



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

⑪ Número de publicación: **2 143 953**

⑫ Número de solicitud: 009801091

⑬ Int. Cl.⁷: G10L 11/00

⑭

PATENTE DE INVENCION

B1

⑮ Fecha de presentación: **26.05.1998**

⑯ Fecha de publicación de la solicitud: **16.05.2000**

Fecha de concesión: **23.10.2000**

⑰ Fecha de anuncio de la concesión: **01.12.2000**

⑱ Fecha de publicación del folleto de patente:
01.12.2000

⑲ Titular/es: **UNIVERSIDAD DE MALAGA**
Plaza de El Ejido s/n
29071 Málaga, ES

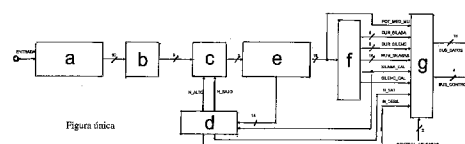
⑳ Inventor/es: **Martín Canales, José Francisco;**
Ríos Gómez, Francisco Javier;
Escaño Quero, Rafael y
Rodríguez Rueda, José

㉑ Agente: **No consta**

㉒ Título: **Circuito Integrado Analizador de Secuencias Silábicas.**

㉓ Resumen:
Circuito Integrado Analizador de Secuencias Silábicas.

Circuito Integrado monolítico Analizador de Secuencias Silábicas caracterizado por incorporar los siguientes módulos funcionales: módulo de muestreo/retención y convertor analógico/digital (a), módulo calculador del valor absoluto (b), módulo de control automático de ganancia (c), módulo detector de nivel (d), módulo calculador de la Potencia Promedio Muestreada (e), módulo detector de sílabas (f), y módulo selector de datos (g). Se presenta como solución para el análisis en tiempo real de la potencia del hablante, discriminador habla-silencio y calculador del ritmo de producción silábica con aplicaciones en codificación, reconocimiento de voz y aplicaciones biomédicas en Foniatría y Logopedia.



Aviso: Se puede realizar consulta prevista por el artº 37.3.8 LP.

Venta de fascículos: Oficina Española de Patentes y Marcas. C/Panamá, 1 - 28036 Madrid

ES 2 143 953 B1

DESCRIPCION

Circuito Integrado Analizador de Secuencias Silábicas.

Sector de la técnica

Campo de la Industria Electrónica y Micro-electrónica orientado hacia la Instrumentación y Medida de parámetros biológicos de la voz para aplicaciones en codificación, reconocimiento de voz y aplicaciones clínicas en Foniatría y Logopedia. La invención se presenta como solución para la medida en tiempo real de la potencia media promedio del hablante, la discriminación habla-silencio y calculo del ritmo de producción silábico.

Estado de la técnica

Las técnicas de Procesado digital de señales y el estado actual de la microelectrónica, permiten la posibilidad de abordar soluciones tales como las que se presentan en esta invención: Un circuito integrado monolítico de aplicación específica (ASIC) que permite la estimación en tiempo real de la potencia media temporal de una señal de voz y sobre ella la implementación de algoritmos de detección de habla-silencio y de identificación silábica.

Las señales generadas por el aparato fonador humano en general, y el ritmo de producción silábica en particular, se puede estudiar en los trabajos de Fant, G.C.M.: '*Acoustic Theory of Speech Production*. Mouton and Co. S.Gravenhage, the Netherlands, 1960, y Flanagan, J.L. '*Speech Analysis Synthesis and Perception*'. Springer-Verlag, Berlin, 1972. Consideremos las siguientes definiciones: "Ritmo de producción" es la relación 'unidades silábicas' -tiempo en un determinado segmento de voz. "Unidad Silábica" es el sonido articulado que constituye un sólo núcleo energético entre dos depresiones sucesivas de emisión de voz. Una unidad silábica es un segmento de voz independiente. Para el idioma español, la unidad silábica así definida casi siempre coincide con el concepto lingüístico de sílaba, pero existen excepciones. Por ejemplo, aplicando las reglas de la fonética clásica a la frase "Ha llegado un camión" nos encontramos con siete sílabas (ha-lle-ga-do-un-ca-mión), mientras que aplicando el concepto energético solo nos encontramos con seis (ha-lle-ga-do-un-ca-mión).

