MINISTERIO DE INDUSTRIA Y ENERGIA

Registro de la Propiedad Industrial

(19) ES (11) (21) NUMERO 464.487 (10) A1

(22) FECHA DE PRESENTACION

ESPAÑA

Concedido al Registro de acuerdo con los datos que figuran en la presente descripción y según el contenido de la Memoria adjunta.

PATENTE DE INVENCION

(30) PRIORIDADES:

(31) NUMERO	(32) FECHA	(33) PAIS
745.066	26 de noviembre de 1.976	EE.UU. de A.

(47) FECHA DE PUBLICIDAD	(51) CLASIFICACION INTERNACIONAL	(62) PATENTE DE LA QUE ES DIVISIONARIA
	G 10 L	

(54) TITULO DE LA INVENCION

SISTEMA DE RECONOCIMIENTO DE VOZ A PARTIR DE CARACTERISTICAS REPRESENTATIVAS DE SONIDOS HABLADOS CONOCIDOS.

(71) SOLICITANTE (S)

WESTERN ELECTRIC COMPANY, INCORPORATED

DOMICILIO DEL SOLICITANTE

222 Broadway, New York, 10038, EE.UU. de A.

(72) INVENTOR (ES)

EDWARD HENRY HAFER

(73) TITULAR (ES)

(74) REPRESENTANTE

D. JOSE MIGUEL GOMEZ-ACEBO Y POMBO.

UTILICESE COMO PRIMERA PAGINA DE LA MEMORIA

Este invento se refiere a un procedimiento para reconocer conversación desarrollando características que representan los sonidos de la conversación que han de ser reconocidos y equiparando las características con características predeterminadas representativas de sonidos hablados conocidos.

Un obstaculo principal en el progreso en el campo de reconocimiento automático de conversación ha sido la gran variación en características de conversación entre individuos, particularmente entre hombres, mujeres y niños. Para poder vencer este obstaculo se han desarrollado sistema que pueden responder a cualquier persona que converse pero que reconocen solamente un vocabulario limitado.

Uno de dichos sistemas ha sido descrito por T.R.Martin "Acoustic Recognition of a Limited Vocabulary in Continuos Speech", Universidad de Pennsylvania, Ph, D.Thesis, 1970 que se puede obtener de University Microfilms, Ltd, Wycorn, Inglaterra, y University Microfilms, Ann Arbor, Michigan, USA, sistema que reconoce un vocabulario limitado extractando características particulares de la señal de conversación y equiparando la secuencia derivada de características con un conjunto de secuencia de características previamente alegidas que representan el vocabulario que se desea reconocer. Las características elegidas son características de los sonidos elementales en la conversación.

En el área del estudio fisiológico de la conversación, se ha determinada que las trayectorias de la lengua de diferentes personas en conversación que pronuncian la misma palabra, son muy similares; en particular, con respecto a la posición clavo del movimiento de la lengua.

Por lo tanto, determinando la posición de la lengua, por medio de un elemento radiante montado en la punta

la lengua de la persona que habla se pueden reconocer palabras habladas, especialmente en combinación con los sistemas como los desarrollados por Martin, v.g., mediante una máquina. Las dificultades que surgen con dichos sistemas combinados, y especialmente con relación a los dispositivos conocidos con anterioridad a este invento, utilizados para determinar la posición de la lengua de la persona que habla (por medio de algún dispositivo montado directamente sobre la lengua de la persona que habla) son bastantes complejas y los sistemas impracticables.

Estas dificultades se resuelven en gran medida, según este invento, en un método de reconocer la conversación que se caracteriza porque se identifican los formantes contenidos en los sonidos de la conversación que se han de reconocer se convierten los formantes identificados en características de posición de la lengua de movimiento de acuerdo con un modelo de tracto vocal, y se equiparan dichas características de posición de la lengua y movimiento con características predeterminadas representativas de sonidos hablados conocidos.

En el dibujo:

La figura 1 representa una vista en sección transversal de la cavidad bucal con un sistema tratado en coordenadas x-y en la misma.

La figura 2 ilustra la trayectoria del cuerpo de la lengua al pronunciar los dígitos en idioma Inglés "eight" (8), "two" (2), "one" (uno), y "five" (cinco), de acuerdo con el sistema de coordenadas de la figura 1.

La figura 3 representa un sistema de coordenadas x-y subdividido empleado para trazar el mapa de las posiciones del cuerpo de la lengua en regiones características de sonidos vocalicos.

La figura 4 es un diagrama de conjuntos de una modalidad de este invento.

La figura 5 ilustra el diagrama de estados del aceptador 300 de la figura 4 perteneciente al vocablo en idioma Inglés "two eight" (dos ocho).

La figura 6 ilustra el diagrama de conjuntos de la memoria exigida en el aceptador 300; y

La figura 7 representa un diagrama de conjuntos del aparato para poner en práctica el diagrama de estados de la figura 5.

En general este invento reconoce conversación conectada de un vocabulario limitado v.g., los diez dígitos, derivando de la señal de un vocablo hablado un cierto número de características, incluyendo una característica de la trayectoria del cuerpo de la lengua, y descifrando de la misma las palabras que se habian pronunciado. De un modo más particular, la señal de conversación se analiza para desarrollar un cierto número de características similares a las empleadas en el pasado, más una nueva característica que caracteriza la posición y movimiento de la lengua de la persona que habla. La derivación de la posición de la lengua se consigue determinando las frecuencias formantes de la conversación y empleando un modelo de tracto vocal humano para hallar la posición de la lengua que coincida mejor con los formantes calculados. Una vez que se obtienen las características de la conversación, la sucesión de características se compara con las secuencias de las características de palabras elegidas y, partiendo de la comparación, se reconocen las palabras habladas.

La figura 1 representa una vista en sección transversal de una cavidad bucal con un eje x-y superpuesto en

la misma. Los ejes x-y de figuras subsiguientes se refieren al eje x-y de la figura 1.

Un estudio de los movimientos del cuerpo de la lengua revela que tanto si la persona que habla es un hombre, como si es una mujer o un niño, el cuerpo de la lengua recorre razonablemente la misma trayectoria cuando se pronuncia un dígito particular entre 0 y 9. La figura 2 representa ejemplos de dichas trayectorias del cuerpo de la lengua para ciertos números del idioma Inglés a partir de los cuales se puede recoger lo siguiente. El dígito "eigth" (ocho), curva 10, se caracteriza porque el cuerpo de la lengua se mueve en una dirección generalmente hacia delante y hacia arriba comenzando en el centro del cuadrante superior delantero de la cavidad bucal. El dígito "two" (dos), curva 20 se caracteriza porque el cuerpo de la lengua comienza en un punto elevado en el centro de la cavidad, se mueve horizontalmente hacia atrás y, cae hacia abajo en la parte posterior de la boca. El dígito "one", curva 30, se caracteriza porque el cuerpo de la lengua se mueve sencialmente hacia abajo en la parte posterior de la boca y después invierte su dirección moviendose hacia arriba. Finalmente, el dígito "five", curva 40 se caracteriza porque el cuerpo de la lengua se mueve hacia abajo en el cuadrante posterior inferior de la cavidad bucal y en ese cuadrante se mueve hacia adelante y hacia arriba en dirección al centro de la boca.

