

MINISTERIO DE INDUSTRIA  
REGISTRO DE LA PROPIEDAD INDUSTRIAL



⑩ ES	⑪ NUMERO	⑩ A 1
	⑪ 453.377	
	⑫ FECHA DE PRESENTACION	
	⑫ 16-11-1976	

PATENTE DE INVENCION

P.- 64.541  
File 22187  
Div. I

⑬ PRIORIDADES:	⑬ FECHA	⑬ PAIS
⑭ NUMERO		
477.080	6-6-74	E.U.A.

⑮ FECHA DE PUBLICIDAD	⑮ CLASIFICACION INTERNACIONAL	⑮ PATENTE DE LA QUE ES DIVISIONARIA
	G 0 9 B	436.945

⑯ TITULO DE LA INVENCION
"UN METODO DE TRATAR INFORMACION"

⑰ SOLICITANTE (S)
NESTOR ASSOCIATES

DOMICILIO DEL SOLICITANTE
Suite 6100, 405 Lexington Avenue, Nueva York, N.Y. 10017, Estados Unidos de América

⑲ INVENTOR (ES)
Leon N. Cooper y Charles Elbaum

⑲ TITULAR (ES)

⑲ REPRESENTANTE
DON OSCAR DE ELZABURU FERNANDEZ

La presente invención se refiere a sistemas adaptables de tratamiento de información, sistemas que son también conocidos como "máquinas de aprender", "redes de neuronas", sistemas "adiestrables", dispositivos autoorganizados y/o sistemas o dispositivos adaptables de memoria.

Durante los últimos años se han venido explorando extensamente los sistemas adaptables de tratamiento de información. Entre algunos de los sistemas más notables se incluyen los de Adaline y Madaline del Laboratorio Electrónico de Stanford, el Perceptron de los Laboratorios Aeronáuticos de Cornell y los Minos I y II del Instituto de Investigación de Stanford. Algunas de las patentes de EE.UU. que se refieren a los sistemas adaptables de tratamiento de información son: la patente núm. 3.287.649 de Rosenblatt; la patente núm. 3.408.627 de Kettler y col.; la patente núm. 3.435.422 de Gerhardt y col.; la patente número 3.533.072 de Clapper; y la patente núm. 3.601.811 de Yoshino. Esta lista de referencias es meramente ilustrativa y constituye sólo una pequeña parte del gran cuerpo de la técnica ya conocida, existente hasta hoy.

Tales sistemas adaptables de tratamiento de información, de la técnica ya conocida, funcionan en general produciendo una respuesta de salida para una señal dada de entrada, respuesta que se mide comparándola con algunas respuestas de salida prefijadas ("correctas"). A estos sistemas

de la técnica ya conocida se les hace que se modifiquen por sí solos, esto es, que "aprendan", a menudo dependiendo ello de la diferencia entre la respuesta de salida real y efectiva y la prefijada, hasta que se consigue la respuesta de salida prefijada. El objeto de un sistema como ésta es el de hacer que el sistema llegue a hallar por su propio camino (mediante algún algoritmo) una relación prefijada:

señal de entrada  $\rightarrow$  respuesta de salida

Hay que hacer notar aquí que siempre que en este estudio se emplee el término "señal de entrada", se tiene la intención de incluir en él la posibilidad de un juego de señales de entrada por separado que se aplican, esencialmente de manera simultánea, a un juego correspondiente de terminales de entrada de un sistema de tratamiento de información. De igual manera, con el término "respuesta de salida" se quiere definir la respuesta entera del sistema a una señal dada de entrada, aun cuando esta respuesta puede comprender una pluralidad de respuestas de salida individuales que aparezcan esencialmente de modo simultáneo en un juego de terminales de salida del sistema.

Un sistema adaptable tipo de la técnica ya conocida es el que se ilustra en la fig. 1 de los dibujos adjuntos. Este sistema comprende, como elementos esenciales, una red de entradas 1, 2, 3 ..., N, respectivamente conectadas a una pluralidad de elementos ponderadores variables  $G_1, G_2,$

$G_3 \dots, G_N$  que tienen un "peso" variable, el cual, por ejemplo, puede ser una ganancia variable en el caso de amplificadores ponderadores o una resistencia variable en el caso de elementos resistivos variables. Las salidas de los elementos ponderadores  $G$  se aplican a un sumador  $S$  que produce una sola salida de circuito o red, en proporción con la suma de las salidas de los elementos ponderadores. El valor de ponderación o "peso" de cada elemento ponderador  $G_1, G_2, G_3 \dots, G_N$  se controla individualmente por medio del denominado "algoritmo de adiestramiento"  $T$  que condiciona al circuito o red haciéndolo responder a una señal de entrada particular con una respuesta de salida deseada.

En el funcionamiento del circuito o red, a las entradas de éste 1, 2, 3 ...,  $N$  se les aplica repetitivamente una señal particular. Tras cada aplicación de la señal específica de entrada, la respuesta de salida del circuito o red se compara con una respuesta de salida prefijada: por ejemplo, por medio de un restador  $D$ ; y la diferencia o error se utiliza en el algoritmo de adiestramiento para modificar el "peso" o acción ponderal de los elementos ponderadores individuales  $G_1, G_2, G_3 \dots, G_N$ .

Cada aplicación de la señal específica de entrada, y la consiguiente modificación de los elementos ponderadores  $G$ , se denomina "ciclo de adiestramiento". A medida que

se producen sucesivos ciclos de adiestramiento, la respuesta de salida del circuito o red se va aproximando cada vez más a la respuesta de salida deseada, hasta que el circuito se condiciona para responder únicamente a la señal particular de entrada que haya de dar la respuesta de salida deseada.

En los sistemas adaptables de tratamiento de información de la técnica ya conocida, se viene insistiendo en hallar un algoritmo de adiestramiento adecuado que permita al sistema "aprender" o adaptarse deprisa a las señales de entrada aplicadas. No hace falta decir que se han ideado numerosos algoritmos a cual más ingenioso; ahora bien, en todos los casos el algoritmo de adiestramiento se viene haciendo, de alguna manera, dependiente, de la salida prefijada deseada que se va a generar en respuesta a una entrada dada.

Es objeto de la presente invención un sistema adaptable de tratamiento de información que tiene la aptitud de construir su propia respuesta distintiva de salida para cualquier señal de entrada dada. En particular, es objeto de la presente invención realizar un sistema con la sorprendente característica de que puede modificarse por sí solo construyendo una transformación o representación interna --- de señal de entrada --- señal de salida --- que funciona como memoria o programa sin intervención o selección alguna exte-

rior en cuanto a cuál es la respuesta de salida deseada o  
cuál es el diseño o pauta de entrada que se presenta. Este  
tipo de procedimiento de "aprendizaje" o automodificación  
del sistema adaptable de tratamiento de información se de-  
5 nominará en lo que sigue "adiestramiento pasivo" o "modi-  
ficación pasiva".

La importancia de esta aptitud de un sistema pa-  
ra modificarse por sí mismo pasivamente se apreciará con-  
siderando un sencillo ejemplo. Por no ser necesario con-  
10 tal sistema conocer de antemano una respuesta de salida  
prefijada y deseada para una señal dada de entrada, es po-  
sible aplicar al sistema señales de entrada que tengan un  
contenido desconocido para éste y, al cabo de un período  
de "adiestramiento", determinar el contenido de información  
15 de las señales de entrada mediante consideración de las  
respuestas de salida. Por ejemplo, si ocurriese que las se-  
ñales de entrada desconocidas fuesen unas señales de infor-  
mación (con alguna estructura desconocida) sepultadas o  
enmascaradas en ruido, como la estructura de las respues-  
20 tas de salida es isomórfica respecto a la de las señales  
de información sepultadas, la estructura desconocida se  
transformará en respuestas de salida y estará representa-  
da por éstas. De esta manera puede ser descifrado por el  
sistema de tratamiento de información el contenido infor-  
25 mativo desconocido de unas señales de entrada cualesquiera.

Es asimismo objeto del presente invento un sistema adaptable de tratamiento de información que, al igual que los sistemas de la técnica ya conocida, es capaz de producir una respuesta de salida deseada, prefijada, a cualquier señal de entrada dada. Este procedimiento, al que en lo sucesivo se denominará aquí de "aprendizaje activo" o "modificación activa", requiere por parte del operador humano un conocimiento de la respuesta de salida deseada que se vaya a asociar a cada señal de entrada individual.

Otro objeto de la presente invención reside en un sistema adaptable de tratamiento de información, en el que el índice de crecimiento o desarrollo del aprendizaje -- es decir, la velocidad a la que el sistema se adiestra por sí solo hasta producir una respuesta particular de salida en función del número de presentaciones de una señal de entrada-- es muy rápida. En particular, es objeto de la presente invención un sistema de tratamiento de información que tiene una tasa o índice de desarrollo del aprendizaje de tipo exponencial, en lugar de lineal u otra más lenta.

Otro objeto de la presente invención reside en un sistema adaptable de tratamiento de información capaz de funcionar como memoria distribuida y altamente invulnerable ante el mal funcionamiento de los componentes individuales. Tal memoria será una memoria adaptable y autoorganizada (que se organiza por sí sola), que tiene la aptitud de adquirir

información únicamente como resultado de la experiencia.

Además, esta memoria distribuida tendrá en general la capacidad, fiabilidad y exactitud de una memoria numérica usual de calculadoras u ordenadores (tal como una memoria de núcleos de ferrita) del tipo que almacena o guarda información en un lugar de emplazamiento local.

Otro objeto de la presente invención reside en un sistema adaptable de tratamiento de información capaz de una gran densidad de almacenaje. Por ejemplo, es de notar que el sistema de tratamiento de información es susceptible de ser realizado por medio de circuitos integrados y no requiere elementos discretos o por separado, tales como núcleos de ferrita.

Es objeto adicional del presente invento el de realizar un sistema adaptable de tratamiento de información capaz de una gran rapidez de funcionamiento; más en particular, un sistema en el que es posible recordar o reclamar y/o tratar, en una sola operación electrónica, del orden de  $2^n$  o más bits de información (siendo  $n$  el número de terminales de salida del sistema).

Finalmente, y lo que es quizá de más importancia, es objeto de la presente invención un sistema adaptable de tratamiento de información capaz de presentar cada una de las siguientes propiedades:

1) De reconocimiento: es la aptitud de producir

una fuerte respuesta de salida ante un suceso o señal de entrada que el sistema haya visto con anterioridad. Como es obvio, el sistema de tratamiento de información responderá inicialmente de modo difuso a una señal particular de entrada. Ahora bien, al cabo de sucesivas presentaciones de esa señal de entrada, el sistema aprenderá a "reconocer" la señal de entrada, dando una respuesta de salida característica.

2) De recordación: es la aptitud de producir una respuesta única de salida para cada una de cierto número de señales de entrada particulares. Esta característica desempeña la función de memoria, puesto que el sistema es capaz, merced a ella, de producir una respuesta única de salida en sus ( $n$ ) terminales de salida (conteniendo del orden de  $2^n$  o más bitios de información) al recibir una señal particular de entrada en su juego de terminales de entrada.

3) De generalización: es la aptitud de extraer un elemento común de entre cierto número de sucesos o señales de entrada diferentes. En particular, si se aplica cierto número de señales de entrada diferentes sucesivamente al sistema de tratamiento de información, el sistema aprenderá a reconocer una característica que sea común a estas señales de entrada. Por ejemplo, si a los terminales de entrada del sistema se aplica repetidamente una señal particular infor-

mativa sepultada en ruido, el sistema extraerá, retendrá y reconocerá a continuación la señal informativa.

5 4) De asociación: es la aptitud de recordar o asociar una primera señal de entrada al recibirse una segunda señal, después de haber sido aplicadas las dos señales de entrada al sistema de tratamiento de información más o menos concurrentemente. Esto es, al aplicarse dos señales de entrada simultáneamente, el sistema no sólo aprenderá estas señales de entrada sino que las "asociará" 10 entre sí. Así, en el futuro, el sistema será capaz de recordar una u otra de las señales de entrada, o ambas, si se aplica una sola de las señales de entrada. Esta característica de asociación puede ser efectiva, por ejemplo, en el análisis de señales desconocidas. Si se aplican al sistema 15 dos señales de entrada desconocidas, el sistema será capaz de determinar si una de ellas está relacionada con la otra de alguna manera.

20 5) De recuperación de entradas parciales (fragmentarias): es la aptitud para recuperar una señal entera de entrada a partir de una porción de dicha señal de entrada. Esta característica puede considerarse como de "autoasociación"; esto es, de "asociación" entre partes de la misma señal. Si se aplica al sistema una señal particular de entrada antes de ser "aprendida", el sistema "asociará" una 25 porción cualquiera de esta señal con la señal entera de mo-

do que, posteriormente, la aplicación de una porción de la señal de entrada traiga como resultado la producción, por parte del sistema, de la señal entera (usualmente con una relación de señal a ruido reducida).

5

En los dibujos adjuntos:

- la figura 1 es un esquema de bloques o funcional de un circuito o red de tratamiento de información que sirve de tipo de la técnica ya conocida;

10

- la figura 2 es un esquema funcional de una red de "nouverón" conforme al presente invento;

15

- la figura 3 es un esquema funcional de un módulo de Nestor<sup>TM</sup> conforme a la presente invención, en el que se emplean numerosas redes de "nouverón" del tipo ilustrado en la fig. 2; para mayor claridad se han omitido las líneas de retroacción que vienen de los sumadores;

20

- la figura 4 es un esquema funcional de un sistema de tratamiento de información que lleva incorporado un módulo de Nestor del tipo ilustrado en la fig. 3;

- la figura 5 es un esquema representativo que ilustra la respuesta de un módulo de Nestor conforme al presente invento, ante un tejido o trama de sucesos exterior;

- la figura 6 es un esquema representativo que ilustra la circulación de señales en un módulo ideal de Nestor conforme a la presente invención;

25

- la figura 7 es un esquema representativo que ilus-

tra un principio particular de funcionamiento en un módulo de Nestor conforme a la presente invención;

5 - la figura 8 es un esquema representativo de un sistema óptico-auditivo que lleva incorporada una pluralidad de módulos de Nestor conforme a la presente invención;

- la figura 9 es un esquema funcional o de bloques que ilustra un aparato determinante de respuestas de salida que puede ser utilizado con un módulo de Nestor conforme a la presente invención;

10 - la figura 10 es un esquema representativo que ilustra la respuesta de dos módulos de Nestor ante el mismo tejido de sucesos exterior;

15 - la figura 11 es un esquema funcional que ilustra un aparato que puede ser empleado con un módulo de Nestor conforme a la presente invención para lograr un tipo específico de respuesta de salida;

20 - la figura 12 es un esquema funcional que ilustra una porción de una red de "nouverón" que lleva incorporado un mnemondedor con arreglo a una forma preferida de realización del presente invento;

- la figura 13 es un esquema de principio de un circuito sumador que puede emplearse en la red de "nouverón" de la fig. 12;

25 - la figura 14 es un esquema de principio de un mnemondedor que puede emplearse en la red de "nouverón" de

la fig. 12; y

- la figura 15 es un esquema funcional de un aparato, que puede emplearse con la red de "nouverón" de la fig. 12, para tratar señales de entrada de dicha red por medio de la salida de la red.

Los diversos objetos de la presente invención, arriba expuestos, pueden lograrse disponiendo un módulo de tratamiento de información que tiene una pluralidad (N) de terminales de entrada 1, 2 ...,  $j$  ..., N, destinados a recibir N señales de entrada  $s_1, s_2 \dots, s_j \dots, s_N$ , respectivamente; una pluralidad ( $\underline{n}$ ) de señales de salida 1, 2 ...,  $i$  ...,  $\underline{n}$ , destinadas a presentar  $\underline{n}$  respuestas de salida  $r_1, r_2 \dots, r_i \dots, r_{\underline{n}}$ , respectivamente; y una pluralidad de elementos de unión, denominados "mnemondedores", que acoplan varios de los terminales de entrada (o multitud de ellos) con varios de los terminales de salida (o multitud de éstos). Con arreglo a la invención, cada mnemondedor efectúa una transferencia de información desde un terminal de entrada  $j$  a un terminal de salida  $i$  según la señal  $s_j$  que aparezca en el terminal de entrada  $j$  y según la denominada "función de transferencia"  $A_{ij}$  del mnemondedor. Se prevén medios para modificar la matriz de funciones de transferencia de los mnemondedores, dependiendo ello del producto de por lo menos una de las señales de entrada y una de las respuestas de salida del módulo.

El término "función de transferencia", tal como aparece en toda la Memoria descriptiva y sus reivindicaciones, se usa en su sentido más lato para definir simplemente una función que modifica de alguna manera la transferencia de información (una señal) desde la entrada a la salida de un mnemodador. En el caso más sencillo, la función de transferencia es equivalente a la ganancia o factor de amplificación de un mnemodador; ahora bien, como se apreciará, la señal que aparezca a la entrada de un mnemodador puede ser modificada de otras maneras para producir una señal de salida de mnemodador adecuada. En todo caso, aquí se especificará la señal de salida  $s_j'$  del mnemodador como producto de sus señal de entrada  $s_j$  y la función de transferencia  $A_{ij}$  del mnemodador, así:

$$s_j' = A_{ij} s_j.$$

En la figura 3 se ilustra un módulo del tipo arriba descrito, al que en lo sucesivo se denominará aquí módulo de Nestor <sup>TM</sup>. La fig. 3 representa un módulo particular en el que cada uno de los  $N$  terminales de entrada va conectado a cada uno de los  $n$  terminales de salida por medio de un solo mnemodador ( $\oplus$ ). Para mayor claridad, se han omitido las líneas de retroacción o realimentación que vienen de los sumadores, y sólo se representan los mnemodadores acoplados

al terminal de entrada  $i$  (mnemondadores  $1j, 2j \dots, ij \dots, nj$ ) y al terminal de salida (mnemondadores  $i1, i2 \dots, ij \dots, iN$ ). Ahora bien, se sobreentiende que en este módulo va dispuesta una formación regular de  $N \cdot n$  mnemondadores de modo que la matriz de funciones de transferencia  $A_{ij}$  será:

$$\begin{bmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ A_{1N} & A_{2N} & \dots & A_{nN} \end{bmatrix}$$

El particular módulo de Nestor, en el que cada uno de los  $N$  terminales de entrada va conectado a cada uno de los  $n$  terminales de salida por un solo mnemondador, presenta lo que se denomina "conectividad  $(N, n)$ ". En la práctica real puede seccionarse una o más de las conexiones de mnemondador entre los terminales de entrada y de salida, sin degradación perjudicial de la función del módulo. De hecho, las conexiones entre los terminales de entrada y los terminales de salida realizadas por medio de los mnemondadores pueden ser enteramente aleatorias, siempre y cuando haya presente un número suficiente de conexiones (mnemondadores) para proporcionar o realizar la transmisión de información desde la entrada a la salida y para el almacenaje de la información en los mnemo-

dedores.