El algoritmo de detección silábico presentado en esta invención ha sido desarrollado partiendo del concepto de Energía *short-time* propuesto por Rabiner L. R. & Schafer R. W. '*Digital Signal Processing of Speech Signals*'. Prentice Hall. New Jersey, 1978 Capítulo 4. Dada una secuencia de muestras de una señal de voz $x(m)$ y dada una ventana rectangular de longitud N, el valor short-time de la energía viene dado por:

$$E_n = \sum_{m=n}^{n+N-1} x^2(m)$$

El valor de la energía así estimado en un segmento de longitud N y su relación con los paquetes energéticos diferenciables en la voz humana depende del valor de N y del nivel de amplitud de la señal $x(m)$.

Explicación de la invención

Circuito Integrado Analizador de Secuencias Silábicas. Estimador del ritmo de producción silábica en una señal de voz continua basado en el cálculo de la Potencia promedio. Incluye interfaz analógico-digital, control automático de ganancia, calculador de Potencia promedio, detector de sílabas-silencios y unidad de control en estructura Monolítica de Silicio. La voz digitalizada, segmentada y normalizada se somete a un proceso de extracción de la energía media temporal. Previa estimación del umbral energético de silencio, se aplica un algoritmo de detección silábico definido como aquel sonido articulado que constituye un paquete energético entre dos depresiones sucesivas de emisión. Los criterios de diferenciación se basan en la detección por umbral.

Este circuito integrado procesa una señal de voz continua y produce en tiempo real los parámetros necesarios para el estudio del secuenciamiento silábico. Posee una entrada analógica y una entrada-salida digital tipo Bus que le proporciona capacidad de funcionamiento autónomo mediante un mínimo de componentes pudiendo funcionar bajo el control de un microcontrolador (μC) o microprocesador (μP).

Resumen de características

Circuito integrado monolítico calculador de los parámetros en tiempo real:

Potencia Promedio

Número de sílabas detectadas desde el comienzo de la medida hasta el instante presente.

Detector del instante de silencio.

Detector de la duración de silencio.

Detector del instante de sílaba

Detector de la duración de la sílaba

Igualmente, como elementos de ayuda para control externo proporciona:

Aviso de nivel de señal de entrada bajo

Aviso de nivel de señal de entrada alto

Otras Características:

Selección del conversor A/D: interno o externo.

Elección de parámetros por usuario disponibles en el Bus de Datos.

Activación o Desactivación del Control Automático de Ganancia

El circuito integrado analizador de secuencias silábicas tiene los siguientes bloques funcionales básicos:

- Módulo de muestreo/retención y conversor analógico/digital (a).

- Módulo calculador del valor absoluto (b).

- Módulo de control automático de ganancia (c).
- Módulo detector de nivel (d).
- Módulo calculador de la Potencia Promedio Muestreada (e).
- Módulo detector de sílabas (f).
- Módulo selector de datos (g).

El algoritmo de detección silábica usado en esta invención está basado en los siguientes presupuestos:

1. La estimación del valor absoluto de la señal de entrada como preproceso de optimización en la estimación de la potencia promedio.
2. La aplicación de un proceso digital de control automático de ganancia como medida de desensibilización de la variabilidad de la amplitud de la señal de voz.
3. Estimación del parámetro Potencia Promedio Temporal P_n definido como:

$$P_n = \sqrt{E_n}$$

La Potencia Promedio P_n se estima en un marco específico o ventana temporal de 600 muestras.

4. La estimación del parámetro derivado '*Potencia Promedio muestreada P_{m_n}* ' con propiedades de menor sensibilidad a la amplitud y con la variabilidad adecuada para la correcta discriminación de los paquetes energéticos propios del habla humano:

Para $n=0, m, 2m, \dots$

P_{m_n} se estima muestreando P_n a intervalos de 302 muestras. Existe por tanto un solapamiento temporal en el cálculo de dos P_{m_n} consecutivos.

5. La aplicación de criterios de detección de mínimos relativos para la discriminación de paquetes silábicos.

Explicación de la figura

Se ha empleado una figura única que representa el sistema a nivel funcional. Cada bloque representa un subsistema del circuito integrado. El nombre de cada bloque indica la funcionalidad específica de cada uno, y el orden de concatenación de los bloques muestra el secuenciamiento del proceso de cálculo seguido por el circuito integrado. La descripción del sistema global así como la de los subsistemas que lo componen, está recogida en el apartado descriptivo de la realización preferente de la invención.