Partiendo de las descripciones de trayectorias anteriores se pueden comprender que las trayectorias del cuerpo de la lengua únicas de varios dígitos hablados, cuando se añaden a otras indicaciones de conversación, pueden mejorar el reconocimiento de dígitos hablados. Por lo tanto, según la forma de enfocar el problema de reconocimiento de conversación de este -

invento, la trayectoria del cuerpo de la lengua de una persona que habla se emplea como característica del sistema de reconocimiento de conversación, junto con una característica de silencio, una impulsión o una característica de consonante explosiva, y una característica de ruido y fricativa de modo de ruido (una para las fricativas sonoras y una para las fricativas sordas).

Respecto a las característica de la trayectoria del cuerpo de la lengua, se ha averiguado que en un sistema para reconocer dígitos, la posición y trayectoria exactas del cuerpo de la lengua no son necesarias para una caracterización apropiada de las características de trayectoria del cuerpo de la lengua, o distintivo. Un distintivo, en el contexto de este invento es la señal que representa la característica. Por el contrario, solamente la región general donde se sitúa el cuerpo de la lengua y su dirección general de movimiento son los rasgos que se tienen que conocer. Por consiguiente, el distintivo de la trayectoria del cuerpo de la lengua en la modalidad ilustrativa descrita en la presente memoria solamente distingue ciertas regiones de la cavidad bucal. La figura 3 representa las diversas regiones que han demostrado ser útiles en un sistema para detectar dígitos hablados indicando cada región la probabilidad de que se hayan pronunciado vocales de un cierto dígito. Por ejemplo, el cuerpo de la lengua situado en la región marcada con un 8 rodeado por un círculo, indica que lo más probable es que se haya pronunciado el sonido de la vocal inicial en el dígito "eight" (ocho).

Para desarrollar el distintivo de la trayectoria del cuerpo de la lengua se tienen que averiguar la posición y dirección de movimiento del lcuerpo de la lengua. La dirección de movimiento se obtiene comparando posiciones sucesivas del cuerpo de la lengua. Las posiciones del cuerpo de la lengua se

obtienen extrayendo las frecuencias de los formantes calculadas en posiciones del cuerpo de la lengua con ayuda del modelo del tracto vocal de Coker. Por "modelo del tractor vocal" se entiende un modelo físico del tracto vocal que se puede alterar de una forma controlada para producir diversos conjuntos de formantes de señales características de la conversación humana. En particular, por cada longitud del tracto vocal y posición de la lengua, dichos modelos generan un conjunto de formantes que caracterizan el sonido que se generaría al hablar una persona. Uno de dichos modelos ha sido descrito por C.H. Cocker en "A Model of Articulatory Dynamics and Control", actas del IEEE, volúmen 64, nº 4 1967, y también en la patente EE.UU 3.530.248, concedida a C.Coker el 22 de septiembre de 1970. Como por cada posición del cuerpo de la lengua estos modelos de tracto vocal proporcionan un conjunto de frecuencias de formantes esperados, utilizando el modelo a la inversa, se puede conocer la posición del cuerpo de la lengua partiendo de cada conjunto de formantes calculados. El empleo, por ejemplo, del modelo de Coker se expondrá con más detalle más adelante conjuntamente con la descripción del aparato utilizado en la práctica de este invento.

En la figura 4 se ilustra un diagrama de conjuntos del aparato para reconocer dígitos hablados según los principios del invento. En este aparato, una señal de conversación entrante que se ha de analizar y reconocer se alimenta al filtro 210 que es un filtro de paso bajo de diseño normal que tiene una banda de paso de 4kHz. A la acción del filtro 210 responde un muestreador y un convertidor analógico a digital 220 que muestrea la señal alimentada, la convierte en una formato digital y remite la señal convertida en segmentos de tiempo llamados cuadros para proceso adicional. El convertidor 220 se controla por el

elemento 200, el elemento de control, que da al convertidor 220 una cronometración de muestreo apropiada, (v.g., 10 kHz) y con cualesquiera otra señales requeridas por el convertidor A/D particular elegido. Se puede emplear cualquiera de un cierto número de convertidores A/D disponibles en mercado en el conjunto 210, v.g., el modelo 4130 de Teledyne Philbrick, Incorporated.

Un extractor 230 responde al convertidor 220 y dicho extracto comprende un detector de silencio 240, un detector de impulsión 250, un detector de fricativas 260 y un procesador de formantes 270. El detector de silencio 240, como su nombre indica, detecta la presencia de silencio en el cuadro probado. El detector de silencio 240 se puede poner en práctica rectificando e integrando la señal probada, del mismo modo que un receptor normal rectifica e integra señales recibidas, y comparando la señal integrada con un umbral fijo. Como variante, se puede emplear un detector de conversación para determinar la ausencia de conversación, como el elemento 24 de la patente EE.UU 3.723.667 concedida a Park et al 27 de Marzo de 1973. Según este invento, cuando se detecta un silencio, se genera un distintivo de silencio y se alimenta al aceptador 300. Es una decisión de si o no. El distintivo de silencio es una señal que tiene un formato predeterminado que puede ser, por ejemplo, una palabra binario de tres bitios con el valor 1_2 (001).

Una impulsión, que tiene lugar entre transiciones de fonema a fonema, se caracteriza por un aumento relativamente brusco de la energía en todo el espectro de la conversación Por lo tanto, para detectar una impulsión, ha sido necesaria la medida del régimen de energía del aumento en toda la banda. Esto se consigue en el detector de impulsión 250 dividiendo la banda de 4 kHz en una pluralidad de bandas auxiliares contíguas y mi-

diendo apropiadamente la energía en las bandas auxiliares. La energía se mide rectificando e integrando la energía en cada banda auxiliar limitando la energía en cada banda auxiliar a un nivel previamente elegido y sumando y diferenciado las salidas de energía limitada de las bandas auxiliares. Debido al proceso de limitación, un aumento grande en la energía de una banda auxiliar no puede producir una señal de suma diferenciada potente mientras que un aumento moderado brusco en toda la banda 4 kHz puede desarrollar una señal de suma diferenciada potente. Así, la señal de suma diferenciada puede servir convenientemente para indicar el régimen de aumento de energía en la banda general de 4 khz.

La puesta en práctica de un detector de impulsiones 250 es muy normal, puesto que las operaciones de proceso en el mismo son operaciones perfectamente conocidas y directas. Por ejemplo, el detector 250 puede contener un conjunto de filtros de paso de bandas contíguos que responden a la señal de conversación un rectificador, un integrador acoplado a un limitador umbral conectado al acceso de salida de cada uno de los filtros de paso de banda, y un adicionador seguido por un diferenciador que responde a cada uno de los limitadores umbrales. Alimentando la señal de salida del diferenciador a otros circuitos umbral se obtiene una señal de salida binaria que representa la presencia o ausencia de una impulsión. Como es lógico, cuando hay presente una impulsión, se genera un distintivo de impulsión.