En una forma preferida de realización del módulo de Nestor con arreglo al presente invento, la respuesta  $r_i$  presente en cada terminal de salida es función lineal de las entradas aplicadas a aquél, de modo que se tiene:

$$r_i = \sum_{j=1}^N A_{ij} s_j .$$

Esta relación lineal, que está indicada en la fig. 3 por medio de los sumadores  $\Sigma_1, \Sigma_2 \dots, \Sigma_i \dots, \Sigma_n$ , es la relación más sencilla que da los resultados deseados. Como se comprenderá, no obstante, el módulo de Nestor no se limita a esta relación lineal y pueden obtenerse otros resultados poderosos haciendo que la respuesta de salida  $r_i$  en uno o más terminales de salida  $i$  sea función distinta de las entradas suministradas a aquél. Por ejemplo, la respuesta de salida  $r_i$  puede hacerse proporcional al producto de las entradas  $A_{ij} s_j$ , en todas las  $j$ .

En el módulo de Nestor ilustrado en la fig. 3, el número de terminales de entrada  $N$  y el de terminales de salida  $n$  pueden adoptar respectivamente un valor cualquiera. El número de terminales de entrada puede exceder del número de terminales de salida, o viceversa, o bien ambos pueden ser iguales ( $N=n$ ). Se sobreentiende que la cantidad de información que puede aparecer en un momento dado cualquiera en los ter-

minales de salida del módulo de Nestor es del orden de  $2^n$  bitios o mayor y que, además, cuanto más alto sea el valor de  $n$  mayor será la relación de señal a ruido del módulo. Por lo tanto, es conveniente hacer que  $n$  sea prudencialmente grande, teniendo en cuenta la necesidad de un aumento correspondiente en el número de elementos de circuito. Así, por conveniencia, en un estudio ulterior se supondrá  $n$  igual a  $N$ .

El módulo de Nestor representado en la fig. 3 puede estar representado por, o considerarse compuesto de, una pluralidad ( $n$ ) de sumadores  $\Sigma_i$  que llevan cada uno asociada una pluralidad ( $N$ ) de terminales de entrada y mnemondadores y un solo terminal de salida  $i$ . Este subcomponente del módulo de Nestor, al que aquí se denominará en lo sucesivo "nouverón", se ilustra en la fig. 2. Comparando la fig. 2 con la fig. 3 puede advertirse que el módulo de Nestor comprende una pluralidad ( $n$ ) de nouverones del tipo ilustrado en la fig. 2.

Como se indica en la fig. 2, cada nouverón produce una sola respuesta de salida  $r_i$  en un terminal de salida  $r_i$ . Esta respuesta de salida es engendrada por el sumador de nouverón  $\Sigma_i$  según sean las salidas  $s'_1, s'_2, \dots, s'_j, \dots, s'_N$  de los  $N$  mnemondadores  $i_1, i_2, \dots, i_j, \dots, i_N$ , respectivamente.

Como se ha dicho más arriba, no es necesario que los mnemondadores del módulo de Nestor acoplen todos y cada uno de los terminales de entrada a todos y cada uno de los ter-

minales de salida del módulo. Por consiguiente, el nouverón  
ilustrado en la fig. 2 puede comprender menos de N mnemondedores,  
de manera que no todos los terminales de entrada 1, 2 ..., j ..., N  
estarán acoplados al sumador  $\Sigma_i$ .

5 También como se ha dicho antes, en el modo de "aprender" la función de transferencia  $A_{ij}$  de por lo menos uno (y preferiblemente de la totalidad) de los mnemondedores  $ij$  del módulo de Nestor se modifica según el producto de por lo menos una de las señales de entrada y una de las respuestas de salida del módulo. Este algoritmo para las modificaciones de A (aparte del decrecimiento o daclinación uniforme que no tiene relación con las entradas y las salidas) puede escribirse así:

$$\delta A_{ij} = f ( s_1, s_2 \dots, s_j \dots, s_N ; r_1, r_2 \dots, r_i \dots, r_n ) ;$$

15 expresión en la que se sobrentiende que la función  $f$  puede depender de sólo una de las señales de entrada  $s_1, s_2 \dots, s_j \dots, s_N$  y sólo una de las respuestas de salida  $r_1, r_2 \dots, r_i \dots, r_n$ . Con el fin de evitar toda complicación innecesaria en el conexionado eléctrico del módulo de Nestor, es preferible efectuar las modificaciones de la función de transferencia  $A_{ij}$  de un mnemondedor de modo que dependan sólo de las señales de entrada y la respuesta de salida del

5 nouverón del cual forma parte dicho mnemondedor. Por lo tanto, en una forma preferida de realización del presente invento, se hace que las modificaciones de  $A_{ij}$  dependan de la señal  $s_j$  y de la respuesta de salida  $r_i$  asociadas al particular mnemondedor; esto es:

$$\delta A_{ij} = f(s_j, r_i),$$

expresión en la que se sobrentiende que la función  $f$  depende siempre de ambas variables  $s_j$  y  $r_i$ .

10 La función  $f(s_j, r_i)$  puede desarrollarse en serie usando la fórmula de Taylor para funciones de varias variables. Haciendo esto, se obtiene:

$$15 \begin{aligned} f(s_j, r_i) = & a_{00} + a_{01} s_j + a_{10} r_i + a_{11} r_i s_j + \\ & + a_{21} r_i^2 s_j + a_{31} r_i^3 s_j \dots + a_{mn} r_i^n s_j^n. \end{aligned}$$

20 Los tres primeros términos de la serie de Taylor no son de interés inmediato para este caso, puesto que se requiere que las modificaciones de  $A_{ij}$  dependan del producto de la señal de entrada  $s_j$  y la respuesta de salida  $r_i$ . Como puede demostrarse, una modificación de  $A_{ij}$  que sea función de uno o más de los tres primeros términos, solamente, de la serie de Taylor no da por resultado un sistema de tratamiento de información que tenga las propiedades de inteligencia exhibidas por el módulo de Nestor.

25

El término de orden más bajo de la serie de Taylor que da realmente por resultado las propiedades deseadas con arreglo a la presente invención es el cuarto término: a saber, el  $a_{11} r_i s_j$ . Este término (al que, por poderse llevar fácilmente a un tratamiento analítico riguroso, se tratará a fondo en el estudio que sigue) da una modificación para  $A_{ij}$  de la forma siguiente:

$$\delta A_{ij} = \eta r_i s_j,$$

donde  $\eta$  es la constante de proporcionalidad.

Ahora bien, se sobretiende que otros términos de la serie de Taylor superiores al tercero pueden también producir resultados poderosos cuando se empleen en el módulo de Nestor. Los términos de potencia par de las variables,  $s_j$  o  $r_i$ , no dan discriminación de polaridad en las modificaciones. Los términos que contienen potencias impares de estas variables, tales como el sexto término  $a_{31} r_i^3 s_j$ , sí que dan esta discriminación. Unos y otros son capaces de comunicar propiedades interesantes al módulo de Nestor. En particular, como los diversos términos de la serie de Taylor proporcionan diferentes "pesos" o índices de ponderación en las modificaciones de A, estos "pesos" pueden usarse con ventaja para obtener propiedades específicas deseadas.

La fig. 4 ilustra una de las maneras en que un

módulo de Nestor conforme a la presente invención puede ser conectado y utilizado en un sistema adaptable de tratamiento de información. El módulo de Nestor representado en esta figura presenta una conectividad  $(N, N)$ ; esto es, el módulo comprende  $N$  terminales de entrada y  $N$  terminales de salida, y cada uno de los terminales de entrada está conectado a cada uno de los terminales de salida por medio de un mnemodador como se representa en la fig. 3.

Las señales de entrada  $s_1, s_2 \dots, s_j \dots, s_N$  que llegan al módulo de Nestor caracterizan un "suceso" en el ambiente designado como "entrada" en la fig. 4. Este suceso puede ser un suceso óptico, tal como la visión de un diseño o pauta; un suceso auditivo, tal como la audición de un tono; o cualquier otro suceso imaginable o inimaginable como, por ejemplo, la recepción de señales de radiación procedentes del espacio exterior. El único requisito para este suceso es el de que sea traducible de alguna manera en una pluralidad de señales de entrada  $s_1, s_2 \dots, s_j \dots, s_N$  que retengan acerca del suceso un detalle suficiente para que sea de interés. Las señales  $s_1, s_2 \dots, s_j \dots, s_N$  son generadas por un "traductor" que efectúa algún tipo de análisis del suceso y produce unas señales en respuesta a este análisis. Como ejemplo, si la entrada es un suceso óptico o "escena", el traductor puede dividir la escena en una pluralidad de elementos de retícula y producir unas señales  $s_1, s_2 \dots, s_j \dots$ .

$s_N$  en proporción con la densidad óptica de cada elemento de  
retícula respectivo. Si la entrada es un suceso auditivo, el  
traductor puede efectuar un análisis en serie de Fourier de  
la información auditiva y producir unas señales  $s_1, s_2, \dots,$   
5  $s_j, \dots, s_N$  en proporción con la amplitud del sonido o cada  
una de las frecuencias de la serie de Fourier. Ahora bien,  
se sobretiende que el dispositivo traductor a utilizar con  
el módulo de Nestor es cuestión enteramente de elección, y  
que se conocen ya en la técnica anterior a este invento nume-  
10 rosos tipos de dispositivos traductores. Por lo demás, como  
el traductor, de por sí, no forma parte de la presente inven-  
ción, no se estudiará aquí con detalle.

Como más arriba se ha indicado, el módulo de Nestor  
produce una pluralidad de respuestas de salida  $r_1, r_2, \dots,$   
15  $r_1, \dots, r_N$  en respuesta a un juego de señales de entrada  
 $s_1, s_2, \dots, s_j, \dots, s_N$ . En una forma preferida de realiza-  
ción del presente invento, estas respuestas de salida son  
unas variables continuas; es decir, pueden adoptar valores  
desde cero hasta cualquier valor máximo positivo y/o nega-  
20 tivo, determinados por las limitaciones o vínculos del mundo  
real que vienen dictados por la construcción del módulo de  
Nestor.

Si así se desea, en lugar de "forzarse" al módu-  
lo de Nestor a adoptar una respuesta de salida particular  
25 (un juego de respuestas individuales  $r_1, r_2, \dots, r_1, \dots, r_N$ )

a la presentación de una señal de entrada particular (un juego de señales individuales  $s_1, s_2 \dots, s_j \dots, s_N$ ), el sistema de tratamiento de información puede ir provisto de una disposición adecuada para aplicar unas respuestas específicas (por ejemplo, unas tensiones eléctricas)  $r_{1A}, r_{2A} \dots, r_{iA} \dots, r_{NA}$  a algunos (seleccionados) o a todos los terminales de salida 1, 2,  $\dots, i \dots, N$ .

De esta manera, el módulo de Nestor puede hacerse funcionar en el modo de "aprendizaje activo" y puede hacerse que dicho módulo proporcione una respuesta de salida deseada, profijada, a cualquier señal de entrada dada.

Asimismo, si se desea, los terminales de salida 1, 2  $\dots, i \dots, N$  de módulo de Nestor pueden ir conectados a una pluralidad de elementos de umbral  $T_1, T_2 \dots, T_i \dots, T_N$ , respectivamente, tales como unos circuitos de báscula de Schmitt o similares, que den una señal de salida si la respuesta de salida aplicada a los mismos excede de un nivel de umbral ajustable  $\theta_1, \theta_2 \dots, \theta_i \dots, \theta_N$ , respectivamente. Estos elementos de umbral convierten efectivamente la respuesta analógica de salida del módulo en una señal numérica de salida que puede usarse convenientemente en un tratamiento sucesivo. Además, estos elementos de umbral desempeñan una función decisoria, para determinar si se ha generado, y en qué momento se ha generado, una respuesta de salida particular.

Los elementos de umbral  $T_1, T_2 \dots, T_i \dots, T_N$  pueden utilizarse también en un modo de funcionamiento, que puede denominarse "modo de supresión", que contribuye a "adiestrar" al módulo de Nestor. Como se estudiará más adelante con mayor detalle, este modo de funcionamiento requiere que la salida de cada uno de los elementos de umbral se devuelva o lleve en retroacción al módulo de Nestor para desactivar todos los sumadores, excepto el sumador que produzca la salida. De esta manera, se suprimirán todas las respuestas de salida  $r_1, r_2 \dots, r_N$ , excepto la  $r_i$  aplicada como entrada al elemento de umbral  $T_i$  que esté produciendo señal de salida. Esto tiene la ventaja de que el módulo de Nestor se "adiestrará" por sí solo rápidamente, debido a ello, hasta producir una pauta de respuestas de salida  $r_1, r_2 \dots, r_i \dots, r_N$  (y, por lo tanto, de respuestas de salida procedentes de los elementos de umbral  $T_1, T_2 \dots, T_N$ ) en la que sólo una de estas respuestas sea distinta de cero al producirse la presentación de un juego dado de señales de entrada  $s_1, s_2 \dots, s_j \dots, s_N$ .

Los terminales de salida del módulo de Nestor, o bien de los elementos de umbral  $T_1, T_2 \dots, T_N$  si los hubiere, pueden ir conectados a un dispositivo u ordenador de salida de cualquier tipo, según la acción que se vaya a tomar a consecuencia o con dependencia de las respuestas de salida del módulo. Si el módulo de Nestor se utiliza

para identificar diseños visuales, por ejemplo (tales como firmas en un cheque bancario), las salidas de los elementos de umbral pueden ir sencillamente conectados a un dispositivo de aviso que indique a un operador humano el momento o la circunstancia en que un diseño particular se haya o no detectado (por ejemplo, que la firma sea válida o fraudulenta). Si el módulo de Nestor se utiliza como memoria pura distribuida, por ejemplo, puede ir directamente acoplado a un ordenador numérico de tipo usual (esto es, sin traductor por separado en el lado de entrada o sin elementos de umbral en el lado de salida). Naturalmente, es preciso disponer un convertidor de numérico en analógico, en el enlace de entrada del módulo, para convertir la salida numérica del ordenador en señales de entrada  $s_1, s_2 \dots, s_N$  analógicas; y es preciso asimismo disponer un convertidor de analógico en numérico en el enlace de salida del módulo, con el fin de cuantificar las respuestas analógicas de salida  $r_1, r_2 \dots, r_N$  para darles entrada en el ordenador.

Como es obvio, el módulo de Nestor es susceptible de ser utilizado para un número de fines o propósitos prácticamente infinito, y el dispositivo u ordenador particular de salida empleado vendrá determinado por las circunstancias particulares de cada caso. Como este dispositivo u ordenador de salida, de por sí, no forma parte de la presente invención, no se describirá aquí con más detalle.

El módulo de Nestor, conforme a la presente invención, puede emplearse en un sistema de tratamiento de información en el que vayan varios módulos conectados entre sí, sea en serie, sea en paralelo, sea en una combinación serie/paralelo. Por ejemplo, los terminales de salida de dos de los módulos pueden estar conectados a los terminales de entrada de un tercer módulo de manera que los dos primeros módulos puedan dar un tratamiento previo ("pretratamiento") a la información recibida del ambiente y pasar esta información al tercer módulo para su tratamiento final y almacenaje. De ese modo, las conexiones en serie y las conexiones en paralelo entre módulos pueden acrecentar el poder "intelectivo" del sistema de tratamiento de información.

El módulo de Nestor, con arreglo a la presente invención, puede estar construido para "aprender" a una velocidad deseada. En el modo de "aprendizaje", las modificaciones que se hagan en las funciones de transferencia  $A_{ij}$  de los mnemodadores han de ser como más arriba se ha definido: en una forma preferida de realización, estas modificaciones adoptan la forma siguiente:

$$\delta A_{ij} = \eta r_i s_j.$$

Mediante ajuste del valor de  $\eta$ , por ejemplo, es posible controlar la velocidad de modificación, o sea la velocidad de "aprendizaje" del módulo. Poniendo  $\eta = 0$  (esto es,  $\delta A_{ij} = 0$ ).

es posible también "desconectar" o anular por completo el modo de aprendizaje del módulo, de manera que el módulo funcione como memoria distribuida pura. El uso del módulo de Nestor en este modo de memoria, en el que las funciones de transferencia  $A_{ij}$  están predeterminadas y fijas, es uno de los rasgos característicos importantes de la presente invención.

El módulo de Nestor, con arreglo a la presente invención, puede estar construido también para "olvidar" a una velocidad prefijada, lo mismo que para aprender. Tal operación puede efectuarse permitiendo que los valores de las funciones de transferencia  $A_{ij}$  decaigan, esto es, decrezcan o declinen, por ejemplo, a velocidad constante. Cuando se está en un modo de aprendizaje, esta pérdida de información almacenada es beneficiosa, puesto que merced a ella el módulo de Nestor puede "olvidar" detalles de su experiencia anterior, y, por tanto, generalizar más rápidamente. Recíprocamente, después de haber sido "adiestrado" el módulo de Nestor y de estar éste funcionando en un modo de memoria, es conveniente reducir toda declinación de las funciones de transferencia  $A_{ij}$  hacia "cero" (esto es, hacia el valor más bajo posible con los componentes existentes), de manera que la información almacenada en el módulo de Nestor pueda conservarse el mayor tiempo posible sin el empleo de una memoria compen-

sadora.

