

REAL ACADEMIA DE CIENCIAS
EXACTAS, FÍSICAS Y NATURALES

DISCURSO INAUGURAL

DEL AÑO ACADÉMICO 1970-1971

LEIDO EN LA SESIÓN CELEBRADA EL DÍA 25 DE NOVIEMBRE DE 1970

POR EL ACADÉMICO NUMERARIO

EXCMO. SR. D. JOSE GARCIA SANTESMASES



M A D R I D

DOMICILIO DE LA ACADEMIA:

VALVERDE, 22.—TELÉFONO 221-25-29

1 9 7 0

Depósito Legal: M. 26.006-1970.

TALLERES GRÁFICOS VDA. DE C. BERMEJO.—J. GARCÍA MORATO. 122.—MADRID

CIBERNÉTICA Y PROCESO DE LA INFORMACIÓN VISUAL EN LOS SERES VIVOS Y EN LAS MÁQUINAS

Excmos. Sres. Académicos:

Señoras y señores:

Sean mis primeras palabras de agradecimiento a la Real Academia por encargarme la honrosa tarea de pronunciar el discurso inaugural de este año académico.

He escogido como tema de mi disertación «La Cibernética y el proceso de la información visual en los seres vivos y en las máquinas», campo de gran interés en sus dos vertientes y en cuyo estudio y desarrollo la aportación de la Cibernética y, podríamos añadir también, el soporte de la Biónica son, sin duda, imprescindibles. Por ello, antes de entrar realmente en el tema, desearía decir unas palabras sobre el estado actual y perspectivas de estas últimas disciplinas.

La Cibernética nació formalmente en 1949 al publicar Wiener su famoso libro con la misma denominación [1]. Como es bien sabido, este término deriva del vocablo griego κυβερνήτης que significa *timonel*. Wiener define la Cibernética como «la ciencia del control y de la comunicación en el animal y en la máquina» (*). Posteriormente se han propuesto otras definiciones pero todas, en

(*) Es de justicia y para mí muy grato señalar las valiosas aportaciones en este campo de nuestro ilustre y llorado compañero Pedro Puig Adam, que expuso magistralmente en su bello y documentado discurso de ingreso en esta Academia, titulado «Matemática y Cibernética».

general, quedan englobadas en la inicial dada por el fundador de esta disciplina.

Creo interesante recordar que el origen de esta ciencia hay que buscarlo en el planteamiento de problemas técnicos. Durante la II guerra mundial, Wiener fue encargado por el Gobierno americano de estudiar la posibilidad de regular automáticamente la dirección de tiro de cañones antiaéreos. Estas investigaciones le condujeron a los sistemas de realimentación y a preguntarse si los neurofisiólogos no se encontraban también con problemas similares, es decir, con procesos de realimentación en los seres vivos. Sus discusiones sobre este tema, primero con el biólogo Rosenblueth y más tarde con un grupo de científicos, le llevaron a la conclusión de que existía, en efecto, una unidad esencial en los problemas del control y de la comunicación en los organismos vivos y en las máquinas. La formulación de leyes generales dio lugar a implicaciones de tipo filosófico. El gran mérito de Wiener reside, sin duda, en avizorar claramente, desde el principio, la filosofía inherente a la idea misma de la Cibernética y exponer con valentía sus consecuencias bajo el punto de vista biológico y social. Tuvo la suerte de vivir lo suficiente para desarrollar sus propias ideas y verlas florecer en sus seguidores.

Actualmente existe una tendencia, bastante generalizada, a considerar la Cibernética como la ciencia que investiga las leyes generales sobre sistemas complejos de control y de tratamiento de la información. Según Pask, sea cual fuere el sistema estudiado, su objetivo es investigar cómo se autorregula, cómo se reproduce, como evoluciona y cómo aprende. En definitiva, estudia cómo se autoorganizan los sistemas. Por ejemplo, un modelo de cerebro no siempre es un modelo cibernético ; pero lo será en tanto manifieste de forma explícita algún aspecto de este modelo que tenga importancia en su propia regulación.

Existe también otra tendencia, más bien filosófica, que considera como misión de la Cibernética la síntesis de gran número de ciencias puras y aplicadas para establecer la unidad esencial entre lo animado y lo inanimado [2]. Sea cual fuere el punto de vista adoptado y exceptuando algunas extrapolaciones discutibles, am-

bas tendencias se hallan en línea con el amplio significado dado a esta ciencia por el propio Wiener.

La Cibernética ha sido muy discutida. Al lado de seguidores excesivamente entusiastas, ha tenido también sus adversarios. A los primeros su ardiente fe en estos principios les ha llevado a extrapolar, a considerar como una realidad lo que sólo es una fantasía, a confundir analogías entre máquinas y sistema nervioso con identidades. Como todas las grandes ideas, ha tenido también sus detractores: hay que reconocer que muchos lo han sido por desconocer los textos originales que han llegado a ellos muy mutilados y deformados por autores de pseudociencia. Afortunadamente, los alegatos de unos y otros no afectan a la gran obra de Wiener (ni a los principios en ella implicados) que permanece intangible, como una de las importantes contribuciones al conocimiento humano.