Al igual que ocurre con el distintivo de silencio, el distintivo de impulsión se alimenta al aceptador 300. El distintivo de impulsión puede tener el mismo formato que el distintivo de silencio, v.g., una palabra bianria de tres bitios - pero con un valor diferente al del distintivo de silencio, v.g., $2_2(010)$. Se pueden encontrar diversos diseños de circuitos úti-

les para poner en práctica el detector 250 en la publicación de Millman and Taub, Pulse Digital and Switching Waveforms, McGraw Hill 1965.

El detector de fricativas 260 genera un distintivo siempre que el cuadro analizado contenga una consonante sonora, como los sonidos generados por las letras z y w en el idioma Inglés, o una consonante sorda como las letras s,f,t,k en el idioma Inglés. Las consonantes sonoras sordas se caracterizan por una concentración de alta frecuencia de energía a modo de ruido, mientras que las consonantes sonoras explosivas se caracterizan por un componente fuerte de energía a bajas frecuencias, v.g., a aproximadamente 500 kHz. T.R, Martin, en la disertación doctoral mencionada, describe elementos para reconocer la presencia de consonantes sonoras explosivas y sordas. Estos elementos se pueden emplear convenientemente en la puesta en práctica de este invento, para proporcionar una señal de salida que tiene un formato binario de bitios múltiples muy parecido al formato del distintivo de la letra explosiva. Por ejemplo, el distintivo de letra fricativa alimentado al aceptador 300 puede tener los valores $3_2(011)$ y $4_2(100)$ cuando se especifica una fricativa sonora y una fricativa sorda, respectivamente. En este memoria, los distintivos de silencio, explosivo y fricativo tienen cada uno un formato de tres bitios, pero los valores son diferentes. Las vías de tres bitios de los elementos 240,250 y 260 se pueden combinar por lo tanto, en una sola vía de tres bitios. Como es lógico, también son posibles otros formatos de señales.

El procesador de formantes 270 analiza las señales de cuadros y extrae de las mismas frecuencias de formantes. Las frecuencias de los formantes son componentes de frecuencia única pronunciada en el espectro de la conversación que están -

presentes de una forma más distintiva cuando se pronuncian sonidos correspondientes a vocales. Aunque la extracción de los formantes no es una tarea fácil, es básica en el arte del análisis y la síntesis de la conversación y, por lo tanto, la ha abarcado perfectamente la literatura. Las técnicas y aparatos útiles para poner en práctica el procesador de formantes 270 se describen entre otras publicaciones, en las siguientes:

1.- B.S. Atal y S.L. Hanauer, "Speech Analysis and Systhesis by Linear Prediction of the Speech Wave", JASA, Volúmen 50, página 637-655, 1971;

2.- Patente EE.UU 3.624.302, concedida a B.S. Atal el 30 de Noviembre de 1971;

3.- S.S. McCandless, "An Algorithm for Automatic Formant Extraction Using Linear Prediction Spectra", IEEE Transactions on Acoustics Speech and Signal Processing, volúmen ASSP 22 nº 2, páginas 135-141 Abril 1974.

4.- J.D, Markel, "Digital Inverse Filtering- A new Tool for Formant Trajectory Estimation, "IEEE Transactions Audio Electric Acoustics", volúmen Au 2, páginas 129-137, 1971.

5.- B.Gold y L.R. Rabiner, "Parallel Processing Techniques for Estimating Pitch Periods os Speech in the Time Domain",-JASA volúmen 46, 1969.

6.- Patente EE.UU 3.649.765, concedida a R.W. Shafer et al el 14 de Marzo de 1972; y

7.- L.R. Rabiner et al "A Hardware Realization of a Digital Formant Synthesizer," IEEE Trans.Comn.Techn, volúmen COM-19, páginas 1016-1020, Noviembre 1971.

Una vez que se han obtenido las frecuencias de los formantes, v.g., empleando los elementos descritos por Rabiner et al en artículo nº 7 arriba referenciado, el procesador de

transformaciones 280 convierten las frecuencias de los formantes obtenidas en posiciones del cuerpo de la lengua, y partiendo de posiciones sucesivas del cuerpo de la lengua el procesador 240 desarrolla los distintivos de las trayectorias del cuerpo de la lengua. El procesador de formantes 270 deriva una señal al procesador de transformación 280 que es representativa de los tres formantes de frecuencia menor hallandos en la señal de conversión Estos formantes se presentan de preferencia simultáneamente, en paralelo, comprendiendo un solo campo yustapuesto. De este modo, cuando cada formante está definido por un código o campo de 8 bitios, el campo de salida yustapuesto del procesador 270 es un campo de 24 bitios. La señal de salida del procesador 280 es un campo paralelo binario que representa el distintivo de posición del cuerpo de la lengua, y de un modo más particular, la región de la cavidad bucal según se define en la figura 3, y la dirección del movimiento del cuerpo de la lengua.

Según se ha indicado anteriormente, el desarrollo de la posición y trayectoria del cuerpo de la lengua se efectúa según el modelo de tracto vocal de Coker. Coker desarrolla los formantes resultantes de cada posición del cuerpo de la lengua. En este caso, se utiliza el modelo de Cocker a la inversa, para desarrollar una posición del cuerpo de la lengua partiendo de un conjunto de las tres frecuencias inferiores de los formantes.

Una descripción simplificada del modelo de Coker y del uso del modelo para desarrollar una posición del cuerpo de la lengua correspondiente a un conjunto de formante presentados se pueden encontrar en "Speech Analysis by Articulatory Synthesis", E.H. Hafer Masters Dissertation, Northwestern University Computer Sciences Department, Evanston, Illinois, Junio de

1974. Esta disertación está disponible para inspección y fotocopiado en la libreria de la Universidad de Northwestern. Las paginas 10-18 de la disertación anterior y los apéndices 1-4 son particularmente reveladores, el texto explica el modelo y el método de derivar los formantes apropiados a partir del modelo, y los apéndices 2-4 presentan los programas FORTRAN que se pueden emplear conjuntamente con un ordenador de uso general para desarrollar la información deseada. Como el procesador 280 puede comprender un ordenador para uso general que emplea los programas descritos en los apéndices citados, la disertación de Hafer se incorpora en la presente a título de referencia y formando parte de esta descripción. Así mismo, como los programas incorporadas unicamente en el certificado de prioridad por no poder ser traducidos son útiles para especificar la manufactura de tablas de averiguación de ROM que se describirán más adelante, los programas se añaden a un ápéndice a este descripción para comodidad de quellos que deban poner en práctica este invento.