Cuando se utilice una pluralidad de módulos de Nestor conectados en serie o en paralelo, los diferentes módulos pueden hacerse funcionar en modos distintos para llevar a cabo funciones diferentes dentro de un sistema de tratamiento de información. Por ejemplo, pueden hacerse funcionar uno o más módulos en un modo de aprendizaje o de tratamiento de información (por ejemplo, en el que tanto  $\eta$  como la velocidad de declinación uniforme sean razonablemente grandes) en tanto que uno o más módulos se pueden hacer funcionar en un modo de "memoria pura" (en el que  $\delta A_{ij}$  y la velocidad de declinación o decrecimiento sean cero). La fig. 6 ilustra un ejemplo de un sistema como éste, que consta de tres módulos de Nestor. Con referencia a dicha figura, se desprende de ella que el número de entradas del grupo H puede ser igual, mayor o menor que la suma de las salidas de  $R_O$  y  $R_A$ , y que cada salida de  $R_O$  y  $R_A$  puede estar conectada a una o más entradas de H, de manera ordenada o aleatoria.

Finalmente, como se apreciará, una vez "adiestrado" un módulo de Nestor tras un período de funcionamiento en el modo de aprendizaje, los valores de las funciones de transferencia  $A_{ij}$  pueden ser almacenados y usados para "adiestrar" instantáneamente otro módulo de Nestor. Este "adiestramiento instantáneo" se consigue simplemente ajus-

tando los valores de las funciones de transferencia  $A_{ij}$  de un módulo de Nestor a unos valores iniciales  $A_{ij}^{(0)}$  antes de poner el módulo en funcionamiento.

5 En una forma preferida de realización del presente invento, el sistema de tratamiento de información está provisto de un dispositivo de memoria compensadora usual (sea analógico o numérico) al cual pueden ser transferidos desde un módulo adiestrado los valores de las funciones de transferencia  $A_{ij}$ , y del cual pueden ser tomados  
10 estos valores cuando las funciones de transferencia del mismo módulo o de un módulo distinto se pongan a sus valores iniciales  $A_{ij}^{(0)}$ .

15 Así declarada y descrita la estructura básica del módulo de Nestor y del sistema de tratamiento de información del presente invento, se hace ahora necesario considerar con detalle la naturaleza y el funcionamiento de esta estructura. Por consiguiente, en la siguiente descripción detallada del invento se explorarán las diversas propiedades exhibidas por este módulo y este sistema.

20 A continuación se describirá con detalle el presente invento, haciendo referencia a las figs. 5 a 14 de los dibujos. Inmediatamente, en lo que sigue, se da un estudio de la base teórica de la invención; sigue luego una descripción de una forma específica preferida de realización del invento.  
25

## I. EXPLICACION TEORICA

### A. Espacio de sucesos y representaciones

Se hace referencia a la fig. 5, que ilustra un módulo de Nestor sometido a un ambiente constituido por cierto número de "sucesos". La duración y la extensión de un "suceso" se definirá finalmente de modo autoconciliable en función de la interacción entre el ambiente y el sistema adaptable que contiene el módulo de Nestor. Ahora bien, para mayor facilidad de la descripción, se procederá inicialmente como si un suceso fuese un acaecimiento objetivo bien definido, considerando un espacio de sucesos  $E$ , designados  $e^1, e^2, e^3 \dots e^k$ . Estos sucesos son "transformados" o "convertidos" por los dispositivos sensorios y de tratamiento precuz o inicial del sistema adaptable, mediante un tratamiento exterior de conversión  $P$  (del inglés "processing") en una distribución de señales, en el espacio  $S$  de entrada del módulo de Nestor, designadas  $s^1, s^2, s^3 \dots s^k$ . El tratamiento exterior de conversión  $P$  está designado por la flecha ancha (de doble trazo) de la fig. 5. Por ahora se supone que este tratamiento exterior de conversión no es modificado por la experiencia.

Aun cuando no es necesario describir aquí con detalle el tratamiento de conversión  $P$ , puesto que el tipo particular de traducción desde el ambiente al espacio de entrada no tiene importancia para el presente estudio, es

de notar que este tratamiento exterior de conversión ha de ser lo bastante rico y detallado para que se conserve una cantidad suficiente de información de interés. En particular, el juego de entradas  $S$  ha de reflejar el grado de "separación" entre sucesos: esto es, el grado de semejanza (u otra relación), en cualquier aspecto, de dos o más sucesos. Se supone, pues, que el tratamiento exterior  $P$  de conversión para pasar de  $E$  a  $S$  tiene la propiedad fundamental de conservar, en cierto sentido, la "contigüidad" o el "alejamiento" de los sucesos.

Se define aquí ahora un juego de señales de entrada  $s^{\nu}$  que corresponde al suceso entrante  $e^{\nu}$  de orden  $\nu$  (que hace el número  $\nu$ ), y un juego de señales de entrada  $s^{\mu}$  que corresponde al suceso entrante de orden  $\mu$ . En esta notación, dos sucesos  $e^{\nu}$  y  $e^{\mu}$  se convierten por tratamiento en unas entradas  $s^{\nu}$  y  $s^{\mu}$  cuya separación viene relacionada con la separación de los sucesos que les dan origen. En una representación vectorial, que es la que se empleará en todo este estudio, imagínese que dos sucesos tan semejantes como pueden ser un gato blanco y un gato gris se convierten en unos vectores casi paralelos, en tanto que dos sucesos tan diferentes como el sonido de un timbre y la vista del alimento se convierten en vectores casi ortogonales entre sí.

Dada en  $S$  la distribución de señales de entrada que viene como resultado de un suceso en  $E$ , imagínese que

esta distribución de señales es convertida interiormente por tratamiento en un juego R de respuestas de salida, por medio de un tratamiento interior de conversión A designado por la flecha sencilla de la fig. 5. Este último tipo de conversión es modificable de la manera que se describirá más adelante con mayor detalle.

Las conexiones reales y efectivas entre las entradas  $s$  y las salidas  $r$  del módulo de Nestor pueden ser aleatorias y redundantes; entre una entrada y una salida particulares puede haber muchas conexiones, o ninguna. Ahora bien, a los fines del presente estudio se idealiza la red de conexiones sustituyendo toda multiplicidad de conexiones entre una entrada y una salida por medio de una sola unión  $\oplus$ , denominada "mnemondedor", que resume lógicamente el efecto de toda la información transferida hacia delante entre el terminal de entrada  $j$  del grupo S y el terminal de salida  $i$  del grupo R. Como se ilustra en la fig. 6, cada una de las N entradas de S está conectada a cada una de las N salidas de R por medio de un solo mnemondedor  $\oplus$ . Los sumadores trabajan de manera que la respuesta o la señal presente en un terminal cualquiera de salida como el  $i$  del grupo R, esto es, la  $r_i$ , es obtenida por conversión partiendo de las señales  $s_j$  presentes en todos los terminales de entrada del grupo S, mediante la expresi

$$r_i = \sum_{j=1}^N A_{ij} s_j,$$

en la que  $A_{ij}$  es la función de transferencia del mnemodador o elemento de unión  $\otimes$  de orden  $ij$ . Es ésta la relación fundamental que da la influencia de las señales de entrada de S sobre las señales de salida de R. Aun cuando el funcionamiento satisfactorio del módulo de Nestor no requiere una suposición tan concreta y específica (esto es,  $r_i$  no tiene que ser función lineal de todas las entradas N), la sencillez de esta relación hace más fácil presentar los resultados en una forma analítica explícita.

10            B. Procedimientos asociativos lógicos, de memoria y de conversión

Es en tratamientos interiores de conversión modificables del tipo A como se almacenan la experiencia y la memoria del módulo de Nestor. En contraste con las actuales memorias de máquina, que son locales (un suceso almacenado en un lugar específico) y a las cuales puede obtenerse acceso por localización (que requiere algo equivalente a unos índices y archivos), la memoria de módulos Nestor es de tipo distribuido, y puede obtenerse acceso a la misma por su contenido o por asociación. Más adelante se demostrará que el tratamiento de conversión A es capaz de tener las propiedades de una memoria no local, a la que se puede obtener acceso por su contenido y en la que la "lógica" es resultado de una asociación y fruto de la propia naturaleza de la memoria.

25            El tratamiento de conversión A se escribe con suma

facilidad a base de la conversión de vectores que ha experimentado el sistema. En un algoritmo preferido, A puede definirse mediante la expresión:

5

$$A = \sum_{\mu, \nu} c_{\mu\nu} r^{\mu} \cdot s^{\nu},$$

10

en la que los juegos correspondientes de señales de salida para los sucesos de orden  $\nu$  y  $\mu$ ,  $e^{\nu}$  y  $e^{\mu}$ , son  $r^{\nu}$  y  $r^{\mu}$ , respectivamente, y el parámetro  $c_{\mu\nu}$  es el coeficiente de acoplamiento entre las señales de entrada  $s^{\nu}$  de orden  $\nu$  y las señales de salida  $r^{\mu}$  de orden  $\mu$ . Como se verá más adelante, el coeficiente  $c$  crece o se desarrolla normalmente con el tiempo, a medida que los sucesivos acaecimientos  $e$  se convierten por tratamiento en entradas  $s$ .

25

El elemento de orden  $ij$  de A de la intensidad o fuerza (el vigor) del mnemondador situado entre la señal entrante  $s_j$  del grupo S y la respuesta saliente  $r_i$  del grupo R. Así, si sólo  $s_j$  es distinta de cero, se tiene:

20

$$r_i = A_{ij} s_j.$$

Como

$$A_{ij} = \sum_{\mu, \nu} c_{\mu\nu} r_i^{\mu} s_j^{\nu},$$

25

la intensidad del mnemondador de orden  $ij$  se compone de la

experiencia entera del sistema, reflejada en las señales de entrada y salida conectadas a este mnemodador. Ahora bien, cada experiencia o asociación ( $\gamma^j$ ) está almacenada en la formación o disposición entera de  $N$ .  $N$  mnemodadores. Es éste el significado esencial de una memoria distribuida. Cada suceso se almacena en toda una parte amplia del sistema, en tanto que en cualquier punto local particular se superponen muchos sucesos.

1) Reconocimiento y recordación: El problema fundamental planteado por una memoria distribuida es la selección para acceso y la exactitud de recordación de los sucesos almacenados. Considérese primero la porción "diagonal" de  $A$ , que viene definida por la expresión siguiente:

$$(A) \text{ diagonal} = R = \sum_j c_{jj} r^j \cdot s^j$$

( en la cual el símbolo  $R$  de letra cursiva significa "reconocimiento y recordación" ).

Un suceso  $a$  arbitrario, convertido por tratamiento en las señales de entrada, generará en  $R$  la respuesta siguiente:

$$r = A s .$$

Si se iguala el reconocimiento con la fuerza de esta respuesta  $r$ , es decir, el valor de

$$(r, r) = \sum_{i=1}^N r_i^2$$

( el "producto. interior" del vector  $\underline{r}$  consigo mismo, esto es, el cuadrado de la longitud de  $\underline{r}$ ), el tratamiento de conversión A distinguirá entonces entre los sucesos que contiene (los  $s^{\nu}$ ,  $\nu = 1, 2 \dots k$ ) y otros sucesos separados de éstos.

5

El vocablo "separado", usado en este contexto, requiere ahora una definición más precisa. En un tipo de argumento usado por J.A. Anderson, Math. Bio-sciences 8, 137 (1970), al analizar una memoria distribuida, los vectores  $s^{\nu}$  se suponen independientes entre sí, y se supone también que satisfacen los requisitos de que, por término medio, sea:

10

$$\sum_{i=1}^N s_i^{\nu} = 0$$

15

$$\sum_{i=1}^N (s_i^{\nu})^2 = 1$$

Dos vectores cualesquiera de éstos tienen componentes aleatorios entre sí, de manera que un nuevo vector,  $\underline{s}$ , presentado a R según lo indicado más arriba da una respuesta semejante a ruido, ya que por término medio ( $s^{\lambda}$ ,  $s$ ) es pequeña. La presentación de un vector visto anteriormente ( por ejemplo el  $s^{\lambda}$ ), en cambio, da por respuesta

20

$$R \ s^{\lambda} = c_{\lambda\lambda} \ r^{\lambda} + \text{ruido.}$$

25

Se pone de manifiesto, pues, que si el número de sucesos impresos  $k$  es pequeño comparado con  $N$ , las relaciones de señal a ruido son razonables.

Definiendo aquí los sucesos separados como aquellos que se convierten por tratamiento en vectores ortogonales, queda claro entonces que una matriz de reconocimiento compuesta de  $k$  vectores ortogonales  $s^1, s^2, \dots, s^k$

$$R = \sum_{\nu=1}^k c_{\nu\nu} r^{\nu} \cdot s^{\nu}$$

distinguirá entre los vectores contenidos,  $s^1 \dots s^k$ , y todos los vectores separados de éstos (perpendiculares a éstos). Además, la respuesta de  $R$  a un vector anteriormente registrado es única y completamente precisa:

$$R s^{\lambda} = c_{\lambda\lambda} r^{\lambda}.$$

En esta situación especial, la memoria distribuida tiene tanta precisión como una memoria localizada.

Además, como ha señalado H.C. Longuet-Higgins, Proc. R. Soc. Lond. B, 171 327 (1968), una memoria distribuida puede tener la interesante propiedad de recordar o reclamar un vector entero de respuesta  $r^{\lambda}$  aun cuando se presente sólo parte de la señal  $s^{\lambda}$ . Es éste el caso de la memoria distribuida aquí estudiada. Sea

$$s^\lambda = s_1^\lambda + s_2^\lambda$$

Si se presenta sólo parte de  $s^\lambda$ , por ejemplo  $s_1^\lambda$ , se obtiene:

5 
$$R s_1^\lambda = c_{\lambda\lambda} (s_1^\lambda, s_1^\lambda) r^\lambda + \text{ruido.}$$

El resultado es, pues, la respuesta entera a la señal  $s^\lambda$  completa con un coeficiente reducido, más ruido.

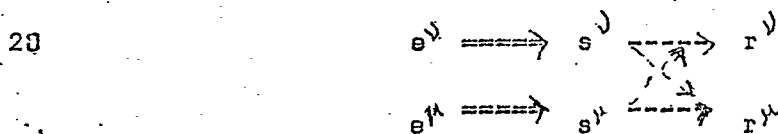
10 2) Asociación: La presentación del suceso  $e^\nu$  que genera el vector  $s^\nu$  da por resultado el reconocimiento y la recordación si

$$R s^\nu = c r^\nu + \text{ruido.}$$

Luego los términos de fuera de la diagonal

15 
$$(A)_{\text{fuera de diagonal}} = A = \sum_{\mu \neq \nu} c_{\mu\nu} r^\mu \cdot s^\nu$$

(donde el símbolo  $\lambda$  en cursiva significa "asociación") puede interpretarse como conducente a la asociación de sucesos inicialmente separados entre sí



donde  $(s^\nu, s^\mu) = 0$ .

25 Pues con tales términos la presentación del suceso  $e^\nu$  engendrará no sólo  $r^\nu$  (que es equivalente al reconocimiento)

de  $e^{\nu}$ ) sino también (quizá más débilmente)  $r^{\mu}$ , que habría de resultar con la presentación de  $e^{\mu}$ . Así, por ejemplo, si  $r^{\mu}$  iniciase alguna respuesta (originariamente, una respuesta a  $e^{\mu}$ ), la presentación de  $e^{\nu}$  cuando  $c_{\mu\nu} \neq 0$  iniciará también esta respuesta.

Por lo tanto, es posible escribir ahora la matriz de asociación:

$$A = \sum_{\mu\nu} c_{\mu\nu} r^{\mu} \cdot s^{\nu} = R + \lambda,$$

en la cual

$$R = (A)_{\text{diagonal}} \equiv \sum_{\nu} c_{\nu\nu} r^{\nu} \cdot s^{\nu} \text{ (reconocimiento)}$$

y

$$\lambda = (A)_{\text{fuera de diagonal}} \equiv \sum_{\mu \neq \nu} c_{\mu\nu} r^{\mu} \cdot s^{\nu} \text{ (asociación)}.$$

Los  $c_{\mu\nu}$  son entonces los coeficientes de reconocimiento y asociación "directos".

3) Generalización: En la experiencia real y efectiva, los sucesos a los cuales se hallaría expuesto el sistema no estarían en general muy separados e independientes, en el sentido estadístico. Por lo tanto, no existe razón alguna para esperar que todos los vectores,  $s^{\nu}$ , impresos en A fuesen ortogonales o ni siquiera estuviesen muy alejados unos de otros. Por el contrario, parece probable que, a menudo, gran número de estos vectores se hallarían muy próximos entre

sí. En estas circunstancias, una memoria distribuida del tipo contenido en A llegará a confundirse y cometer errores: "reconocerá" y "asociará" sucesos que, de hecho, nunca se hayan visto o asociado antes.

5                    Para ilustrar con algún ejemplo, supóngase que el sistema ha sido expuesto a una clase de sucesos no separados  $\{e^1 \dots e^k\} : \{e^\alpha\}$  que por tratamiento se convierten en los  $k$  vectores  $\{s^1 \dots s^k\} : \{s^\alpha\}$ . La contigüidad de los sucesos así convertidos puede expresarse en un espacio lineal por medio del concepto de "comunidad". Se define aquí la comunidad de un juego de vectores, tal como el  $\{s^\alpha\}$  arriba indicado, como la cota o límite inferior de los productos interiores  $(s^u, s^t)$  de dos vectores cualesquiera de este juego. Concretamente, la comunidad del juego de vectores  $\{s^\alpha\}$  es  $\tau$ ,  $\tau = \inf [s^\alpha] = \tau$ , si  $\tau$  es el límite inferior de  $(s^u, s^t)$  para todos los  $s^u$  y  $s^t$  contenidos en  $\{s^\alpha\}$ .