Realización preferente de la invención

La figura única muestra el diagrama general de bloques de la invención en lo que respecta a sus módulos funcionales básicos, representando cada módulo un subsistema del circuito integrado. Estos son:

- Módulo de muestreo/retención y conversor analógico/digital (a). Se encarga de muestrear la señal analógica de voz previamente filtrada a 4.5 KHz. Se compone de un muestreador retenedor y un conversor analógico/digital según la técnica de aproximaciones sucesivas de 10 bits con codificación unipolar. La frecuencia de muestreo es de 10KHz.
- Módulo calculador del valor absoluto (b). Calcula el valor absoluto de la señal de entrada. Dada una señal codificada unipolarmente, el bit i -ésimo z_i de la salida se obtiene a partir del bit i -ésimo de la entrada x_i de acuerdo con la operación:

$$z_i = x_i \oplus \bar{x}_9 \quad 0 \leq i \leq 8$$

siendo el símbolo \oplus la operación OR EXCLUSIVA

- Módulo de control automático de ganancia (c). Su función es adecuar el nivel de la señal que será entregado al módulo calculador de la Potencia Promedio. Este módulo posee una entrada de datos procedente del módulo (b) de 9 bits y dos entradas de control procedentes del módulo detector de nivel (d). Este módulo junto con el módulo calculador de la Potencia Promedio Muestreada (e), conforman un sistema realimentado negativamente. Las señales de control son de tipo flag. Este módulo posee tres niveles de control de ganancia que se obtienen mediante una selección de 7 bits de los nueve bits de entrada, por un proceso de desplazamiento a la izquierda (atenuación) o a la derecha (amplificación). Este proceso posee la ventaja del mantenimiento de la precisión de cálculo.
- Módulo detector de nivel (d). Unidad detectora del nivel de la señal de entrada que sensa la Energía media en un segmento previo de voz mediante, generando cuatro señales de control tipo flag, dos de ellas de nivel y otras dos de pulso. Las dos primeras muestran un funcionamiento incorrecto del sistema debido a un mal acondicionamiento de los niveles de la señal de entrada. Las otras dos actúan sobre el módulo de control automático de ganancia (c) para controlar su ganancia efectiva de acuerdo con lo descrito en el apartado anterior. La realimentación se realiza de forma sincronizada con la detección de una sílaba, lo que evita la aparición de falsos mínimos debidos a un reajuste de ganancia.
- Consta de un sumador, dos registros, cuatro detectores de umbral, y dos circuitos detectores de pulsos.
- Módulo calculador de la Potencia Promedio Muestreada (e). Se encarga de estimar la

Potencia Promedio Muestreada (P_{m_n}) de acuerdo con la expresión:

$$P_{m_n} = P_n$$

para $n=0, m, 2m, ..$

donde $P_n = \sqrt{E_n}$

Este módulo posee una entrada de datos de 7 bits y dos salidas de datos, una correspondiente al valor de la Potencia Promedio Muestreada (16 bits) y otra correspondiente al valor de la entrada al cuadrado (14 bits), siendo esta última utilizada por el módulo (d) para la determinación del nivel de ganancia óptimo.

El cálculo de la Potencia Promedio Muestreada, es realizada por este módulo mediante las operaciones: elevación al cuadrado, acumulación o suma, y extracción de la raíz cuadrada. La operación de acumulación se realiza teniendo en cuenta el solapamiento temporal del cálculo de la Potencia Promedio Muestreada.

- Módulo detector de sílabas (f). Partiendo de la señal de Potencia Promedio Muestreada, éste módulo detecta Sílabas y Silencios. Dentro del proceso de producción de voz, definimos sílaba como aquel sonido articulado que constituye un paquete energético entre dos depresiones sucesivas de emisión. La sílaba así determinada se caracteriza por su duración temporal y por su instante temporal en el que concluye. Los criterios de diferenciación se basan en la detección por umbral en la señal de Potencia Promedio Muestreada de acuerdo con los siguientes criterios:

1. Si la Potencia Promedio Muestreada está por debajo de un umbral mínimo significa que estamos detectando un periodo de silencio.
2. Si la Potencia Promedio Muestreada está por encima de ese umbral mínimo significa que existe producción de voz.