Resumiendo brevemente el modelo y su uso, el modelo de tracto vocal es una representación paramétrica de un plano medio sagital del aparato articulatorio humano. Se emplean seis parámetros en el modelo para controlar la posición de tres articuladores (cuerpo de la lengua, punta de la lengua y labios). Estos articuladores, determinan el área en sección transversal a lo largo del tracto. Se consigue una aproximación de la función del área del tracto vocal mediante 36 secciones transversales uniformente separadas que están definidas en planos perpendiculares a la línea central de la cavidad bucal. Según se comprenderá estudiando la figura 1, el área de la sección transversal de la cavidad bucal varía con la posición del cuerpo de la lengua. Por lo tanto, determinando el área en sección transversal de la cavi

dad a partir de las frecuencias de los formantes, se puede determinar la posición del cuerpo de la lengua. En situaciones en las cuales el ordenador para uso general sea la modalidad preferible del procesador 280, los programas que se adjuntan en la prioridad se pueden emplear para determinar la posición del cuerpo de la lengua de la persona que habla. Los programas actúan de una manera interactiva. En primer lugar, se supone que el cuerpo de la lengua se encuentra en un estado previamente elegido, y se deriva un conjunto de formantes característicos de dicho estado. El estado supuesto es la última posición conocida del cuerpo de la lengua. Partiendo del estado supuesto del cuerpo de la lengua se compara el conjunto derivado de formantes con los formantes alimentados (desarrollados en el procesador 270), y se evalúa una función de error para determinar la diferencia entre los formantes derivados y los formantes de la persona que habla. Dicha función de error determina los cambios que se han de hacer en el estado del modelo del tracto vocal para reducir el valor de la función de error. El modelo se cambia, los formantes se calculan y se evalúan de nuevo la función de error. Una vez que se ha determinado que la función de error es suficientemente pequeña, se analiza la forma del modelo del tracto vocal para dar lo que se ha demostrado como una aproximación razonable de la posición del cuerpo de la lengua para la mayoría de las vocales.

En aquellas situaciones en las cuales el ordenador de uso general puede que no sea la forma preferible de enfocar el problema para poner en práctica el procesador de transformación 280, se puede realizar de un modo diferente para los fines de este invento calculando previamente, por los programas que se adjuntan en la prioridad, los conjuntos de formantes desarrollados por el modelo de Coker para todas las posiciones del

cuerpo de la lengua y longitudes del tracto vocal de interes y almacenado los formantes evaluados en una tabla de averiguación. Se puede emplear una memoria de lectura solamente como tabla de averiguación, y disponerse para que el campo de localizaciones indique la posición del cuerpo de la lengua y la longitud del tracto empleada por el modelo y contenido de cada lugar de la memoria indicará los formantes generados por el modelo en respuesta al estado de modelo caracterizado porque el campo de localizaciones. El empleo de dicha tabla de averiguacion es iterativo porque los formantes asociados con las posiciones elegidas del cuerpo de la lengua y longitudes del tracto se tendría que comparar con los formantes derivados por el procesador 270.

Una tabla de averiguación de ROM se construye preferiblemente comprendiendo los formantes la variable independiente en lugar de la variable dependiente. O sea, los tres formantes derivados por el modelo se yustaponen para formar un solo campo y dicho camp sirve como campo de localización a una memoria en la cual los lugares contienen las posiciones del cuerpo de la lengua y longitudes del tracto que corresponden a los formantes que comprenden las localizaciones correspondientes. Con dicha tabla de averiguación, no es necesaria una operación iterativa.

La señal de salida del procesador de transformación 280 es un dispositivo de la trayectoria del cuerpo de la lengua que comprende la posición del cuerpo a la lengua y una medida del movimiento de la lengua. La información de posición se obtiene, según se ha descrito, de la tabla de averiguación. La indicación de movimiento se deriva comparando la posición obtenida con la posición anterior. Se puede realizar almacenado las posiciones de las coordenadas x e y y restandolas de las po-

siciones de las coordenadas x e y y recien determinadas. Como solamente se necesita discriminar 10 regiones para obtener una indicación de posición suficiente (vease la figura 3), el formato del distintivo del cuerpo de la lengua puede ser una palabra binaria de 8 bitios, indicando los primeros cuatro bitios la posición de la lengua e indicando los dos bitios siguientes el movimiento en la dirección x, e indicando los dos últimos bitios el movimiento en la dirección y.

Resumiendo la modalidad preferible del procesador 280 según se ha descrito el modelo de Coker se emplea a la inversa para desarrollar una tabla de posiciones del cuerpo de la lengua que corresponden mejor en cada conjunto de tres formantes inferiores. La tabla se desarrolla del modo más sencillo utilizando el programa que se adjunta, en el certificado de prioridad puesto que los principios del modelo de Coker se incorporan en dicho programa. No obstante, el empleo de este programa no es obligatorio y, además, en lugar de la tabla de averiguación de memoria, se pueden utilizar medios para valorar el modelo de Coker a la inversa. En la modalidad descrita, se utiliza una tabla porque es la forma más sencilla de ejecución y dicha tabla está almacenada permanentemente en una memoria ROM. La memoria ROM se localiza por el única campo que está compuesto de los tres campos yustapuestos que definen los tres formantes de frecuencia inferiores. La señal de salida del procesador 280 es un campo de 8 bitios, indicando los primeros 4 bitios la posición de la lengua, indicando los dos bitios siguientes el movimiento horizontal del cuerpo de la lengua e indicando los dos últimos bitios el movimiento vertical del cuerpo de la lengua.

La señal de salida del procesador 280, como las señales de salida de los elementos 240, 250 y 260, se alimentan

al aceptador 300, estando indicadas las diversas vías por el número de referencia 302 en las figuras 4 y 6.

Si fuera cierto que las señales que corresponden solamente a dígitos válidos se alimentaran al sistema de reconocimiento de palabras de este invento, el aceptador 300 no tendría que ser una máquina muy compleja. El aceptador 300 tendría un estado inicial a partir del cual se ramificaría a una de las secuencias de distintivos que representan el dígito hablado y cuando se completa la detección del dígito v.g., se detecta la secuencia completa de distintivos, el aceptador 300 reintrodu cirfa el estado inicial, dispuesto para descodificar el dígito siguiente. Desgraciadamente, el aceptador 300 debe poder aceptar palabras, pronunciaciones y sonidos distintos a dígitos válidos sin desactivarse. Por consiguiente, el aceptador 300 debe poder suponer que cualquier distintivo es el comienzo de una secuencia de dígitos válidos y debe poder pasar a una iniciación de una nueva secuencia siempre que se desactiva. La exigencia del paso o retroseguimiento se podrá comprender mejor por el ejemplo siguiente donde las secuencias de distintivos 110, 011, 101, 111, 110 y 011, 101, 111, 1001 son secuencias válidas y donde las secuencias de distintivos 110, 011, 101, 1001. Cuando el aceptador 300 prosigue a través de los distintivos 110, 011, 101, 111 en la secuencia encontrada, supone que la secuencia 110, 011, 101, 111, 110, se detecta y, por lo tanto sigue en dicho trayecto. - Cuando se alcanza el disntintivo 1001, el aceptador 300 debe poder determinar que la secuencia 110, 011, 101, 111, 1001, no es una secuencia válida y que, por lo tanto, debe retroseguir o volver a una nueva iniciación de secuencia. Volviendo a seguir a partir del distintivo 1001 hasta el distintivo 001 (eliminando el primer distintivo 110) la secuencia 001, 101, 111, 1001 se -

detecta apropiadamente por el aceptador 300 como una secuencia válida.