15                    Si cada exposición da por resultado la adición a A (o a R) de un elemento de la forma  $c_{\nu\nu} r^\nu \cdot s^\nu$ , la respuesta entonces a un suceso  $s^u$  de esta clase,  $s^u \in \{s^\alpha\}$ , es:

20                    
$$R s^u = r = \sum_{\nu} c_{\nu\nu} r^\nu (s^\nu, s^u) = c_{uu} r^u + \sum_{\nu \neq u} (s^\nu, s^u) c_{\nu\nu} r^\nu,$$

donde  $(s^\nu, s^u) > \tau$ .

25                    Si  $\tau$  es suficientemente grande, la respuesta a  $s^u$ , por con-

siguiente, no se distingue muy claramente de la de otra cualquiera  $s$  contenida en  $\{s^\alpha\}$ .

Si se presenta al sistema un nuevo suceso,  $e^{k+1}$ , no visto anteriormente, y este nuevo suceso está próximo o contiguo a los demás de la clase  $\alpha$  (por ejemplo, supóngase que  $e^{k+1}$  se convierte por tratamiento en  $s^{k+1}$ , que es un miembro de la comunidad  $\{s^\alpha\}$ ),  $R s^{k+1}$  dará entonces una respuesta no demasiado diferente de la producida para uno de los vectores  $s^u \in \{s^\alpha\}$ . Por lo tanto, el suceso  $e^{k+1}$  será reconocido, aunque no haya sido visto antes.

Esto, naturalmente, es en potencia un error muy valioso, pues la memoria asociativa reconoce y luego atribuye propiedades a unos sucesos que caen dentro de la misma clase que los sucesos ya reconocidos. De hecho, si los vectores contenidos en  $\{s\}$  tienen la forma

$$s^{\nu} = s^0 + n^{\nu},$$

en la que  $n^{\nu}$ , el factor de ruido, varía de manera aleatoria, llegará un momento en que  $s^0$  será reconocido más fuertemente que cualquiera de los  $s^{\nu}$  particulares presentados real y verdaderamente. De este modo, por ejemplo, es posible extraer de un ruido aleatorio una señal repetida.

Se tiene aquí, pues, una realización explícita de lo que en cierto modo y libremente pudiera denominarse una "lógica" -- que, naturalmente, no es lógica en manera alguna.

Más bien, lo que ocurre podría describirse como resultado de una directriz intrínseca para generalizar. La memoria asociativa, por naturaleza, adopta la fórmula de conducta

5 
$$s^0 + n^1, s^0 + n^2 \dots s^0 + \dots \rightarrow s^0$$

que, en términos de lenguaje, podría describirse como pasar de lo particular (por ejemplo, gato<sup>1</sup>, gato<sup>2</sup>, gato<sup>3</sup> ...) a lo "general ("gato").

10 La rapidez con que se adopte y realice esta fórmula depende (como se verá en el apartado siguiente) de los parámetros del sistema. Mediante el recurso de alterar estos parámetros, es posible construir o realizar conversiones por tratamiento que varían desde aquellas que conservan todos los detalles a los cuales están expuestas hasta aquellas que pierden los detalles y conservan solamente elementos comunes -- el vector central de cualquier clase.

15

Además de "errores" de reconocimiento, la memoria asociativa comete también errores de asociación. Por ejemplo, si todos los vectores (o muchos de ellos) de la clase  $\{s^k\}$  con una comunidad razonablemente amplia llevan asociada alguna respuesta  $r^b$  particular, de modo que el tratamiento de conversión A contiene términos de la forma

20

25 
$$\sum_{j=1}^k c_{\beta D} r^{\beta} \cdot s^D,$$

con  $c_{\beta}^{\nu} \neq 0$  en gran parte de  $\nu = 1, 2 \dots k$ , el nuevo suceso  $s^{k+1}$ , que se convierte por tratamiento en  $s^{k+1}$  como en el ejemplo anterior, no sólo será entonces reconocido, por ser  $(R s^{k+1}, R s^{k+1})$  grande, sino que también llevará asociada  $r^{\beta}$

5

$$A s^{k+1} = c r^{\beta} + \dots$$

tan fuertemente como cualquiera de los vectores contenidos en  $\{s^{\alpha}\}$ .

Si los errores de reconocimiento conducen al proceso descrito en términos de lenguaje como "pasar de lo particular a lo general", los errores de asociación podrían describirse como "pasar de lo particular a lo universal": por ejemplo, el gato<sup>1</sup> maúlla, el gato<sup>2</sup> maúlla ..., --> todos los gatos maúllan.

10

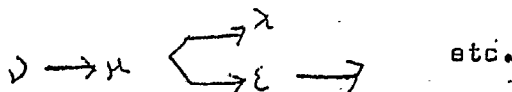
15

No existe, naturalmente, "justificación" alguna para este proceso. Se realiza como consecuencia de la naturaleza del sistema. Sea cualquiera la eficacia que tenga, ésta dependerá del orden del mundo en que se encuentre el sistema.

20

Mediante una sucesión de tratamientos de conversión de la forma indicada (o bien mediante retroacción o realimentación de la salida de A hacia sí misma), se obtiene una trama de sucesos y conexiones.

25



tan rico como sugerente. Se ve fácilmente la posibilidad de un flujo de actividad eléctrica influido tanto por los tratamiento de conversión interior de la forma A como por la entrada exterior. Este flujo viene regulado no sólo por los coeficientes de asociación directa  $c_{\mu\nu}$  (que pueden aprenderse explícitamente, como se describirá a continuación) sino también por asociaciones indirectas debidas a la superposición de los sucesos convertidos por tratamiento, como se indica en la fig. 7. Además, es posible imaginar fácilmente que surjan situaciones en las cuales se haya perdido el acceso directo a un suceso, o a una clase de sucesos ( $c_{\mu\nu} = 0$  en la fig. 7), mientras la existencia de este suceso o tal clase de sucesos en A influye en el flujo de actividad eléctrica.

4) Separación de vectores: Todo estado de una memoria distribuida es, en general, una superposición de diversos vectores. Así, pues, se tiene que emplear unos medios merced a los cuales se distingan entre sí los sucesos (o las entidades en las cuales hayan sido convertidos).

Existen varias posibilidades: No es difícil, ni mucho menos, imaginar dispositivos de umbral, o no lineales, que separen un vector de otro. Pero la aparición de un vector de la clase  $\{s^d\}$  en una memoria distribuida da por resultado un juego de respuestas de salida en todo un gran número de salidas  $r_i$ , cada una de las cuales está lejos del umbral. Por

tanto, se plantea el problema lógico de asociar el umbral de una única respuesta a una señal distribuida como ésta. En un apartado más adelante se describirá de qué modo puede hacerse esto.

5

Además de la aparición de tales salidas de umbral, puede haber cierta separación de las señales convertidas por tratamiento, debido a la localización real y efectiva de las áreas en las que estas señales se producen. Por ejemplo, las señales ópticas y auditivas podrían someterse mucho a tratamiento antes de encontrarse en realidad en un módulo de Nestor. Es posible permitir que se produzcan primero la identificación de señales ópticas o auditivas (como ópticas o auditivas); las conexiones entre un suceso óptico y uno auditivo podrían entonces producirse a continuación, en un segundo nivel de tratamiento, yendo desde el grupo de respuesta R a un segundo grupo de respuesta H, como se sugiere en la fig. 8.

10

15

### C. Aprendizaje, y modificación de módulos

El elemento de orden  $ij$  de la conversión asociativa A

20

$$A_{ij} = \sum_{\mu\nu} c_{\mu\nu} r_i^\mu \cdot s_j^\nu \quad (1)$$

es una suma ponderada de las componentes  $j$  de todas las señales  $s_j^\nu$  convertidas por tratamiento, y de las componentes  $i$  de las respuestas  $r_i^\mu$ , apropiada para la recordación o la

25

asociación.

Esta conversión por tratamiento puede conseguirse, naturalmente, mediante ajuste del "peso" de cada mnemodador, de manera que su valor sea igual al correspondiente  $A_{ij}$  según lo que antecede. Esta es el modo más sencillo en que puede funcionar el módulo de Nestor.

Una característica sumamente importante del módulo de Nestor es su capacidad de automodificación. Cuando está funcionando en un modo de aprendizaje, el módulo de Nestor modifica los "pesos" de sus mnemodadores de manera que (aparte de un decrecimiento o declinación uniforme que se describe más adelante) se tiene:

$$\delta A_{ij} \sim r_i s_j$$

Este  $\delta A_{ij}$  es proporcional al producto de la entrada  $s_j$  y la salida  $r_i$ . También son posibles las alteraciones en la intensidad de la unión, proporcionales solamente a  $s_j$  o a la respuesta de la unión inmediata  $s_j$ ; ahora bien, tales modificaciones no dan por resultado las diversas propiedades aquí estudiadas. La adición de tales cambios de  $A$  indicados por la proporcionalidad (2) arriba escrita, para todas las asociaciones  $r^k \cdot s^l$  da por resultado, también, una conversión por tratamiento con las propiedades estudiadas en el apartado precedente.

Para efectuar las modificaciones de  $A_{ij}$

$$\int A \sim r^{\nu} \cdot s^{\nu} \quad (3)$$

5 por medio del procedimiento de automodificación del módulo de Nestor, el sistema ha de tener en su grupo S la distribución de señales  $s^{\nu}$ , y la distribución  $r^{\nu}$  en su grupo R, donde  $s^{\nu}$  se encuentra convertida por tratamiento mediante P, a partir del suceso  $s^{\nu}$ .

10 En lo que aquí se denomina "aprendizaje activo", al módulo de Nestor puede serle presentada una entrada  $s^{\lambda}$  y el módulo puede ser forzado a producir la respuesta de "correcto", esto es,  $r^{\omega}$ . Esto puede hacerse, por ejemplo, con una disposición o aparato del tipo ilustrado en la figura 9, en la que los valores de respuesta deseados  $r_{1A}, r_{2A}, \dots, r_{iA}, \dots, r_{NA}$  pueden ser aplicados a las salidas 1, 2, ..., j ..., N para vincular o constreñir las señales de salida 15  $r_1, r_2, \dots, r_i, \dots, r_N$  a que sean iguales a estos valores deseados. Como las señales de salida se utilizan en el módulo de Nestor en la modificación de los elementos  $A_{ij}$  con arreglo a la proporcionalidad (2) arriba indicada; esto es, con arreglo a

$$20 \quad \int A_{ij} = \eta r_i^{\omega} s_j^{\lambda}$$

(expresión en la que  $\eta$  es la constante de proporcionalidad), al producirse la aplicación repetida de la entrada  $s^{\lambda}$ , el 25 módulo acumula con gran rapidez una memoria detallada y pre-

cisa de la respuesta de salida  $r^w$  a la entrada  $s^\lambda$ .

El aprendizaje activo indica también un tipo de aprendizaje en el que una respuesta del sistema a un juego de entradas se hace concordar con una respuesta esperada o deseada, y se juzga correcta o incorrecta. En este caso, si al sistema se le presenta una entrada  $s^\lambda$ , su respuesta  $r^\lambda$  de salida puede compararse con la respuesta adecuada o de "correcto"  $r^w$ , y puede hacerse que los elementos  $A_{ij}$  sean incrementados en el sentido que dé lugar a una respuesta más próxima a  $r^w$  si se vuelve a aplicar  $s^\lambda$ .

Es evidente que, con el aprendizaje activo, se requiere que el operador humano del sistema conozca la respuesta  $r$  apropiada a las diversas entradas  $s$ . Ahora bien, el módulo de Nestor es capaz de otro modo de funcionamiento, al que se denominará aquí "aprendizaje pasivo", que no requiere intervención humana. En particular, el módulo de Nestor es capaz de identificar, reconocer, recordar, generalizar o asociar unas características del ambiente al cual se halla expuesto, aun cuando no se haya hecho análisis ni identificación previos de los rasgos característicos del ambiente exterior. En este tipo de aprendizaje, el operador humano no tiene por qué darse cuenta siquiera de las características correspondientes del ambiente que se están extrayendo y y tratando en el módulo de Nestor.

Para llegar a un algoritmo que produzca el apren-

dizaje pasivo, se utiliza aquí una distinción entre formar una representación interna de sucesos del mundo exterior y, como antagonista, producir una respuesta a estos sucesos que concuerde en relación con lo que se espera o desea del mundo exterior.

La idea, sencilla pero importante, es que la actividad eléctrica interior que en un determinado módulo señala la presencia de un suceso exterior no es necesariamente la misma actividad eléctrica que señala la presencia de ese mismo suceso para otro módulo. No hay nada que requiera que el mismo suceso exterior sea convertido en las mismas distribuciones de señal por módulos diferentes. El suceso  $s^d$ , que para un determinado módulo se convierte por tratamiento en las distribuciones de señales  $r^d$  y  $s^d$ , en otro módulo puede ser convertido en  $r'^d$  y  $s'^d$ . Lo que se necesita para el eventual acuerdo entre módulos, en su descripción del mundo exterior, no es que las señales eléctricas convertidas por tratamiento sean idénticas, sino más bien que la relación de las señales entre sí y con los sucesos del mundo exterior sea la misma. La fig. 10 ilustra este principio de forma gráfica.

1) Aprendizaje pasivo: Sea  $A^{(t)}$  la matriz  $A$  (esto es, el juego de funciones  $A_{ij}$ ) tras la presentación de  $\underline{t}$  sucesos (" tiempo"  $\underline{t}$ ). Se escribe:

$$A(t) = \gamma A(t-1) + \delta A(t),$$

donde

$$\delta A(t) = \eta r^t \cdot s^t.$$

En esta ecuación, como antes se ha dicho,  $\eta$  es la constante de proporcionalidad y  $\gamma$  es una "constante de decrecimiento o declinación" sin dimensiones, que da la medida de la declinación uniforme de la información en cada lugar o emplazamiento (una especie de "olvido"). Usualmente se tiene  $0 < \gamma \leq 1$ .

También se introduce ahora el parámetro  $\xi$ , definido como el valor de  $\eta$  cuando las entradas  $s^t$  son normalizadas.  $\xi$ , que da la medida de la velocidad o régimen a que se hacen modificaciones en  $A$  (velocidad de aprendizaje), se utilizará en cálculos ilustrativos realizados para entradas  $s^t$  normalizadas. Los valores de los parámetros  $\gamma$ ,  $\eta$  y  $\xi$  pueden ajustarse a discreción del usuario con el fin de producir las propiedades deseadas del sistema. Por ejemplo, durante el período de adquisición de la información (aprendizaje o inscripción de programa),  $\eta$  o  $\xi$  podrían ser razonablemente mayores que cero (por ejemplo,  $\eta \approx \xi \approx 1/10$ ), y  $\gamma$  podría ser razonablemente menor que uno (por ejemplo,  $\gamma \approx 9/10$ ), de manera que el sistema adquirirá información y perderá detalles. Tras el período de adquisición, puede ser útil fijar  $\eta = \xi = 0$  y  $\gamma \approx 1$ , de modo que el sistema dejará de "aprender" pero retendrá durante un período o intervalo de tiempo arbitrario toda la información que haya adquirido. En un módulo en

funcionamiento, este tiempo de almacenaje viene determinado por las constantes de tiempo características de los circuitos. Por razones de economía o conveniencia, éstas pueden elegirse de manera que permitan el almacenaje durante períodos del orden de minutos o de horas. Para almacenajes más duraderos, en tales circunstancias, sería posible transferir el contenido de la memoria distribuida (por ejemplo, los valores de la  $A_{ij}$ ) a una memoria compensadora que lo devolviese cuando fuera necesario.

En general, un sistema en el que sea  $\gamma < 1$  pierde detalles y tiene mayor capacidad para generalizar. Resulta, pues, que los valores de  $\gamma$  iguales o ligeramente menores que 1 son, pues, de sumo interés. Con el fin de mantener el sistema libre de saturación, también es conveniente realizar la modificación cero (dejando  $\eta = 0$ ) cuando la salida,  $r = As$ , supera un máximo especificado; es decir:  $(r, r) = (As, As) > \text{máximo especificado}$ .

En lo que sigue, se normalizan todos los vectores  $(s, s) = 1$ . de manera que  $\gamma$ , que se torna ahora para que sea constante, carece de dimensión.

Si ahora se dice que  $r^t$  es:

$$r^t \gamma = A^{(t-1)} s^t + r_R^t + r_A^t,$$

como puede verse, la respuesta total se compone de tres términos; una respuesta pasiva,  $\gamma A^{(t-1)} s^t$ , un término activo pero aleatorio  $r_R^t$ , y una respuesta activa,  $r_A^t$ . Para el aprendizaje puramente pasivo, se tiene en cuenta sólo el primer término, de manera que

$$\xi A^{(t)} = \xi r^t \cdot s^t = \xi \xi A^{(t-1)} s^t \cdot s^t .$$

Aquí las respuestas son justamente las producidas por el tratamiento de conversión existente,  $A^{(t-1)}$ , cuando el vector  $s^t$  del grupo S se convierte por tratamiento en R:

5

$$r^t \gamma = A^{(t-1)} s^t .$$

El algoritmo de aprendizaje pasivo es entonces

$$\begin{aligned} A^{(t)} &= A^{(t-1)} (\gamma + \xi \xi s^t \cdot s^t) = \\ &= \xi A^{(t-1)} (1 + \xi s^t \cdot s^t) , \end{aligned}$$

10

donde, en general,  $\xi$  sería usualmente mucho menor que la unidad. Antes de haberse presentado cualquier suceso exterior, A tiene la forma  $A^{(0)}$ , que podría ser aleatoria. El efecto de  $A^{(0)}$  sobre la conversión por tratamiento interior se analizará más adelante.