La condición de sílaba detectada se puede deber a dos situaciones:

- a. Un mínimo relativo en un segmento de voz.
- b. Un paso de segmento de voz a segmento de silencio.

La condición de silencio detectado se produce en una única situación:

- a. Un paso de segmento de silencio a seg-

mento de voz.

Este módulo consta de tres registros, un comparador detector de condición mayor o igual, un comparador detector de condición estrictamente mayor, un circuito combinatorial discriminador de umbral Habla-Silencio, y tres contadores.

La señal de Potencia Media Muestreada se almacena consecutivamente en los tres registros (valor en curso, valor anterior y valor anterior del anterior). El cálculo se realiza sobre el valor de la Potencia Promedio Muestreada anterior, por lo que los resultados están retrasados del tiempo real en una ventana (20 milisegundos). El valor en curso y el valor anterior del anterior son los datos de comparación para la toma de decisión del carácter de la potencia almacenada. El contenido del registro asociado a la Potencia Promedio Muestreada en curso es evaluado por un el circuito combinatorial para determinar si se trata de un segmento de voz o de un segmento de silencio. Finalmente, los tres contadores, se encargan respectivamente de llevar la cuenta de la longitud temporal del silencio en curso, sílaba en curso y número total de sílabas detectados hasta el momento.

- Módulo selector de datos (g). Se encarga de presentar a través de los dos buses de salida y de forma selectiva, el conjunto de resultados correspondientes a las operaciones realizadas: en el bus de datos (16 bits) se presentan los resultados de forma secuencial: número de sílabas en curso (16 bits), duración del último silencio calculado (8 bits), duración de la última sílaba calculada (8 bits), y valor de la Potencia Promedio Muestreada (16 bits). Los Datos se presentan en el bus de datos de forma excluyente, salvo los de longitud 8 bits que son simultáneos. Todos los datos están disponibles en cualquier momento siendo el bus de control el que proporciona las señales para la generación de la toma de decisiones. El bus de control (4 bits), presenta simultáneamente dos conjuntos de señales: las primeras se corresponden con el modo de funcionamiento normal del circuito integrado y las segundas son de tipo test.

Consta de tres multiplexores que se encargan de suministrar las señales tal como hemos descrito.

Todos estos Módulos funcionales se encuentran gestionados por una unidad de control centralizada que proporciona las señales de sincronismo para garantizar el funcionamiento encauzado (PipeLine).

REIVINDICACIONES

1. Circuito Integrado monolítico Analizador de Secuencias Silábicas **caracterizado** por incorporar los siguientes módulos funcionales: módulo de muestreo/retención y conversor analógico/digital (a), módulo calculador del valor absoluto (b), módulo de control automático de ganancia (c), módulo detector de nivel (d), módulo calculador de la Potencia Promedio Muestreada (e), módulo detector de sílabas (f), y módulo selector de datos (g).

2. Circuito Integrado Analizador de Secuencias Silábicas, según la reivindicación 1 **caracterizado** por:

- El procedimiento digital por el cual se ajusta el nivel de la señal de entrada al módulo calculador de la Potencia Promedio Muestreada: Control del desplazamiento de una palabra de 9 bits para conseguir una señal de siete bits con el nivel adecuado.
- El procedimiento para la detección del nivel de la señal de entrada al Módulo Calculador de la Potencia Promedio Muestreada: Basado en la detección de la Energía media en un segmento previo de señal generándose cuatro señales de control indicativas de nivel saturado, alto, normal, bajo y débil.

- El procedimiento utilizado para el Cálculo de la Potencia Promedio Muestreada según la expresión:

$$P_{m_n} = P_n$$

para $n=0, m, 2m, \dots n=600$ y $m=302$

donde $P_n = \sqrt{E_n}$

Mediante el uso de dos acumuladores compartiendo una única etapa multiplicadora por un adecuado proceso de sincronización que alterna la inicialización y selección de forma alternativa entre uno y otro acumulador.

- El procedimiento de Detección de Sílabas y Silencios en cuanto a su producción y en cuanto a su duración. Criterios y condiciones de diferenciación silábica.