Para desarrollar las operaciones requeridas, el aceptador 300 se construye como una máquina en secuencia de estados definidos que comienza en un estado inicial y prosigue a través de varias transiciones de estados hasta una de 10 conclusiones satisfactorias (detectando cada uno de los 10 dígitos). Dichas máquinas de secuencias, que a veces se llaman detectores de secuencia, son muy comunes. El diseño de dichas máquinas, para conseguir los diagramas de estados predeterminados se describe, por ejemplo, en Switching Theory por P.E. Wood, Jr. MaGraw-Hill Book Co., 1968, capítulo 5, y en el área de la conversión, dicha máquina de secuencias ha sido descrita por Martin en su disertación doctoral mencionada y se ha descrito, entre otras publicaciones, la patente EE.UU 3.700.815 concedida a Rowland et al, el 24 de Abril de 1972. Cualquier desviación de un trayecto aceptable conduce de nuevo al estado inicial. Esto se ilustra, para los fines de esta descripción, por el diagrama de estado de la figura 5 que describe las transiciones de estado necesarias para detectar la pronunciación en lengua inglesa "two eight" (dos ocho). El diagrama de estados completos del aceptador 300 depende, como es lógico de la lista exacta de palabras que se pretende detectar (dígitos 0-9, palabras de unión como "hundred", (ciento). El diagrama de estado de la figura 5 y los elementos para su ejecución, ilustrados en la figura 7, se consideran representativas. El estado 1 del aceptador 300, que representa en la figura 5 como número 1 dentro de un círculo, es el estado inicial del aceptador 300. Es el estado en el cual el aceptador 300 entra en acción siempre que se completa con éxito o sin exito una prueba. El aceptador 300 permanece en el estado 1 hasta

que se recibe un distintivo que corresponde al comienzo de cualquiera de las palabras reconocibles, v.g., dígitos. El rayo indicado como A en la figura 5 representa a los trayectos de salida desde el estado 1 en la dirección de los dígitos distintos a "two" y "eight" (dos y ocho).

Cuando se pronuncia en lengua inglesa el dígito "two" (dos), el sonido /t/ de "two" da por resultado un distintivo explosivo que hace que el aceptador 300 avance al estado 2. Esto está indicado en la figura 5 por el rayo marcado B (equivalente a "Burst" (explosivo)) se extiende desde el estado 1 hasta el estado 2. El aceptador 300 permanece en el estado 2 en tanto que se alimente un distintivo explosivo pero sale del estado 2 a través del rayo marcado * siempre que se alimenta un distintivo que no es consonante con la continuación de la pronunciación "two" (dos). Una salida marcada por * indica un retorno al estado 1 en un modo de operación de retroseguimiento. Cuando se pronuncia el dígito "two" (dos) sigue un segmento de vocal a la explosión de /t/. La parte inicial del segmento de la vocal representa un cuerpo de la lengua situado en la segunda parte divisoria de la figura 3. Por lo tanto, en respuesta a un distintivo que indica una zona divisoria dos de la posición del cuerpo de la lengua (p=2), el aceptador 300 avanza al estado 3 según se representa en la figura 5. El aceptador 300 permanece en estado 3 hasta que el cuerpo entra en la zona divisoria 6 y comienza moverse en la dirección x positiva . Cuando esto ocurre, se reconoce el dígito 2, indicado en la figura 5 por el rayo marcado D=2, y el aceptador se repone al estado 1 como medida preparativa para el dígito siguiente.

Según se ha indicado anteriormente, la segunda parte de la pronunciación "two" (dos) contiene un segmento de

vocal que produce un cuerpo de lengua situado en la zona divisoria 6 y que avanza en la dirección x positiva. Como no existe un dígito cuyo segmento inicial coloque el cuerpo de la lengua en la zona divisoria 6, el aceptador 300 permanece en su estado inicial durante la parte final de la pronunciación "two" (dos), hasta el comienzo de la pronunciación "eight" (ocho).

La pronunciación "eight" (ocho) en lengua Inglesa comienza con un segmento de vocal en la zona divisoria 8. Por lo tanto, cuando el cuerpo de la lengua se mueve en la zona divisoria 8, el aceptador 300 sale del estado 1 y entra en el estado 4. Continuando en las direcciones x e y positivas, el cuerpo de la lengua asciende en la región divisoria 3, en cuyo instante el aceptador 300 avanza al estado 5 donde permanece hasta que llega el distintivo explosivo de la pronunciación "eight" (ocho), en cuyo instante, se reconoce el dígito "eight" (ocho) y el aceptador se repone al estado 1, dispuesto para el dígito siguiente.

En la ejecución del aceptador 300, se han de considerar dos elementos principales: medios para proporcionar la capacidad de retroseguimiento y medios para poner en práctica el diagrama de estados del aceptador.

Para la capacidad de retroseguimiento, se necesita una memoria que almacene las secuencias de distintivos alimentadas por el aceptador 300. Esta memoria se debe organizar de modo que se puedan extraer datos antíguos y volverse a elaborar mientras se introducen nuevos datos. Dicho dispositivo se pone en práctica almacenando los distintivos alimentados en una memoria normal bajo control de un contador de localizaciones de disitintivos que funciona en una arítmetica de modulos iguâl o menor que el tamaño de la memoria (por ejemplo, con

un contador de localizaciones de 10 dígitos se emplea por lo menos una memoria de 1024 palabras). Con dicho dispositivo, se insertan disitintivos alimentados en secuencia en la memoria según determina el contador de localizaciones de distintivos y cuando, por ejemplo, se llena el lugar 1023 de la memoria (si se utiliza un contador de 10 bitios) el lugar siguiente de la memoria se ha de llenar (borrando la información antígua) es el lugar de la memoria 0.

Dos contadores más, que funcionan en el mismo módulo que el contador de localizaciones de distintivos, se incluyen para una utilización apropiada de la memoria: un contador de iniciación de secuencias (contador A) y un contador de localizaciones corrientes (contador B). El contador A indica el lugar del primer distintivo en la secuencia probada y el contador B indica la localización corriente del distintivo en la secuencia en que se ha probado. Un diagrama de conjuntos de este dispositivo se ilustra en la figura 6. En la figura 6, una memoria 301 almacena los distintivos alimentados al aceptador 300 en el conductor 302 y envía los distintivos previamente almacenados exigidos por el aceptador 300 por el conductor 317. La escritura y la lectura de la memoria 301 se realiza en respuesta a las ordenes de control de lectura y escritura proporcionadas por el elemento de contro 200 (figura 4) por los conductores 303 y 304. La localización apropiada se proporciona a la memoria 301 por el bloque de selección 305 que, a su vez, responde al contador 306 (contador de localizaciones de distintivos) y al contador 307 (contador B). El contador 308 (contador A) interacciona con el contador 307 por la línea de vía 309 y su interacción se mantiene bajo los conductores de control 310, 311, 312 y 313. Una señal en el conductor de control 310 hace avanzar el contador

308 una unidad, la señal en el conductor de control 311 duplica el valor del contador 307 en el contador 308, una señal en el conductor de control 312 ha-ce avanzar una unidad al contador 307, y una señal en el conductor de control 313 duplica el valor del contador 308 en el contador 307. El conductor 314 controla el contador 306, haciéndolo avanzar cada vez que se alimenta un nuevo distintivo.