15

Con este algoritmo, al cabo de  $k$  sucesos, A tiene la forma:

$$A^{(k)} = \xi^k A^{(0)} \prod_{\nu=1}^k (1 + \xi s^{(\nu)} \cdot s^{(\nu)}) ,$$

en la que  $\pi$  es un producto ordenado, en el que los factores que tienen índices inferiores quedan a la izquierda:

20

$$\prod_{\nu=1}^k s^{(\nu)} = s(1) s(2) \dots s(k) .$$

Esto puede escribirse también así:

25

$$A^{(k)} = \gamma^k A^{(0)} \left[ 1 + \xi \sum_{\nu=1}^k s^\nu \cdot s^\nu + \xi^2 \sum_{\nu < \mu} s^\nu \cdot s^\mu \right. \\ \left. (s^\nu, s^\mu) + \dots + \xi^k s^1 \cdot s^k (s^1, s^2) (s^2, s^3) \right. \\ \left. (s^3, s^4) \dots (s^{k-1}, s^k) \right].$$

5

El algoritmo de aprendizaje pasivo genera su propia respuesta  $A^{(0)} s^\nu$  al vector entrante  $s^\nu$ , respuesta que dependen de la configuración originaria de la red que recorre  $A^{(0)}$ , y del vector  $s^\nu$  obtenido por conversión a partir del suceso  $s^\nu$ . Por ejemplo, si  $s^\nu$  es el único vector presentado, A adopta eventualmente la forma

10

$$A \rightsquigarrow r^\nu \cdot s^\nu.$$

en la cual

15

$$r^\nu \equiv A^{(0)} s^\nu.$$

2) Casos especiales de A: Se presenta ahora la forma de A en varios casos especiales; en todos ellos, se supone que  $\xi$  es constante y pequeña.

20

(a) Si los k vectores son ortogonales, A se convierte en

$$A^{(k)} = \gamma^k \left( A^{(0)} + \xi A^{(0)} \sum_{\nu=1}^k s^\nu \cdot s^\nu \right).$$

Haciendo  $A^{(0)} s^\nu \equiv r^\nu$ , el segundo término adopta la forma de la parte "diagonal" de A:

25

$$(A)_{\text{diagonal}} \equiv R = \sum_{j=1}^k r^j \cdot s^j .$$

5 y servirá para el reconocimiento de los vectores  $s^1 \dots s^k$ .  
 (hay que hacer notar que los vectores  $r^j$  asociados no se  
 dan por adelantado; son generados por la red.). Ahora bien,  
 si  $\zeta$  es pequeño, esto podría ser inadecuado para el recono-  
 cimiento, puesto que el término de reconocimiento sería dé-  
 bil. Además, por lo general resultará más útil que el reco-  
 10 nimiento se haga aparecer sólo tras una repetida exposi-  
 ción al mismo suceso.

(b) El siguiente ejemplo demuestra que el algorit-  
 mo de aprendizaje pasivo acumula o forma realmente coeficien-  
 tes de reconocimiento a velocidad exponencial para repetidas  
 15 entradas del mismo suceso. Si se presenta  $l$  veces el mismo  
 vector  $s^0$ ,  $A$  llega eventualmente a ser

$$A^{(l)} = \gamma^l A^{(0)} (1 + e^{\zeta l} s^0 \cdot s^0) .$$

Si  $l$  es lo bastante grande para que  $e^{\zeta l} \gg 1$ , llegará un momen-  
 to en que el término de reconocimiento domine. Cuando  $e^{\zeta l}$  se  
 20 haga lo bastante grande, puede resultar deseable ajustar el  
 valor de  $\zeta$  de modo que no haya más desarrollo o crecimiento.  
 Esto puede lograrse haciendo que  $\zeta$  sea función de la respues-  
 ta al vector entrante, para que más allá de cierto valor má-  
 ximo no haya más aumento del coeficiente.

25 (c) La presentación de  $m$  vectores ortogonales  $l_1, l_2$

...,  $l$  veces da lugar a una sencilla generalización del segundo resultado. Si se hace  $\gamma = 1$  para simplificar,

$$A^{(l_1 + l_2 + \dots + l_m)} = A^{(0)} \left( 1 + \sum_{\nu=1}^m e^{l_\nu \epsilon} s^\nu \cdot s^\nu \right)$$

5

$$s^\nu \cdot s^\nu$$

que es precisamente una matriz asociativa de reconocimiento y recordación

$$A = \sum_{\nu=1}^m c_{\nu\nu} r^\nu \cdot s^\nu$$

10

si

$$e^{l_\nu \epsilon} = c_{\nu\nu} \gg 1.$$

(d) Parte del efecto de la falta de ortogonalidad puede ser presentada, mediante el recurso de calcular el resultado de una entrada consistente en  $l$  vectores ruidosos distribuidos de manera aleatoria en torno a una  $s^0$  central

15

$$s^\nu = s^0 + n^\nu.$$

Aquí  $n^\nu$  es un "vector estocástico" (esto es, un vector que varía aleatoriamente) cuya magnitud es pequeña comparada con la de  $s^0$ . Se obtiene, pues

20

$$A^{(l)} = \gamma^l A^{(0)} e^{l \epsilon} n^2/N \left( 1 + e^{l \epsilon} s^0 \cdot s^0 \right)$$

25

donde  $\underline{n}$  es la magnitud media de  $n^\nu$ . Como se puede ver, la

$A^{(\lambda)}$  generada, con el factor adicional debido al ruido, es precisamente de la forma adecuada para el reconocimiento de  $s^0$ . Así, la repetida aplicación de un vector ruidoso de la forma arriba indicada da por resultado una función A que reconoce el vector central  $s^0$ . Esto proporciona también

3) Estructura del espacio convertido por tratamiento: Las comunidades o clases separadas, de los espacios de señales o de los exteriores, E o S, serán las mismas del espacio R convertido por tratamiento, si

$$(r^\alpha, r^\beta) = (s^\alpha, s^\beta),$$

donde

$$r^\alpha = A^{(0)} s^\alpha.$$

Esto sucederá si  $A^{(0)}$  satisface la relación

$$(A^{(0)})^T A^{(0)} = I \text{ (la matriz de identidad)}$$

o bien

$$\sum_{i=1}^N A_{ij}^{(0)} A_{ik}^{(0)} = \delta_{jk},$$

donde

$$\begin{aligned} \delta_{jk} &= 1 & j &= k \\ &= 0 & j &\neq k, \end{aligned}$$

de donde se sigue entonces

$$(r^\alpha, r^\beta) = (A^{(0)} s^\alpha, A^{(0)} s^\beta) = (s^\alpha, s^\beta).$$

Esto puede disponerse fácilmente. Si, por ejemplo, se elige

$$A^{(0)} = I,$$

se satisface entonces la igualdad (4); y el espacio S se convierte por tratamiento en sí mismo:

$$r^{\alpha} = s^{\alpha}.$$

Es interesante notar que incluso una  $A^{(0)}$  aleatoria, por término medio, satisfará el requisito de la igualdad (4). Supóngase que  $A^{(0)}$  es una matriz simétrica aleatoria y satisface las condiciones

$$\sum_{i=1}^N A_{ij}^{(0)} \approx 0$$

para todas las  $j$ ,

$$\sum_{i=1}^N (A_{ij}^{(0)})^2 \approx 1$$

luego

$$\sum_{i=1}^N A_{ij}^{(0)} A_{ik}^{(0)} \approx 0 \quad j \neq k$$

en tanto que

$$\sum_{i=1}^N A_{ij}^{(0)} A_{ij}^{(0)} = \sum_{i=1}^N (A_{ij}^{(0)})^2 \approx 1 \quad j = k.$$

Así, se satisface la condición (4); por lo tanto, una  $A^{(0)}$

aleatoria, según lo anterior, conducirá a un espacio convertido por tratamiento, con las mismas comunidades y clases que el espacio de señales originario.

4) Términos de asociación: Los términos asociativos o fuera de diagonal pueden ser generados del siguiente modo: Supóngase que A ha alcanzado la forma

$$A = \sum_{\nu=1}^k A^{(0)} s^{\nu} \cdot s^{\nu} = \sum_{\nu=1}^k r^{\nu} \cdot s^{\nu}.$$

10 Preséntese ahora los sucesos  $e^{\alpha}$  y  $e^{\beta}$  de manera que estén "asociados", para que los vectores  $s^{\alpha}$  y  $s^{\beta}$  aparezcan o salgan juntos por "conversión". (Las condiciones precisas que den por resultado tal conversión simultánea de  $s^{\alpha}$  y  $s^{\beta}$  dependerán de la construcción del sistema. La situación más sencilla de imaginar es aquellas en la cual el vector  $(s^{\alpha} + s^{\beta})$  es convertido por tratamiento si  $e^{\alpha}$  y  $e^{\beta}$  se presentan al sistema lo bastante juntos entre sí en el tiempo.) Se puede suponer que  $e^{\alpha}$  y  $e^{\beta}$  están separados de manera que  $(s^{\alpha}, s^{\beta}) = 0$ .  
15 En el grupo S, si el vector está normalizado por conveniencia, se tiene entonces  $1/\sqrt{2} (s^{\alpha} + s^{\beta})$ .

20 Tras una presentación de  $e^{\alpha}$  y  $e^{\beta}$  tal como ésta, A se convierte (dándole a  $\gamma$  el valor 1 con fines de amplificación), en:

$$A^{(1)} = \sum_{\nu=1}^k r^{\nu} \cdot s^{\nu} + \frac{\gamma}{2} (r^{\alpha} \cdot s^{\alpha} + r^{\beta} \cdot s^{\beta}).$$

El segundo término de la asociación entre  $\alpha$  y  $\beta$  con el coeficiente

$$c_{\alpha\beta} = \beta\alpha \quad c = \epsilon/2,$$

que generalmente (excepto en circunstancias especiales) sería sumamente útil de ser pequeño. Si  $s^\alpha$  y  $s^\beta$  no vuelven a aparecer en asociación,  $c_{\alpha\beta}$  o  $c_{\beta\alpha}$  (aun cuando se desarrollan realmente al presentarse  $s^\alpha$  o  $s^\beta$  por separado) siguen siendo pequeños en comparación con los respectivos coeficientes de reconocimiento  $c_{\beta\beta}$  o  $c_{\alpha\alpha}$ . Ahora bien, si  $(s^\alpha + s^\beta)/\sqrt{2}$  aparece con frecuencia (por ejemplo,  $\ell$  veces), el coeficiente del término en cruz se convierte en

$$c_{\alpha\beta} \approx \frac{\epsilon}{2} \ell$$

y se hace tan grande como el coeficiente de reconocimiento.

Con los resultados que anteceden se ha establecido que los espacios de señales y de respuestas, en unión de la conversión por tratamiento que los conecta, combinen una estructura análoga a la primitiva del ambiente exterior o del espacio de sucesos, E. Esto significa lo siguiente:

(1) Las clases o las comunidades del espacio de respuestas R son las mismas de los espacios exterior o de señales, E y S.

(2) Las clases o sucesos asociados en el espacio exterior (los que aparecen en asociación durante

un período de aprendizaje) llegan a asociarse en el espacio de respuesta de manera que, tras el período de aprendizaje, la aparición de un determinado miembro de los sucesos o clases asociados en el espacio exterior E, y por lo tanto en el espacio de señales S, convertirá ambos miembros de las clases o sucesos asociados del espacio de respuesta R, aun cuando se trata de sucesos de tipos muy diferentes.

5) Separación de sucesos - Dispositivos de umbral:

En lo que antecede se ha hablado y tratado de espacios y conversiones lineales. Como consecuencia, un estado es en general una superposición de varios vectores. Para distinguir entre sí los vectores -- o las señales en las que se han convertido -- es posible incorporar al sistema un dispositivo de umbral o de otro tipo no lineal. También puede haber una separación de las señales de conversión, debida a localización de las áreas en donde aparecen estas señales. Por ejemplo, como se ilustra en la fig. 8, las señales ópticas y auditivas pueden someterse a tratamiento en módulos separados antes de que realmente se encuentren en un módulo común. Así, la identificación de las señales ópticas o auditivas se produciría primero a partir del módulo en el cual se conviertan inicialmente. En el módulo común surgirían entonces las asociaciones entre un suceso óptico y un suceso auditivo.

A continuación se describe un ejemplo de dispositivo de umbral. Como una señal en una memoria distribuida se

5 extiende o disemina por todo un gran número de entradas o de salidas, incluso una señal grande [ (  $s^N$ ,  $s^N$  ) grande ] podría estar constituida por unas componentes,  $s_i^d$ , cada una de las cuales fuese bastante pequeña. Por lo tanto, un problema fundamental es el de cómo asociar el umbral de un solo dispositivo a una señal distribuida como ésta. Ello puede lograrse añadiendo dispositivos de umbral  $T_1, T_2 \dots, T_i \dots, T_N$  al módulo básico representado en la fig. 11. Por ejemplo, el dispositivo de umbral  $T_i$  da una salida si el valor absoluto de su entrada  $|r_i|$  excede de algún valor prefijado

$$|r_i| > \theta_i,$$

10 siendo  $\theta_i$  el umbral de señal del dispositivo. Es importante hacer notar que la entrada primitiva o de origen  $s^d$  podría ser, bien una entrada primitiva procedente de  $E$ , o bien la salida de un módulo anterior.

15 Una entrada repetida del diseño o pauta  $s^d$  se convierte por tratamiento en el diseño de salidas  $s^N$  que, por repetición del argumento dado más arriba, crece exponencialmente. Como consecuencia, una entrada arbitraria pero repetida  $s^d$  llegará eventualmente a activar un dispositivo de umbral que responda a esa entrada. Es importante observar que:

- (a) la entrada no necesita ser conocida de antemano;
- (b) la entrada podría ir "sumergida" en ruido aleatorio; y

25

(c) tampoco necesita conocerse de antemano cuál es el dispositivo de umbral que responde al diseño  $s^k$ . (Con el algoritmo arriba indicado, la componente más grande de la respuesta  $r^k$  es la que determina cuál es el dispositivo que responde.) Mediante una sencilla variación, un dispositivo particular de umbral podría ser designado para responder a un diseño particular.

Con la adición de conexiones laterales como se indica en la figura 11, el "cebado" o activación de un solo dispositivo de umbral en respuesta al diseño  $s^k$  suprimiría entonces la respuesta de los demás dispositivos de umbral a este diseño. Si el parámetro  $\gamma$  fuese menor que la unidad durante el período de adquisición, y  $s^k \rightarrow T_i$ , la respuesta a  $s^k$  sería entonces modificada debido a la acción combinada de la declinación ( $\gamma < 1$ ) y de la supresión lateral, de modo que sólo la componente de orden  $i$  de la respuesta  $r^k$  a la entrada  $s^k$  seguiría siendo sustancialmente mayor que cero. En un estado final, se tendría:

$$s^k \rightarrow r_i \rightarrow T_i \rightarrow \text{señal de salida.}$$

Así, un solo dispositivo de umbral (o tantos como se quisiese) podrían responder a un solo diseño.

Además es útil no modificar más (permitiendo, sin embargo, su declinación) los mnemodadores asociados con el elemento de umbral de ordas  $i$  ( $1i, 2i \dots Ni$ ) cuando este elemento de orden  $i$  produce una señal de salida superior a un máximo especificado. Esto puede conseguirse finalizando las modificaciones en los mnemodadores antes mencionados para un número fijo

de ocurrencias cada vez que dicha salida supere el máximo especificado.

Si a un sistema como éste se le presentan señales separadas u ortogonales durante un período de aprendizaje, los dispositivos de umbral, con la rapidez exponencial anteriormente descrita, llegarán a ser activados por los diferentes diseños.

Así, por ejemplo N señales entrantes ortogonales repetidas darían tras el período de aprendizaje, una respuesta en N dispositivos de umbral diferentes.

De esta manera los dispositivos de umbral podrían aprender a responder a características repetidas del ambiente, aun cuando éstas no fuesen conocidas para el usuario.

Además, la asociación de estos dispositivos a unos diseños de salida de un módulo ya conocido serviría para la separación de sucesos o vectores arriba mencionada.

\* \* \*

En conclusión, de la explicación teoría de la presente invención arriba expuesta se desprende que el módulo de Nestor es un instrumento extremadamente poderoso para tratar información. En particular, este módulo es capaz de exhibir facultades de reconocimiento, recordación, generalización y asociación, según lo definido más arriba, sin necesidad de intervención humana en ningún proceso de decisión ni de aprendizaje. A continuación se describirá una forma específica de realización del módulo de Nestor en la que se utilizan sólo elementos de circuito de tipo usual y que se presta a una realización con métodos técnicos de microminiaturización ya conocidos.

## II. UNA FORMA ESPECIFICA DE REALIZACIÓN REPRESENTATIVA

Como se apreciará partiendo de la explicación teórica y estructural del presente invento arriba expuesta, la invención puede realizarse de varias maneras. En lo que sigue se da una descripción de lo que actualmente se considera una forma preferida de realización del aparato para poner en práctica la presente invención, aparato en el que se usan sólo componentes eléctricos normales, tales como resistencias, condensadores, diodos y transistores. Ahora bien, se sobrentiende que esta forma de realización preferida se describe solamente con fines de explicación y no con intención de limitar el ámbito del presente invento.

Como se recordará, el nouverón de orden o lugar  $i$  (fig. 2) de un módulo de Nestor comprende  $N$  entradas  $s_1, s_2, \dots, s_N$  que conducen a  $N$  mnemodadores  $i_1, i_2, \dots, i_N$  conectados a un sumador  $\Sigma_i$  que da una salida  $r_i$ . El mnemodador de orde  $ij$  tiene una función de transferencia  $A_{ij}$ ; esto es, la salida de este mnemodador es  $s_j' = A_{ij} s_j$ , donde  $s_j$  es la entrada de orden  $j$  al módulo de Nestor.

En la forma preferida de la invención, que se va a describir, se supondrá que las señales de información  $s_j, s_j'$  y  $r_i$  están representadas, en todos los casos, por unos niveles de tensión eléctrica o voltaje. Asimismo, como se recordará aun a riesgo de incurrir en repeticiones, se sobrentiende que las señales de información pueden estar

también representadas por variaciones en algún otro aspecto de las señales eléctricas del mundo real. Por ejemplo, las señales de información pueden estar representadas por frecuencias (frecuencias de repetición de impulsos o frecuencias sinusoidales), por anchuras de impulsos, por corrientes, por campos magnéticos, por magnetizaciones o incluso por una combinación de ellas.

