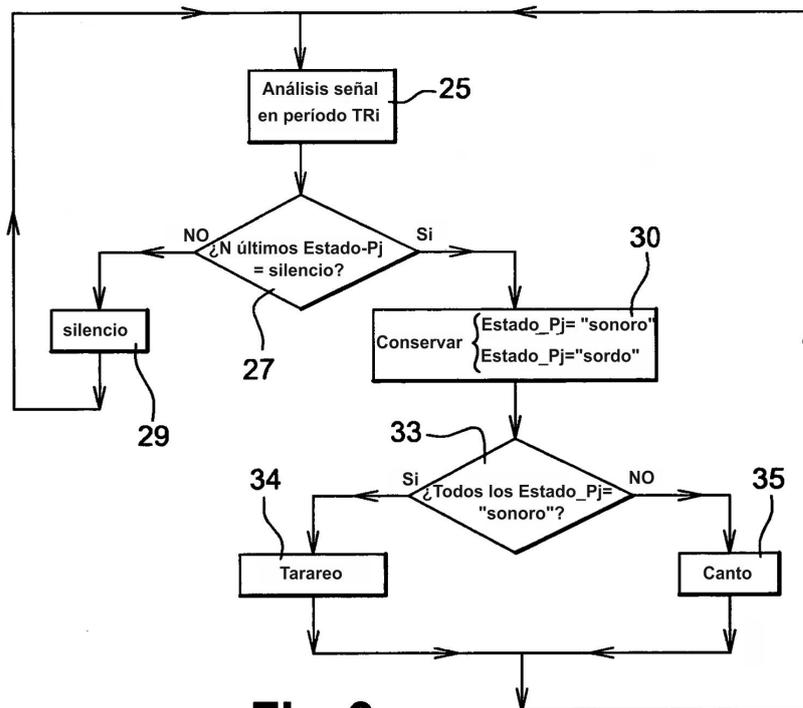


Fig. 3