En la práctica, cuando se inicia una prueba de secuencia, ambos contadores A y B localizan el mismo lugar, haciendo que l primer disitntivo de la secuencia probada se extraigs de la memoria 301. En tanto que laprueba prosiga satisfactoriamente, el contador 307 avanza de uno en uno mientras que el contador 308 permanece sin cambiar. Cuando la prueba termina con éxito al final de la secuencia, el contador 308 avanza a la posición del contador 307 y se inicia una nueva prueba. Cuando la prueba termina sin éxito (con una entrada de * al estado l), el contador 308 avanza una unidad y el contador 307 se coloca igual que el contador 308, iniciando de nuevo una nueva prueba.

Para poner en práctica el diagrama de estado del aceptador 300, se puede emplear técnicas tradicionales. No obstante, para poder completar la figura 7 ilustra una modalidad para poner en práctica la parte operativa del diagrama de estado representado en la figura 5.

Como solamente hay presentes cinco estados en la figura 5, la figura 7 representa cinco basculadores representativos de los estados (701-705). Cada basculador se conecta a un conjunto lógico asociado (711-715), y los conjuntos lógicos 711-715 responden todos a la vía de señal 317 que surge en la memoria 301 (figura 6).

Cada uno de los conjuntos lógicos 711-715 gene-

ra una señal de salida combinatoria diferente que se ha concebido en particular para ejecutar una parte del diagrama de estados. Por ejemplo, el conjunto lógico 711 desarrolla las señales de salida necesarias para sacar al aceptador 300 del estado 1 e introducirlo en los estados, 2, 4 o A. Por consiguiente, el conjunto 711 tiene tres salidas: una señalque dirige la entrada al estado A (conductor 721), una señal que dirige la entrada en el estado 4 (conductor 722) y una señal que dirige la entrada en el estado 2 (conductor 723). Según la figura 5, la entrada en el estado 4 ha de tener lugar solamente cuando tiene lugar b=8. Por lo tanto, la expresión boolean para la salida en el conductor 722 es (estado 1) (p=8). La primera variable (estado 1) se deriva del basculador 701 y la segunda variable, p=8, se deriva de una descodificación en la información en la vía 317. Así, se utiliza una puerta Y de dos entradas para generar la señal de salida del conductor 722. Las señales de salida de los elementos 711-715 se derivan de una manera análoga.

Según se ha indicado anteriormente, siempre que el diagrama de estados de la figura 5 indica una salida * , el aceptador 300 debe reintroducir el estado 1 y debe modificar de un modo particular los contadores 307 y 308. Con este fin, la puerta 0 731 recoge todas las salidas y las combina para formar una señal de salida en el conductor 732 que controla al contador 307 y 308. Las salidas D exigen también una reintroducción del estado 1 pero con una modificación diferente de los contadores 307 y 308 (según se ha descrito anteriormente). Con este fín, se emplea la puerta 0 733 para generar una señal de salida en el conductor 734. Las señales de control de salida y D se combinan en la puerta 0 735 que control la entrada en el estado 1.

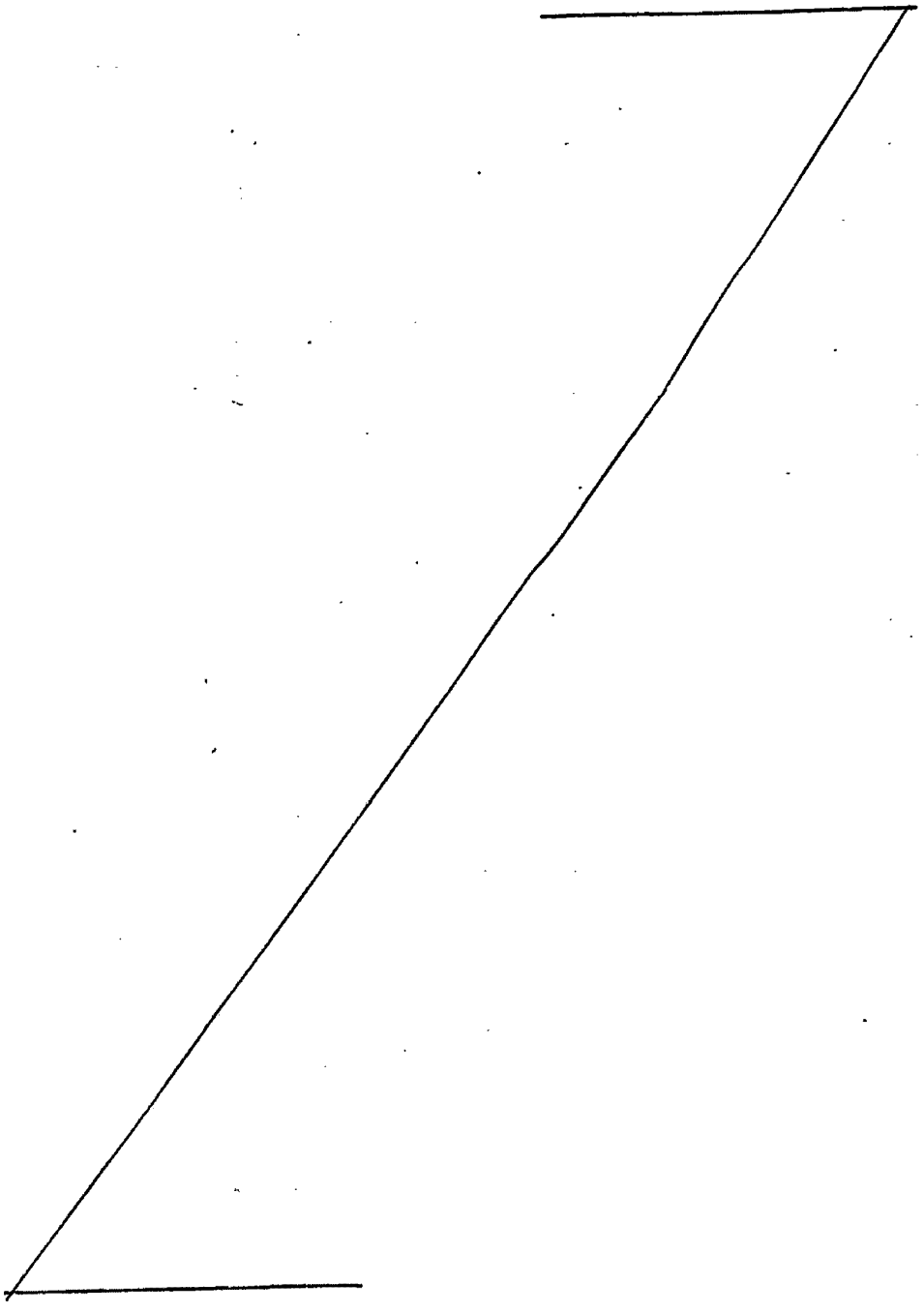
La entrada en cualquier estado particular, co-