Ahora bien, suponiendo que las señales de información vayan a estar representadas por tensiones, la función de transferencia  $A_{ij}$  puede considerarse como una "amplificación" (a veces amplificación, a veces reducción o disminución, a veces cambio de signo) y, en particular, como una "amplificación", que se va a modificar, dependiendo ello de la entrada  $s_j$  que llegue al mnemodador y de la salida  $r_i$  del nuevo-  
rón de orden  $i$ . En un algoritmo preferido, la función de modificación se describe mediante la ecuación siguiente:

$$\int A_{ij} = \eta r_i s_j .$$

Con el fin de realizar un amplificador cuya ganancia venga controlada por unos incrementos producidos en alguna ganancia precedente, es necesario proveer medios para almacenar información de la ganancia precedente y medios para sumar y/o restar incrementos a este valor almacenado. La función de almacenaje, en la forma preferida de realización del invento, viene realizada por un elemento, tal como un conden-

sador, destinado a almacenar una carga eléctrica  $Q$ ; la función de incrementación en esta forma de realización viene, pues, realizada por un aparato que haga variar la carga  $Q$  desde cero a  $+Q_0$  o  $-Q_0$ , límites del elemento de almacenaje.

5                    Con referencia ahora a la fig. 12, se representa en ella el mnemondedor 100 de orden  $ij$  del nouverón 102 de orden  $i$ . Una fuente de tensión 104 aplica una tensión  $V_{ij}$  por medio de una línea 106 a un amplificador 108, para controlar su ganancia  $A_{ij}$ . El amplificador 108, por lo tanto, modifica o hace variar la señal de tensión  $s_j$  que aparece  
10                    en la línea 110 y produce en la línea 112 una tensión de salida  $s_j' = A_{ij} s_j$ .

                    La tensión  $V_{ij}$ , que regula la ganancia, se deriva de  $Q_{ij}$ , la carga almacenada en el mnemondedor de orden  $ij$ ,  
15                    siendo, por lo tanto, proporcional a  $Q_{ij}$ . Las líneas 114 y 116 de trazo interrumpido indican, respectivamente, que esta carga puede hacerse variar (aumentarse o reducirse) según o con dependencia de la señal de entrada  $s_j$  que aparezca en la línea 110 y la respuesta de salida  $r_i$  que aparezca en  
20                    la línea 118. (La línea 116 se representa como yando a la fuente 104, con el fin de indicar simbólicamente la idea de modificación por parte de  $r_i$ ; en la presente forma de realización, la realimentación  $r_i$  está realmente dirigida a un dispositivo que modifica la anchura de los impulsos de entrada, de alturas  $s_1, s_2, \dots, s_N$ . Véase la fig. 14 y la descrip-  
25

ción de la misma.)

Para modificar  $Q_{ij}$  y, por lo tanto,  $V_{ij}$  y  $A_{ij} = s_j'/s_j$  con arreglo a la ecuación (5) arriba indicada, es necesario modificar la carga almacenada  $Q_{ij}$  por medio del producto  $r_i s_j$ .

5 [Nótese que las  $N$  cargas del nouverón de orden  $i$  ( a saber,  $Q_{i1}, Q_{i2} \dots, Q_{ij} \dots, Q_{iN}$ ) deben modificarse en proporción a las entradas respectivas que reciben (esto es, a  $s_1, s_2 \dots, s_j \dots, s_N$ ) y a la salida común  $r_i$  del nouverón.] Para conseguir la modificación deseada, se usa aquí como señal  $s_j$  la  
10 altura de un impulso de tensión 120, en tanto que la anchura  $t_i$  de este impulso se hace proporcional a  $|r_i|$ . Si  $r_i$  es menor que cero, se invierte el impulso de tensión 120. Así:

$$\delta Q_{ij} \sim s_j t_i$$

15

$$t_i \sim |r_i|$$

$$\frac{r_i}{|r_i|} = \begin{cases} +1 \rightarrow \text{no hay inversión} \\ -1 \rightarrow \text{hay inversión} \end{cases}$$

20

Los circuitos específicos para el aspecto como sumador del nouverón 102 y el mnemodador 100 se representan esquemáticamente en las figuras 13 y 14. El sumador 122 del nouverón 102 se representa en la figura 13 como un circuito clásico de sumación en el que se hace uso de un amplificador

25

operacional 124 conectado para recibir señales de entrada  $s_1', s_2' \dots, s_j' \dots, s_N'$  por medio de una pluralidad de resistencias valoradas por igual 126, 128 ..., 130 ..., 132, respectivamente. El amplificador operacional está provisto de una resistencia 134, de manera ya conocida, y da por la línea 136 una señal de salida  $r_i$  proporcional a la suma de las diversas señales  $s_1', s_2' \dots, s_j' \dots, s_N'$ , que son las salidas de los mnemodadores individuales del nouverón 102. La señal de salida  $r_i$  viene suministrada por una línea 138 de realimentación a cada mnemodador del nouverón.

El mnemodador de orden  $i_j$  representado en la fig. 14 recibe la señal de entrada por la línea 110 y da su señal de salida  $s_j'$  por la línea 112. La entrada de orden  $j$  está en la forma de un impulso de amplitud  $s_j$  (que puede ser positivo o negativo) y de una anchura de impulso normalizada, determinada por el ordenador o por la memoria compensadora de entrada que conecta el mnemodador, el nouverón y el módulo de Nestor al mundo exterior (el ambiente). Como se ha dicho más arriba, los dispositivos de entrada y salida empleados con el módulo de Nestor no forman parte alguna del presente invento, ni, por lo tanto, se describen aquí con detalle. La forma de construcción del ordenador de entrada dependerá, naturalmente, del tipo de sucesos que vayan a ser convertidos por tratamiento en el módulo de Nestor, en tanto que el ordenador de salida estará construido

con arreglo a los objetivos finales del sistema: esto es, la acción que se vaya a tomar en respuesta a las salidas del módulo de Nestor.

5 A los fines de esta descripción, basta decir que el ordenador de entrada suministra una pluralidad de impulsos de entrada a los terminales de entrada 1, 2 ..., j ..., N del módulo de Nestor. Los impulsos de entrada, que se suministran simultáneamente para cada suceso, tienen todos una anchura normalizada (por ejemplo, de varios microsegundos) y tienen amplitudes proporcionales a las variables  $s_1, s_2 \dots, s_j \dots, s_N$ , respectivamente, que se vayan a tratar. La amplitudes de impulso pueden ser negativas para reflejar valores negativos de las señales  $s_1, s_2 \dots, s_j \dots, s_N$ .

10 Con referencia ahora a la fig. 14, el impulso de entrada de amplitud  $s_j$  se aplica a un modulador 140 de anchura de impulso, que simultáneamente recibe por la línea 138 una señal  $r_i$  tomada de la salida del sumador  $\Sigma_i$  (esto es, el aparato ilustrado en la fig. 13). La señal  $r_i$  es el resultado de la sumación efectuada por el sumador  $\Sigma_i$  con las señales  $s_j$  entregadas por los mnemodadores.

15 El modulador de anchura de impulso transmite por la línea 142 un impulso que tiene la misma amplitud absoluta que el impulso de entrada presente en la línea 110 (pero invertida si  $r_i$  es menor que cero) y tiene una anchura de impulso  $t_i$  proporcional a la variable  $|r_i|$ .

Los impulsos positivos y negativos que aparecen en la línea 142 son amplificados y separados, con arreglo a la polaridad (signo), por medio de un amplificador operacional 144 y de sus resistencias 146, 148 y 150 y unos diodos 152 y 154 asociados al mismo. Si el impulso que aparezca en la línea 142 es positivo, se canaliza como impulso positivo hasta una línea 156; si es negativo, se canaliza como impulso negativo a una línea 158. Sea positivo o negativo, como se apreciará, el área comprendida bajo el impulso (visto éste gráficamente) es proporcional a la magnitud absoluta del producto de  $s_j$  y  $r_i$ .

Los impulsos positivos presentes en la línea 156 se aplican a un amplificador operacional 160 inversor, que incluye unas resistencias 162, 164 y 166, y finalmente a la base de un transistor 168 de tipo PNP. Los impulsos que llegan a la base del transistor 168 tienen así la polaridad apropiada (el signo adecuado) para activar el transistor y para cargar o descargar un condensador 170 por medio de una resistencia 172. La cantidad de carga depositada en el condensador 170 o retirada de éste por cada impulso positivo es proporcional al producto de la amplitud del impulso, que determina la conductancia efectiva del transistor 168, y la anchura del impulso, que determina la duración de la operación de carga o descarga.

Los impulsos negativos presentes en la línea 158

se suministran a un amplificador operacional inversor 174, con sus resistencias 176, 178 y 180. De manera análoga a la descrita para el transistor 168 de tipo PNP, se activa de ese modo un transistor 182 de tipo NPN, por medio de los impulsos negativos invertidos (esto es, positivos) aplicados a la base del transistor NPN. Como consecuencia, el condensador 170 se descarga o se carga a través de una resistencia 184. La cantidad de carga retirada del condensador 170 o añadida a éste por cada impulso negativo es proporcional al producto de la amplitud del impulso, que determina la conductancia efectiva del transistor 182, y la anchura del impulso, que determina la duración de la operación de descarga o de carga.

Como consecuencia de la operación arriba descrita, la carga  $Q_{ij}$  en bornes del condensador 170 y, a su vez, la tensión  $V_{ij}$ , es el resultado de un estado inicial de carga, que puede aplicarse a un terminal de entrada 186 antes de ponerse en funcionamiento el mnemodador, y de la suma total de todos los incrementos y decrementos que aparecen como resultado de la aplicación repetida de impulsos a la línea de entrada 110. Como se apreciará, el condensador 170 puede ser cargado con una u otra polaridad, así como cambiar de polaridad, dentro de los límites de las posibilidades de tensión positiva y negativa de la fuente de alimentación.

Con el fin de permitir que el mnemodador "olvide"

la información almacenada a lo largo de un período o intervalo de tiempo, la tensión  $V_{ij}$  puede dejarse declinar o decrecer a un régimen (velocidad) apropiado. Este régimen de decaimiento tiene relación con la constante de declinación  $\gamma$  arriba estudiada en la explicación teórica de este invento, en la que un tiempo infinito de declinación (a circuito abierto) es equivalente a  $\gamma = 1$ , y un tiempo de declinación cero (cortocircuito) es equivalente a  $\gamma = 0$ . Como se ha dicho, los valores de  $\gamma$  próximos a 1 son del máximo interés en la práctica, de manera que la constante de tiempo de declinación o decaimiento ha de hacerse bastante grande. A este fin, se seleccionan valores apropiados de capacidad para el condensador 170 y de resistencia o impedancia para todos los elementos, con el fin de obtener las constantes de tiempo deseadas.

Finalmente, la tensión  $V_{ij}$  que hay en bornes del condensador 170 se aplica por medio de una línea 192 a la entrada de control de un amplificador 194 controlado en ganancia. De ser necesario, puede intercalarse un paso de amplificación entre el condensador 170 y el amplificador de ganancia controlada 194. Aquí también se elige para este amplificador una impedancia de entrada adecuadamente alta, con el fin de obtener la constante de declinación  $\gamma$  deseada. El amplificador 194 recibe también los impulsos de amplitud  $s_j$  procedentes de la línea 110 y "amplifica" (aquí, de nuevo, el término "amplificación" incluye [ los conceptos de ] amplificación, dis-

minución y cambios de polaridad) estos impulsos, con arreglo al nivel controlado de ganancia, hasta producir impulsos de salida de amplitud  $s_j$  en la línea 112.

5 El tratamiento de las entradas  $s_j$  por medio de las salidas  $r_i$  sumadas se analiza a continuación con cierto detalle, haciendo referencia a la fig. 15. Es de notar en primer lugar que  $r_i$  puede ser positivo, negativo o cero, y que en ausencia de provisiones en contrario,  $r_i$ , tal como se obtiene a la salida del sumador 122 de la fig. 12, está en  
10 forma de impulsos. Si bien esta forma puede ser satisfactoria a los fines de los ordenadores que vayan a ir conectados a la línea 136 de la fig. 13, la operación de realimentación o retroacción se efectúa de manera más conveniente, en la presente forma de realización, por medio de una forma de  $r_i$  de  
15 tipo cuasi continua. A este fin, la realimentación indicada con la línea 138 en las figs. 13 y 14 va conectada a un dispositivo 196 de "seguimiento y retención" en la fig. 15. Este dispositivo tiene por objeto prolongar la duración de cada impulso de  $r_i$  durante un período aproximadamente igual a la separación en el tiempo entre impulsos consecutivos, sin  
20 modificar la amplitud de los impulsos. Como se indica esquemáticamente en la fig. 15, el dispositivo 196 de "seguimiento y retención" es puesto en acción por las entradas (señal  $s_j$  pulsada), a través del circuito de báscula 198. (El circuito  
25 de báscula 198 puede incluir provisiones para la amplifica-

ción de señales, su perfilamiento, etc., como resultará evidente para las personas versadas en la materia.) La salida pulsada  $r_i$ , por tanto, se convierte en una señal esencialmente continua de amplitud variable con el tiempo, a la que en lo sucesivo se designará aquí como  $r_i'$ . Esta señal  $r_i'$  es llevada a su vez a un rectificador 200 de "onda completa" y luego a un convertidor 202 de "tensión en anchura de impulso". Así, la señal cuasicontinua que llega al convertidor 202, en el punto 204, es positiva (o cero).

El convertidor 202, una vez activado por el circuito de báscula 198, según está indicado en la fig. 15, produce impulsos de una altura y una anchura normalizadas, siendo la anchura pequeña en comparación con la de las entradas pulsadas  $s_j$ . La señal  $r_i'$  tiene por función la de ensanchar estos impulsos estrechos (producidos por el convertidor 202) en proporción con la amplitud de  $r_i'$ . Los impulsos (de amplitud constante) producidos por el convertidor 202 se llevan a otro dispositivo de "seguimiento y retención" 205 y a una puerta 208, como se indica en la fig. 15. Las señales entrantes  $s_j$ , inicialmente de una anchura de impulsos normalizada, entran primero en el dispositivo de "seguimiento y retención" 206, que amplía o prolonga esta anchura de impulsos hasta llevarla a la duración, proporcional a  $r_i'$ , del impulso generado por el convertidor 202. Las señales de entrada  $s_j$  entran a continuación por la puerta 208, que se mantiene

abierta durante el mismo tiempo, determinado también por la anchura de impulso procedente del convertidor 202. Si, debido a ser pequeña la amplitud de  $r_i'$ , la anchura de estos últimos impulsos se reduce por debajo de la anchura normalizada de  $s_j$ , la puerta 208 permanece abierta sólo en la duración reducida de los impulsos que vienen del convertidor 202. Es este tiempo de apertura de la puerta 208 el que determina entonces la anchura de la señal de entrada  $s_j$ .

Con el fin de conservar el signo algébrico del producto  $s_j r_i'$ , según lo requerido por la teoría, la señal  $r_i'$  va canalizada desde el lugar 210, (donde todavía aparece con ambas polaridades) a un conmutador 212. En este conmutador se invierte la señal entrante  $s_j$  si  $r_i'$  es negativa, y se deja pasar con la misma polaridad que trae si  $r_i'$  es positiva. El conmutador 212 ha sido colocado detrás de la puerta 208 en la fig. 15, por conveniencia de la descripción secuencial. Es técnicamente preferible que este conmutador se coloque delante del dispositivo 206 de "seguimiento y retención".

Finalmente, entre el lugar 204 y la puerta 208 se incluye una puerta de coincidencia (puerta Y) 214 para suprimir el paso de la señal entrante (y con ello del producto  $s_j r_i'$ ) cuando  $r_i'$  sea tan pequeña que pida una anchura de impulso menor o más estrecha que la del impulso normalizado generado por el convertidor 202. Es de notar que de esa mo-

do se produce en el producto  $s_j r_i$  una discontinuidad entre el valor determinado por los impulsos más estrechos obtenidos del convertidor 202, y cero. Este rasgo característico no es sino un aspecto de lo que en términos generales podría denominarse "ruido" y, como puede deducirse de la teoría, el módulo de Nestor es particularmente invulnerable al ruido, de hecho, al funcionamiento imperfecto de los componentes individuales. Además, es ya bien conocido en el campo de los circuitos electrónicos el recurso de mejorar la relación de señal a ruido en general, y en particular de reducir la discontinuidad mencionada.

Es de notar asimismo que, como las modificaciones de la función  $A_{ij}$  dependen sólo de las señales de entrada al nouverón del cual forma parte el mnemodador en cuestión, y de la respuesta de salida del mismo, las anchuras de impulso y la posible inversión debida a  $|r_i|$  y a  $r_i/|r_i|$  tienen aplicación de igual manera a cada una de las señales pulsadas  $s_1, s_2 \dots, s_N$  que entran en el nouverón. Por lo tanto, la mayoría de los circuitos o medios electrónicos indicados en la fig. 15, como resultará evidente para los técnicos en la materia, sirven al mismo tiempo para todos los mnemodadores de un nouverón y, por tanto, se necesitan sólo una vez para cada nouverón.

Se termina así la descripción de la forma específica de realización del presente invento. Si bien esta forma

de realización se ha descrito haciendo referencia a señales y cargas eléctricas como portadoras del contenido informativo de un sistema de tratamiento de información, a las personas versadas en la materia se les ocurrirán otros numerosos métodos para representar señales y para almacenar información. También se sobrentiende que la presente invención es susceptible de por sí de diversos cambios, modificaciones y adaptaciones que caen dentro de su ámbito y de su espíritu. Por ejemplo, el algoritmo para las modificaciones de A de los mnemodadores no tiene por qué limitarse al cuarto término de la serie de Taylor -- es decir, al término incorporado a la ecuación (5) precedente; antes bien, otros términos, tales como el sexto, pueden dar resultados igualmente poderosos. Por consiguiente, se tiene la intención de que el presente invento sólo pueda considerarse limitado por las reivindicaciones que siguen y sus equivalentes.