3. Uso del Circuito Integrado Analizador de Secuencias Silábicas **caracterizado** en las reivindicaciones 1 a 2, en aplicaciones de Codificación, reconocimiento de voz y aplicaciones clínicas en Foniatría y Logopedia tanto en soluciones de sobremesa basadas en microcontroladores y/o microprocesadores, como en soluciones autónomas de tipo portable.

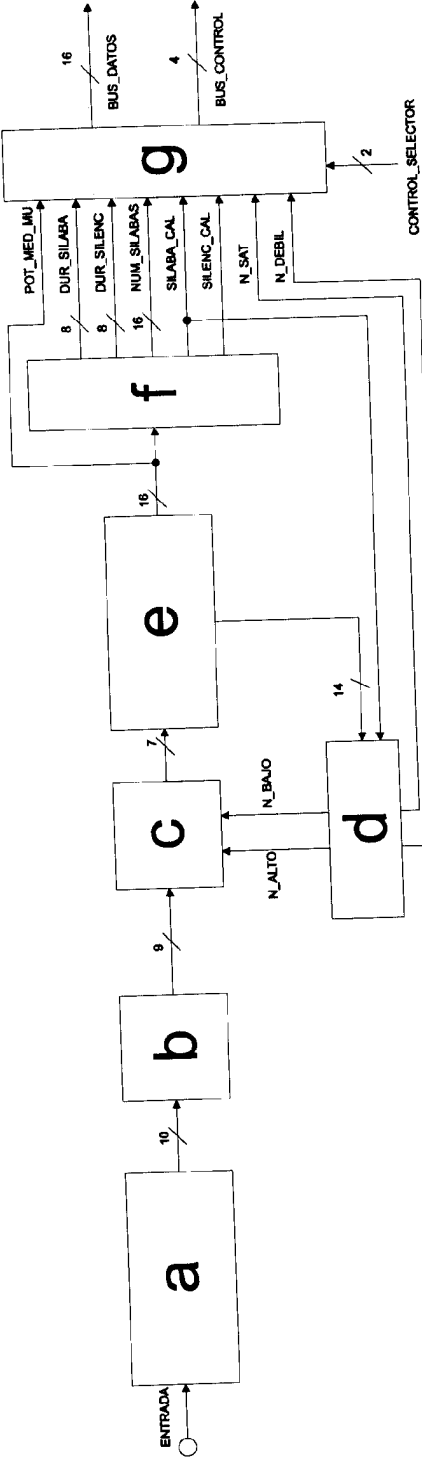


Figura única



OFICINA ESPAÑOLA
DE PATENTES Y MARCAS
ESPAÑA

- ⑪ ES 2 143 953
⑫ N.º solicitud: 009801091
⑬ Fecha de presentación de la solicitud: 26.05.1998
⑭ Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TECNICA

⑮ Int. Cl.⁷: G10L 11/00

DOCUMENTOS RELEVANTES

Categoría	Documentos citados	Reivindicaciones afectadas
A	US 4703504 A (VITTORELLI, V.) 27.10.1987, todo el documento.	1-3
A	US 4908864 A (TOGAWA et al.) 13.03.1990, resumen; columna 1, línea 47 - columna 2, línea 19; columna 2, línea 58 - columna 3, línea 40; figura 1.	1-3
A	US 5056150 A (YU et al.) 08.10.1991, todo el documento.	1-3
A	PATENT ABSTRACTS OF JAPAN, CD-ROM MIJP 9608 PAJ 1996 - 12[023] [08-196101/08-224000], JP 08-211893 A (TOSHIBA CORP) 20.08.1996, resumen; figura.	1-3
A	PATENT ABSTRACTS OF JAPAN, CD-ROM MIJP 9511 PAJ 1996 - 3[014] [07-289001/07-312900], JP 07-302098 A (KENWOOD CORP.) 14.11.1995, resumen; figura.	1-3

Categoría de los documentos citados

X: de particular relevancia

Y: de particular relevancia combinado con otro/s de la misma categoría

A: refleja el estado de la técnica

O: referido a divulgación no escrita

P: publicado entre la fecha de prioridad y la de presentación de la solicitud

E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

El presente informe ha sido realizado

☒ para todas las reivindicaciones

☐ para las reivindicaciones n.º:

Fecha de realización del informe

08.04.2000

Examinador

O. González Peñalba

Página

1/1