mo es lógico, debe ir acompañada por la salida de todos los demás estados. Por lo tanto, cuando se colocan en posición inicial los basculadores 701-705, se deben reponer todos los demás basculadores. Esto se consigue en la figura 7 con la ayuda de los conjuntos lógicos 741-745 o con la puerta O 746. La puerta O 746 desarrolla una señal siempre que se produce cualquier transición de estado y dicha señal se alimenta a las entradas R de los conjuntos lógicos 741-745. Cada uno de los conjuntos lógicos 741-745 se disponen para que proporcione una señal de salida en el terminal Q cuando se alimenta una señal a la entrada R, y una señal de salida en el terminal $\overline{Q}$ cuando se alimenta una señal a la entrada R y S. De esta manera, los conjuntos 741-745 se combinan con la puerta 746 para reponer todos los basculadores excepto el basculador que se ha colocado en posición inicial.

El control del sistema de la figura 4 se consigue mediante el elemento de control 200. Proporciona el reloj de muestreo al convertidor A/D 220, las señales de control de lectura y escritura (conductores 303 y 304) a la memoria 301, las ordenes de colocación y avance (conductores 310-314) a los conductores 306, 307 y 308, y todas las demás señales de control necesarias para ell funcionamiento apropiado del extractor de características 230. El elemento 200 puede ser de construcción tradicional que comprende un multivibrador aestable para desarrollar una señal de cronometración básica, basculadores interconectados al multivibrador para desarrollar sus múltiplos de la señal de cronometración básica y varias puertas interconectadas para formar el circuito lógico combinatorio apropiado por cada señal de control requerida. De este modo, la circuitería necesaria es muy directa y los detalles de las interconexiones de puertas lógicas se dejan a la elección de los expertos en la materia

que pongan en práctica al invento.

Descrita suficientemente la naturaleza del invento, así como la manera de realizarlo en la práctica debe hacerse constar que las disposiciones anteriormente indicadas son susceptibles de modificaciones de detalle en cuanto no alteren su principio fundamental.



REIVINDICACIONES

1.- Sistema de reconocimiento de voz a partir de características representativas de sonidos hablados conocidos, caracterizado porque comprende: un extractor de características, que responde a la muestra de representación, para determinar características de voces contenidas en la muestra de representación, incluyendo una característica correspondiente a la posición del cuerpo de lengua y su dirección de movimiento: y un aceptor, que responde al extractor de característica, para aparear la secuencia de las características determinadas con características de secuencias predeterminadas que corresponden a palabras seleccionadas.

2.- Sistema según la reivindicación 1, caracterizado porque las características determinadas por el extractor de características comprenden un indicativo de silencio, un indicativo de impulsión, un indicativo fricativo, y un indicativo de la trayectoria del cuerpo de lengua.

3.- Sistema según la reivindicación 1, caracterizado porque el extractor de características comprende: un primer medio que responde a la muestra de representación para computar las características de silencio; un segundo medio que responde a la muestra de representación para computar características de impulsión; un tercer medio que responde a la muestra de representación para computar características fricativas; un cuarto medio que responde a la muestra de representación para computar frecuencias formadoras de las aplicaciones de voces; y un quinto medio que responde al cuarto medio para convertir las frecuencias formadoras a características de trayectoria del cuerpo de lengua.

4.- Sistema según la reivindicación 3, carac-

terizado porque el quinto medio emplea un modelo de trayecto vocal para convertir las frecuencias formantes a características de trayectoria del cuerpo de lengua.

5.- Sistema según la reivindicación 3, caracterizado porque el quinto medio comprende una tabla de memoria de búsqueda.

6.- Sistema según las reivindicaciones anteriores, caracterizado porque cuando incluye medios para desarrollar una secuencia de características que representan señales aplicadas y medios para aparear la sucesión de características a secuencias de características predeterminadas que representan palabras preseleccionadas, el medio para desarrollar una secuencia de características preseleccionadas determina los formantes contenidos en las señales aplicadas y convierte dichos formantes a características de trayectoria del cuerpo de lengua.

7.- Sistema según la reivindicación 6, caracterizado porque el medio de extracción convierte los formantes a las características de trayectoria del cuerpo de lengua de acuerdo con el modelo de trayecto vocal de Coker.

8.- Sistema de reconocimiento de voz a partir de características representativas desonidos hablados conocidos, tal y como queda sustancialmente descrito en la presente Memoria.

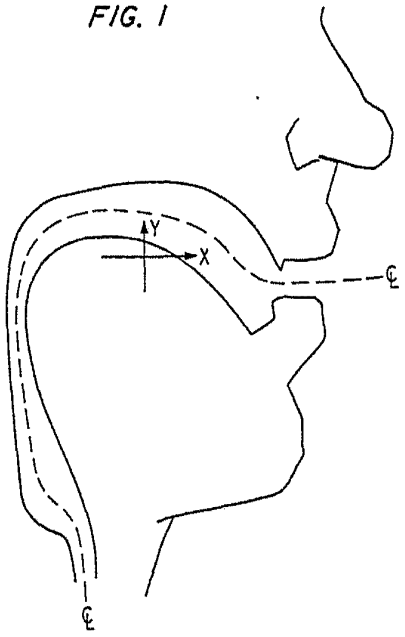
Esta Memoria consta de 27 hojas escritas a máquina por una sola cara.

Madrid, 16 ENE. 1979

WESTERN ELECTRIC COMPANY, INCORPORATED

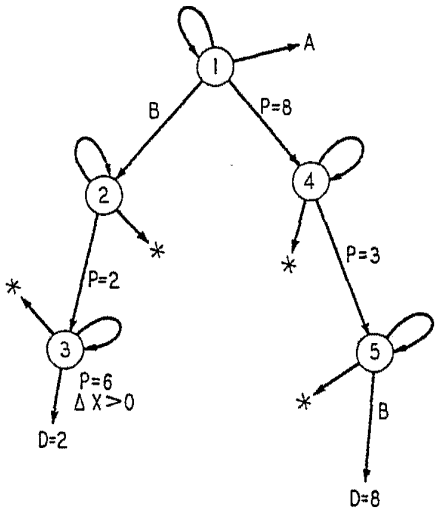
J. M. GOMEZ ACEBO Y POMBO
p. p. Firmado: J. Suarez Diaz