La presente solicitud, que corresponde a la presentada en Estados Unidos de América, el 6 de Junio de 1974, bajo el número 477.000, se acoge a los beneficios del Artículo 51 del vigente Estatuto sobre Propiedad Industrial.

## - REIVINDICACIONES -

10

Los puntos de Invención propia y nueva que se presentan para que sean objeto de esta solicitud de Patente de Invención en España, por VEINTE años, son los que se recogen en las reivindicaciones siguientes:

15

1a.- Un método de tratar información, que comprende las etapas de: (a) recibir una pluralidad de señales de entrada  $s_1, s_2 \dots, s_j \dots, s_N$ ; (b) producir una pluralidad de señales intermedias  $s'_{i1}, s'_{i2} \dots, s'_{ij} \dots, s'_{iN}$ , de las cuales cada  $s'_{ij}$  depende de una (respectiva) de dichas señales de entrada  $s_j$  y de una función de transferencia  $A_{ij}$  asociada; (c) producir una respuesta de salida  $r_i$  que depende de por lo menos una de dichas señales intermedias  $s'_{ij}$ ; y (d) modificar por lo menos una de dichas funciones de transferencia  $A_{ij}$ , cuando se está en un

20

25

modo de "aprender", según el producto de por lo menos una de dichas señales de entrada  $s_j$  y dicha respuesta de salida  $r_i$ .

5           2a.- El método de la reivindicación 1a, en el que cada señal intermedia  $s'_{ij}$  es igual al producto de la respectiva señal de entrada  $s_j$  y la función de transferencia  $A_{ij}$  asociada ( $s'_{ij} = A_{ij} s_j$ ).

10           3a.- El método de la reivindicación 1a, en el que dicha respuesta de salida  $r_i$  depende de todas las citadas señales intermedias  $s'_{ij}$ .

4a.- El método de la reivindicación 1a, en el que dicha respuesta de salida  $r_i$  es proporcional a la suma de una pluralidad de dichas señales intermedias  $s'_{ij}$ .

15           5a.- El método de la reivindicación 4a, en el que dicha respuesta de salida  $r_i$  es proporcional a la suma de todas las citadas señales intermedias  $s'_{ij}$ .

20           6a.- El método de la reivindicación 1a, en el que la etapa (d) incluye la etapa de modificar cada una de dichas funciones de transferencia  $A_{ij}$  según el producto de por lo menos una de dichas señales de entrada  $s_j$  y la citada respuesta de salida  $r_i$ .

25           7a.- El método de la reivindicación 1a, en el que la etapa (d) incluye la etapa de modificar por lo menos una de dichas funciones de transferencia  $A_{ij}$  según el producto de la señal de entrada  $s_j$ , asociada a ella, y la

citada respuesta de salida  $r_i$ .

5           8a.- El método de la reivindicación 7a, en el que la etapa (d) incluye la etapa de modificar cada una de dichas funciones de transferencia  $A_{ij}$  según el producto de la señal de entrada  $s_j$ , asociada a ella, y dicha respuesta de salida  $r_i$ .

10           9a.- El método de la reivindicación 1a, en el que la etapa (d) incluye la etapa de modificar por lo menos una de dichas funciones de transferencia  $A_{ij}$  en proporción con el producto de la señal de entrada  $s_j$ , asociada a ella, y dicha respuesta de salida  $r_i$ .

15           10a.- El método de la reivindicación 1a, en el que la etapa (d) incluye la etapa de modificar cada una de dichas funciones de transferencia  $A_{ij}$  en proporción con el producto de la señal de entrada  $s_j$ , asociada a ella, y la respuesta de salida  $r_i$ .

20           11a.- El método de la reivindicación 1a, en el que la etapa (c) incluye la etapa de producir una pluralidad de respuestas de salida  $r_1, r_2 \dots, r_i \dots, r_n$ , de las cuales cada respuesta  $r_i$  depende de por lo menos una de dichas señales intermedias  $s'_{ij}$ , en la que  $j$  varía de 1 a N.

25           12a.- El método de la reivindicación 11a, en el que la etapa (c) incluye la etapa de producir una pluralidad de respuestas de salida  $r_1, r_2 \dots, r_i \dots, r_n$ , de las cuales cada respuesta  $r_i$  depende de todas las citadas seña-

les intermedias  $s'_{ij}$ , en las que  $j$  varía de 1 a N.

5 13a.- El método de la reivindicación 11a, en el que la etapa (c) incluye la etapa de producir una pluralidad de respuestas de salida  $r_1, r_2 \dots, r_i \dots, r_n$ , de las cuales cada respuesta  $r_i$  es proporcional a la suma de una pluralidad de dichas señales intermedias  $s'_{ij}$ , en las que  $j$  varía de 1 a N.

10 14a.- El método de la reivindicación 13a, en el que la etapa (c) incluye la etapa de producir una pluralidad de respuestas de salida  $r_1, r_2 \dots, r_i \dots, r_n$ , de las cuales cada respuesta  $r_i$  es proporcional a la suma de todas las citadas señales intermedias  $s'_{ij}$ , en las que  $j$  varía de 1 a N.

15 15a.- El método de la reivindicación 11a, que comprende además la etapa de suprimir todas las respuestas de salida  $r_1, r_2 \dots, r_n$ , excepto la respuesta  $r_i$  que exceda de un umbral prescrito.

20 16a.- El método de la reivindicación 11a, que comprende además la etapa de fijar por lo menos una de dichas respuestas de salida  $r_i$  a un valor igual a una respuesta de salidas  $r_{iA}$  específica deseada.

25 17a.- El método de la reivindicación 11a, que comprende además la etapa de fijar cada una de dichas respuestas de salida  $r_1, r_2 \dots, r_i \dots, r_n$  a un valor respectivamente igual a una respuesta de salida  $r_{1A}, r_{2A} \dots, r_{iA} \dots, r_{nA}$  específica deseada.

18a.- El método de la reivindicación 1a, que comprende además la etapa de guardar cada una de dichas funciones de transferencia  $A_{ij}$  en un dispositivo de almacenaje.

5 19a.- El método de la reivindicación 18a, que comprende además la etapa de transferir por lo menos una de las funciones de transferencia  $A_{ij}$  almacenadas a otro dispositivo de almacenaje.

20a.- El método de la reivindicación 18a, que comprende además la etapa de transferir a otro dispositivo de almacenaje cada una de las funciones de transferencia  $A_{ij}$  almacenadas.

21a.- El método de la reivindicación 1a, que comprende además las etapas de: (1) recibir una segunda pluralidad de señales de entrada  $s^2_1, s^2_2, \dots, s^2_j, \dots, s^2_N$ ; (2) producir una segunda pluralidad de señales intermedias  $s'^2_{i1}, s'^2_{i2}, \dots, s'^2_{ij}, \dots, s'^2_{iN}$ , de las cuales cada señal  $s'^2_{ij}$  dependen de una (respectiva) de dichas segundas señales de entrada  $s^2_j$  y de dicha función de transferencia  $A_{ij}$  asociada a una (respectiva) de las primeras señales intermedias  $s'_{ij}$  producidas en la etapa (b); y (3) producir una segunda respuesta de salida  $r^2_i$  que depende de por lo menos una de dichas segundas señales intermedias  $s'^2_{ij}$ .

22a.- UN METODO DE TRATAR INFORMACION.

25 Tal y como se ha descrito en la Memoria que an

tecede, representado en los dibujos que se acompañan y con los fines que se han especificado.

Esta Memoria consta de ochenta y tres hojas es critas a máquina por una sola cara.

5

Madrid, 21. MAR 1977

P.A. Oscar de Elizaburu  
Por Poder. *[Signature]*

10

15

20

25

14.5.75

- 83 -

FMM./

FIG. 1

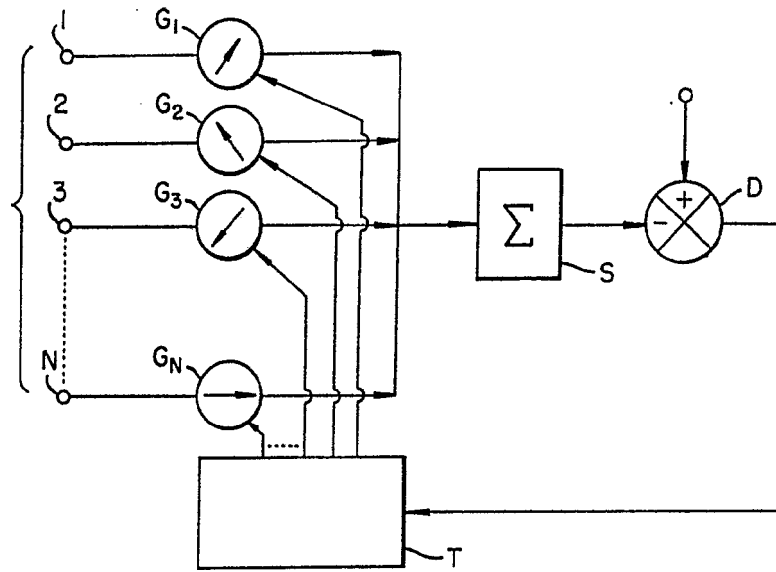
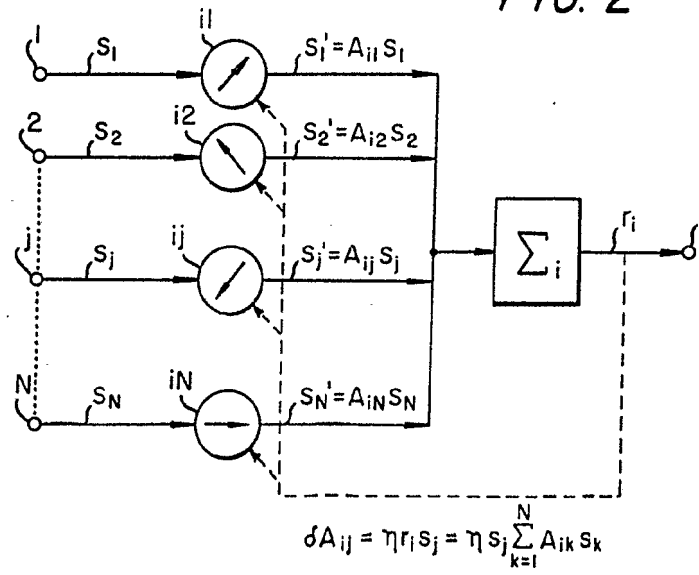


FIG. 2



Oscar de Elzaburro  
Por Poder




FIG. 3

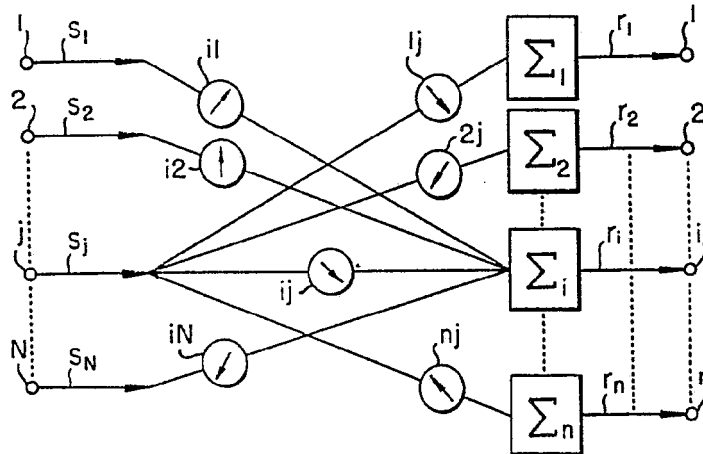


FIG. 4

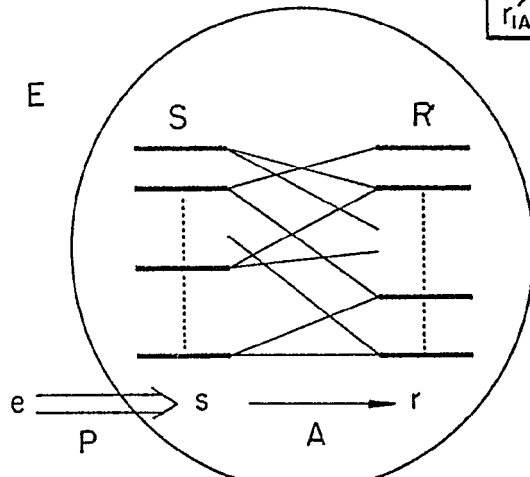
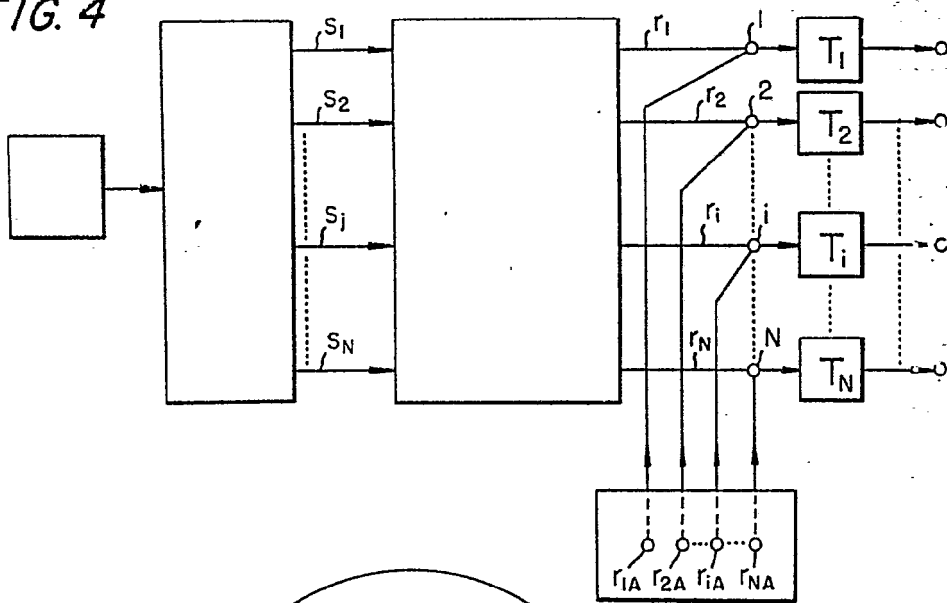


FIG. 5

Oscar de Vizcarra  
Por Poder

FIG. 6

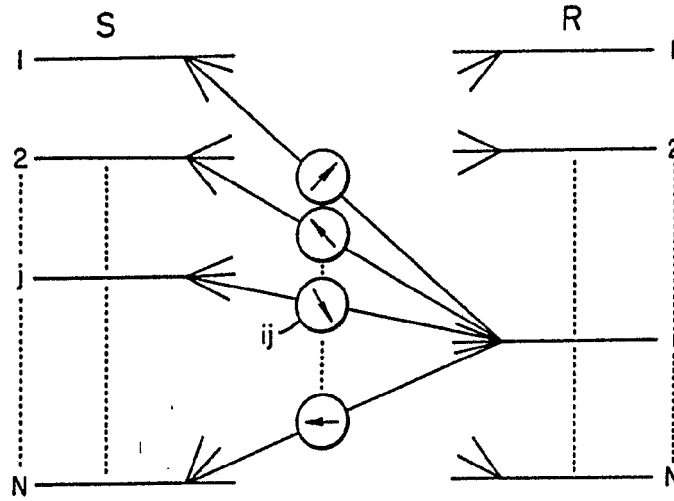


FIG. 7

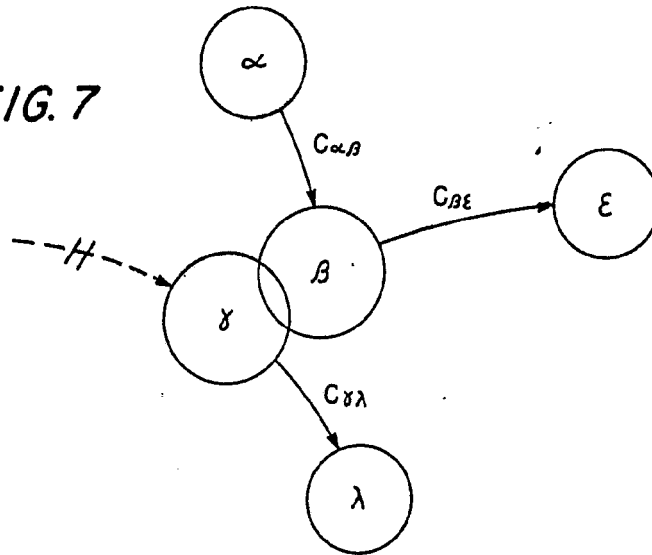
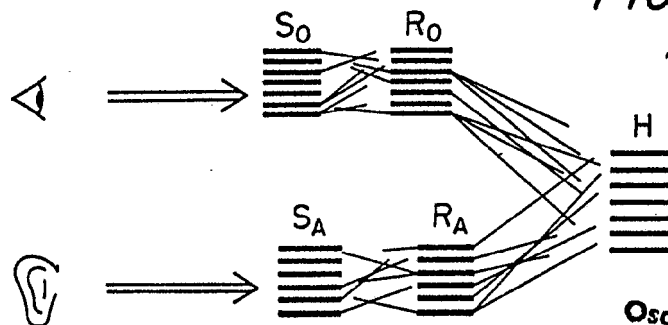


FIG. 8



Oscar de Elzabur  
For Poder.

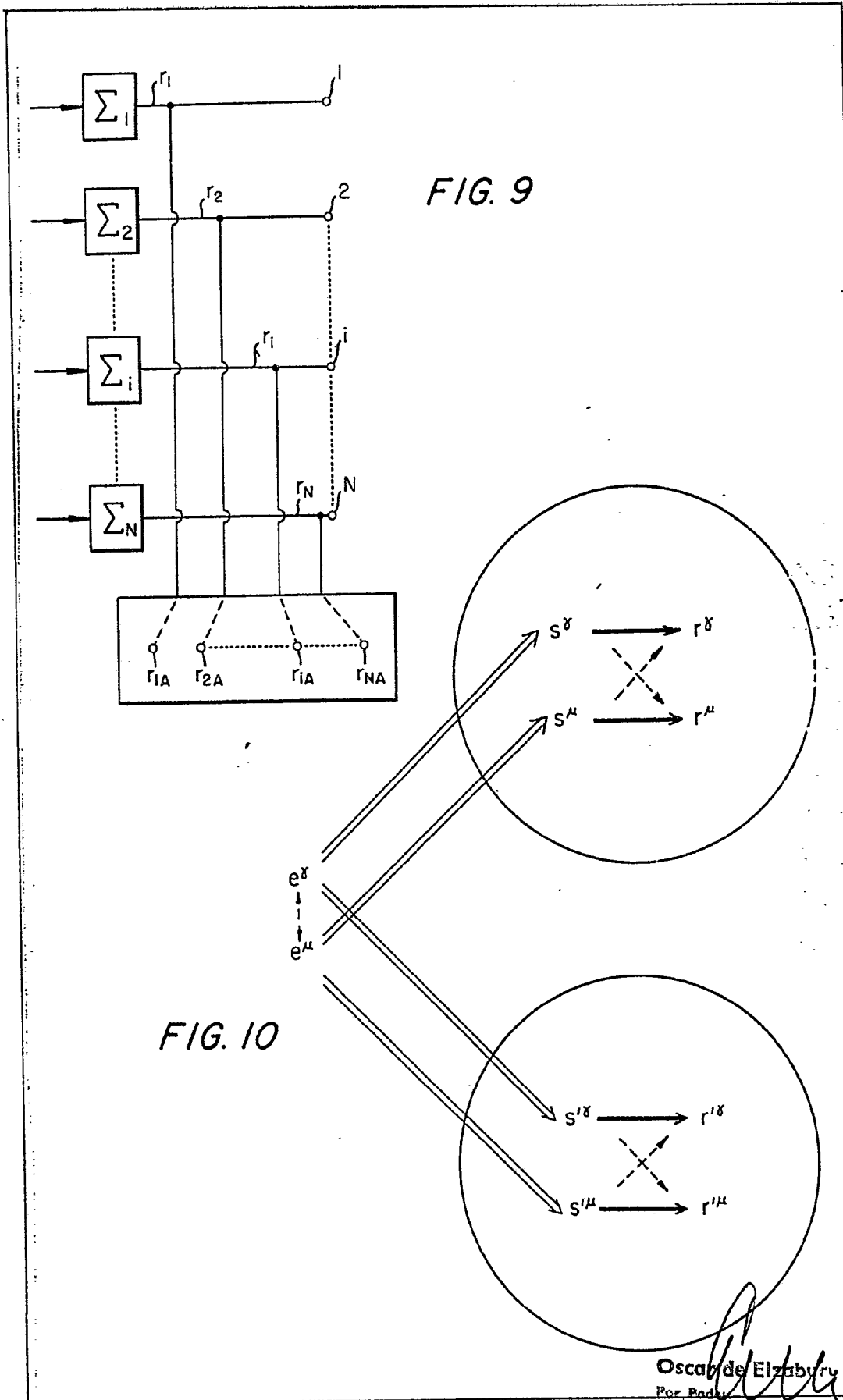
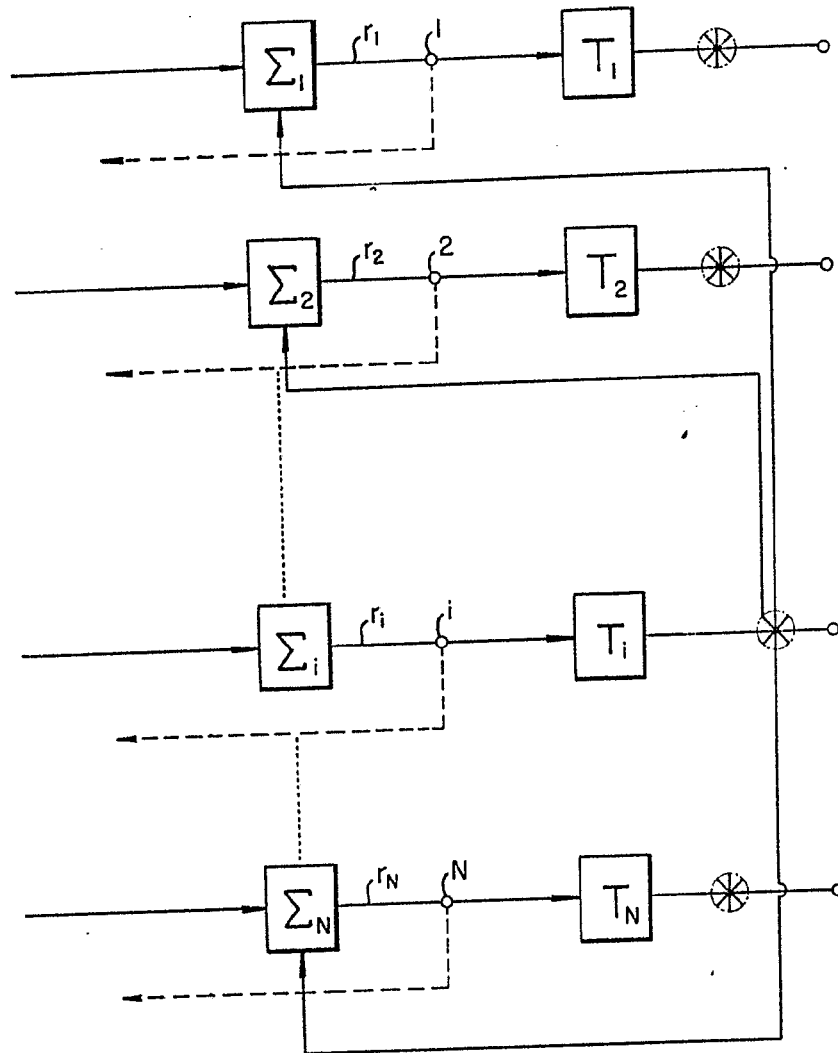


FIG. 9

FIG. 10

Oscar de Elizaburu  
 For. Bodu

FIG. 11



Oscar de Elzaburu  
 Por Padre

FIG. 12

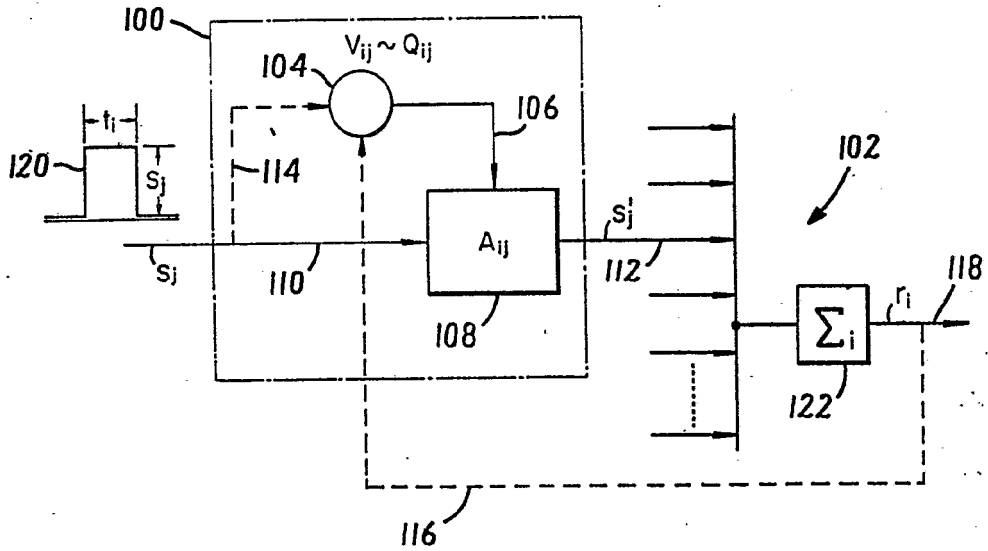
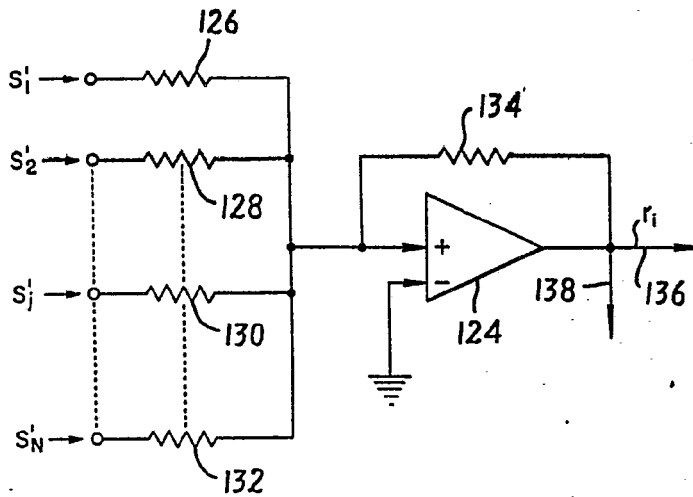


FIG. 13



Oscar de Elzaburu  
Por Poder.

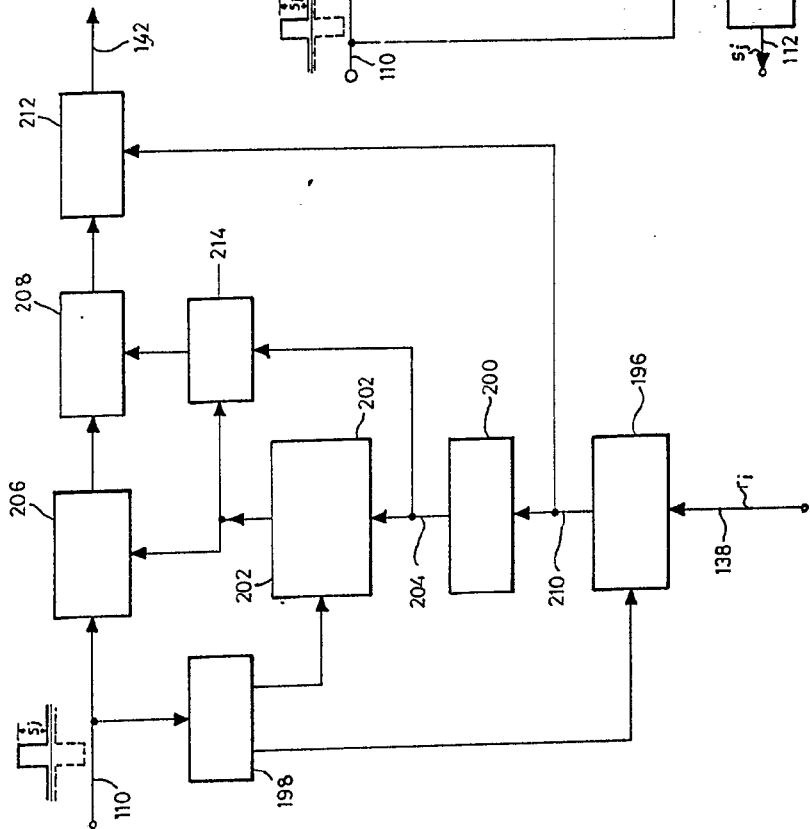
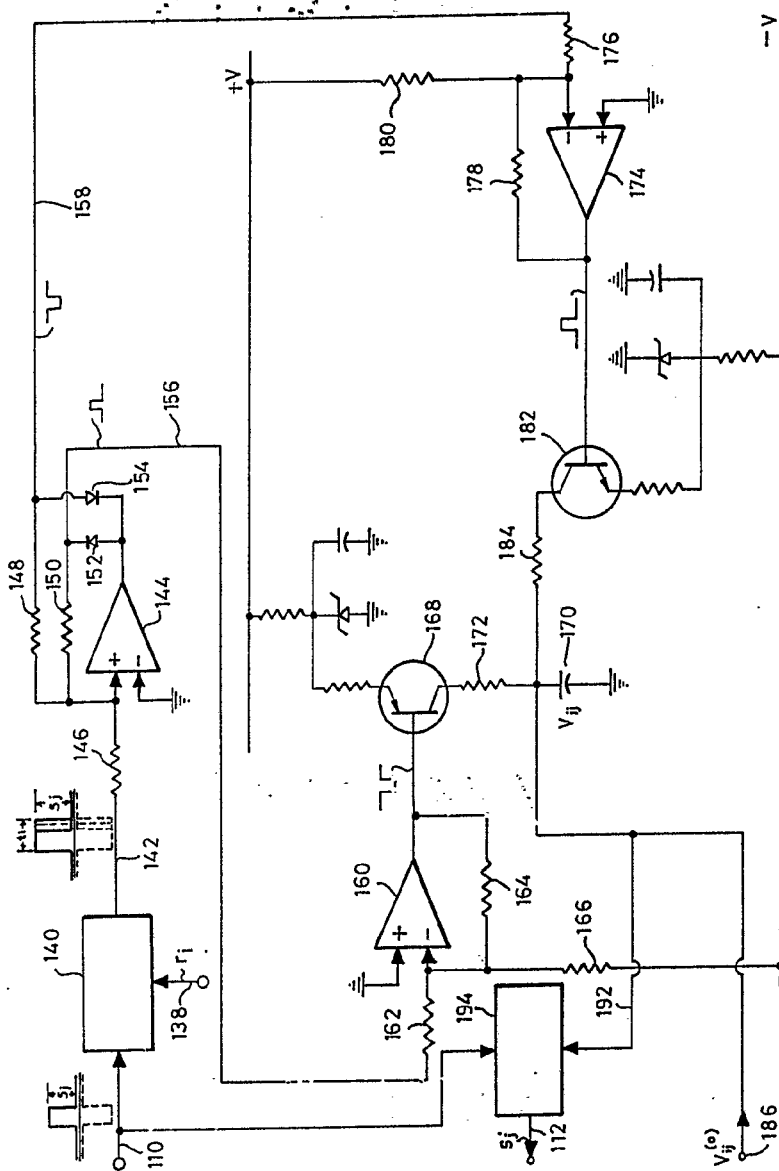


FIG-15

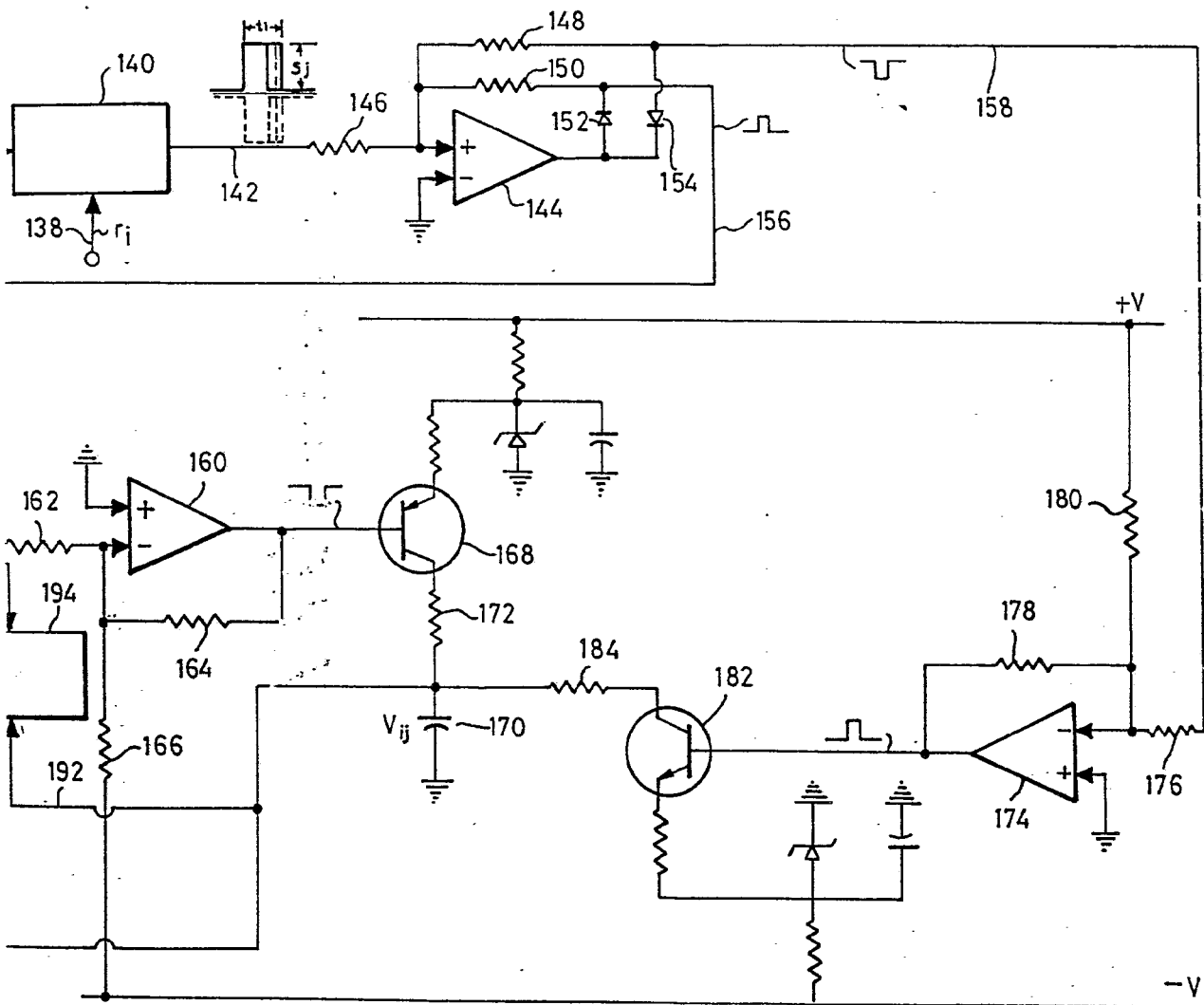
FIG-14



Oscar S. Elzohary  
For Patent



FIG. 14



Oscar de Ezaburu  
Por Poder