FIG. 1



ESCALA VARIABLE

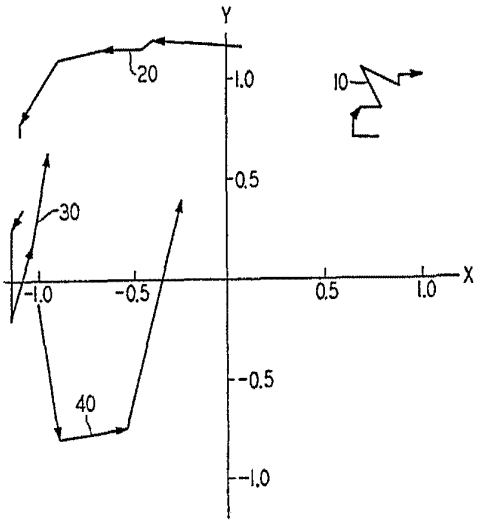
FIG. 5



Madrid 21 AGO. 1979

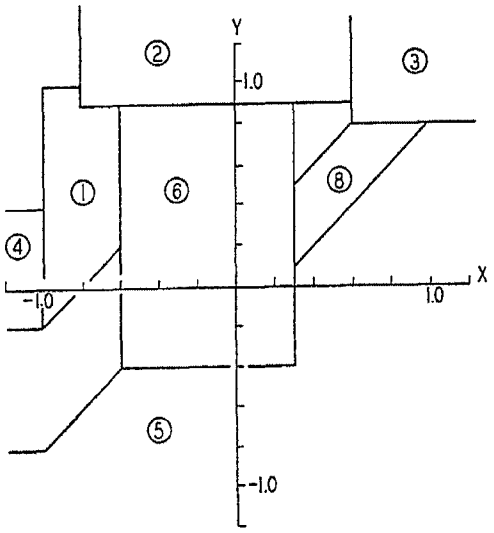
J. M. GOMEZ ACEBO Y POMBO
p. p. Firmado: Alejandro Carro López

FIG. 2



ESCALA VARIABLE

FIG. 3

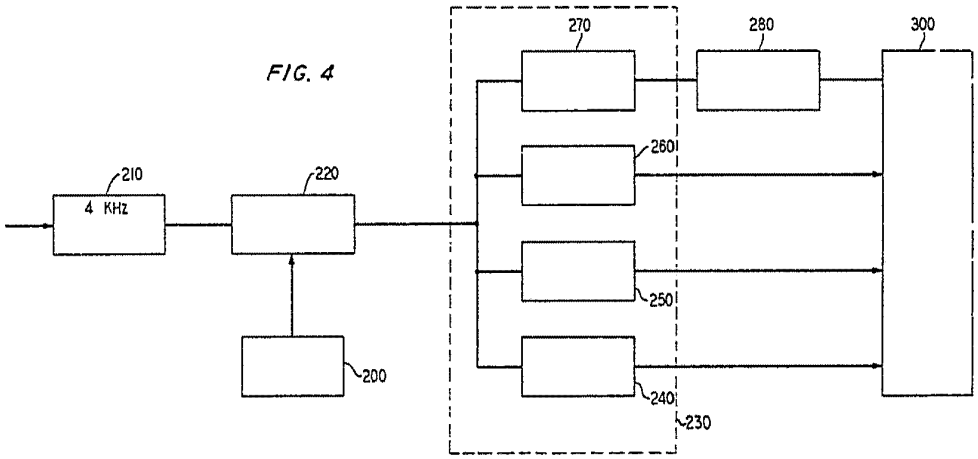


Madrid AGO. 1978

J. M. GOMEZ ACEBO Y POMBO
p. p. Firmado: Alejandro Calle López

ESCALA VARIABLE

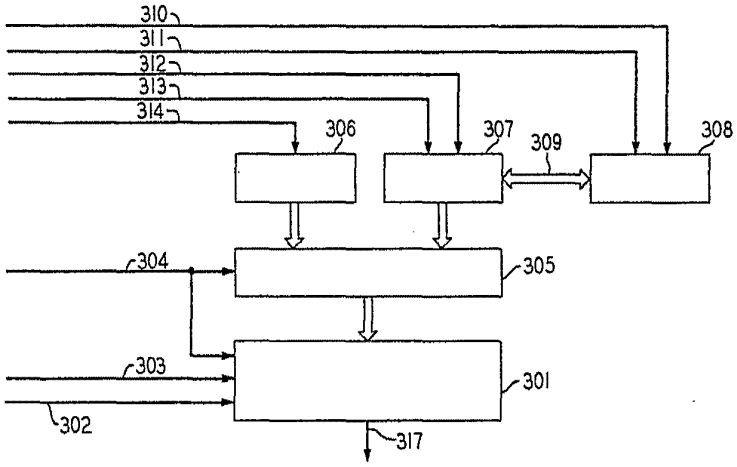
FIG. 4



Madrid 21 AGO. 1978

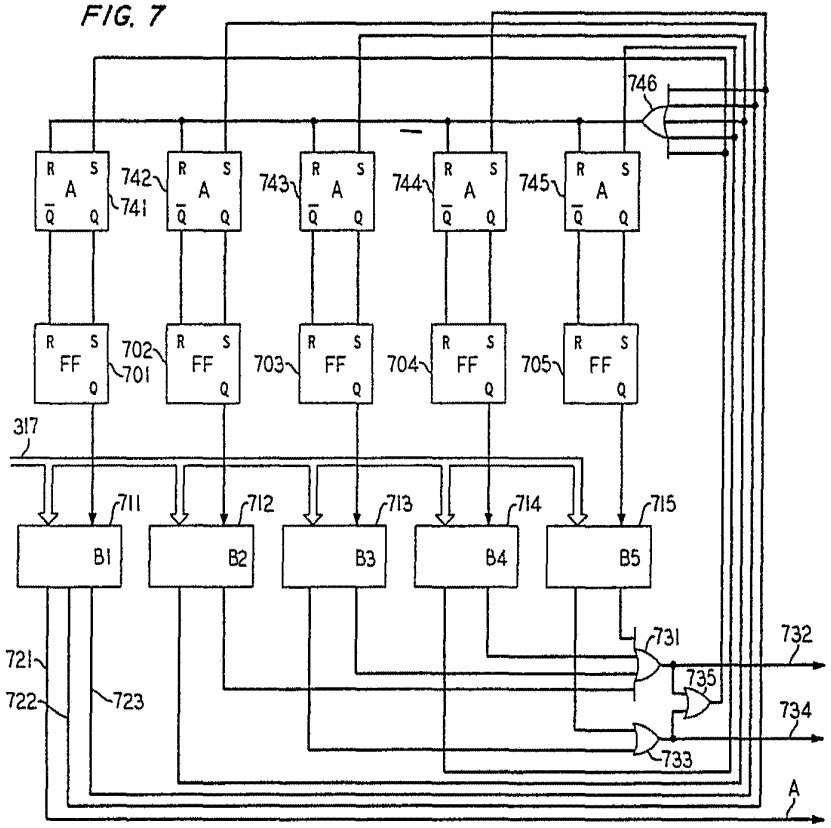
J. M. GOMEZ ACEBO Y POMBO
p. p. Firmado: Alejandro Cabello López

FIG. 6



ESCALA VARIABLE

FIG. 7



Madrid, 21 AGO. 1978

J. M. GOMEZ ACEBO Y COMBO

p. p. Firmado: Alejandro Calle López