El campo de la Cibernética es tan amplio, las materias comprendidas inicialmente bajo esta denominación han crecido tan deprisa y han adquirido un volumen y una personalidad tan grandes, que han surgido nuevas denominaciones tales como «Sistemas de inteligencia artificial», «Sistemas autoorganizados», «Sistemas adaptivos», etc., cuyos contenidos encajan tanto dentro de la Cibernética como de la Automática y, hasta en ciertos aspectos, de la Informática, pudiendo crear un estado de confusión al no especializado. Conviene hacer esta observación no para discriminar y señalar los objetivos y alcance de las materias que se definen bajo estos títulos, sino para demostrar que sus límites no están estrictamente determinados, que hay solapamientos entre las materias de aquellas denominaciones. A las ya nombradas hay que añadir otra que creemos de gran interés y que ha nacido hace pocos años, concretamente en 1960, y cuyo nombre es el de *Biónica* [3]. Esta denominación fue acuñada por el mayor J. E. Steele de las Fuerzas Aéreas americanas. Se puede definir como la ciencia que trata del estudio de los principios y características de los organismos vivos con objeto de aplicarlos a la resolución de problemas técnicos.

Los objetivos de esta rama de la Ciencia, básicamente inter-

disciplinaria, son muy ambiciosos y su campo de actividad muy amplio ; se investigan posibles aplicaciones a los sistemas de inteligencia artificial mediante la simulación de neuronas y redes neuronales ; se estudian los mecanismos sensoriales y su simulación, así como las funciones de autoadaptación de los organismos vivos y su aplicación al guiado y control. Los elementos sensoriales de ciertos animales (vista, oído, etc.) son extraordinariamente sensibles, a veces superiores a los del hombre y, por consiguiente, su estudio ofrece un gran aliciente como fuente de información para intentar la realización de máquinas y aparatos de automática provistos de elementos receptores que permitan captar y después interpretar lo percibido del mundo exterior.

A pesar de los progresos evidentes de los sistemas electrónicos actuales, es posible que en el futuro se tengan que investigar nuevos derroteros, principios distintos de los clásicos, si se desea llegar a máquinas francamente superiores a las actuales. La Biónica ofrece amplios horizontes porque la naturaleza viva pone a nuestra disposición modelos muy perfeccionados y, por consiguiente, un manantial inagotable para nuevas ideas.

En este sentido hay que destacar, por su gran interés, sus posibles aplicaciones en Astronáutica. En efecto, en los vehículos espaciales existen calculadoras y otros aparatos automáticos que deben poseer gran seguridad funcional con dimensiones reducidas. Pero, probablemente, en el futuro se necesitarán además otras condiciones para realizar progresos tangibles en Astronáutica. Nos referimos a que los sistemas automáticos antes citados posean facultad de autoadaptación a medios distintos y difíciles de prever ; a que su reacción ante estos ambientes desconocidos sea variar su estructura para adaptarse a los mismos, etc. La Biónica, estudiando y desentrañando los mecanismos de adaptación de la naturaleza viva, puede señalar caminos para ello.

Muchas son las propiedades de los organismos vivos cuya investigación puede ser útil desde el punto de vista tecnológico, como por ejemplo los sistemas biológicos de propulsión, pero existen otros campos que se refieren principalmente a los procesos de apren-

dizaje, reconocimiento de formas, autoorganización, etc., en los que su investigación puede ser más fructífera.

¿Cuáles son los métodos de trabajo en Biónica? ¿Cuáles son los procedimientos de investigación para lograr sistemas tecnológicos de interés partiendo del estudio de los seres vivos? Evidentemente no hay regla única; cada problema que se presenta es distinto. Pero sin embargo sí se puede señalar que la filosofía de la Biónica no es realizar simples copias del proceso que se trata de simular. Ello podría ser, en ocasiones, muy difícil y no por eso de resultados más útiles. Lo importante es comprender el principio de funcionamiento y partiendo de ello, en general, serán necesarios grandes esfuerzos, bajo el punto de vista matemático, físico y tecnológico, para llegar a proyectar y desarrollar modelos que respondan de forma análoga al organismo vivo. En algunos casos, este proceso da lugar a simulaciones que tienen características inferiores al prototipo vivo, pero en otros éste está superado por las condiciones de funcionamiento del prototipo artificial. Cuando se trata de modelos complejos, como por ejemplo el sistema nervioso, la información que se puede extraer es, desgraciadamente, pequeña y se refiere a la definición cuantitativa del problema y a la prueba de que existe una solución.

Llegados a este punto, cabe preguntarse: ¿cuál es la diferencia entre Biónica y Cibernética?

Es difícil establecer una frontera entre ambas ciencias pues tienen un punto en común: el estudio de los organismos vivos. La diferencia hay que verla más en los objetivos que se proponen que en las bases científicas interdisciplinarias que constituyen su estructura. La Cibernética, según hemos visto, a pesar de su universalidad, o tal vez por eso mismo, estudia e investiga (además de los sistemas físicos) en modelos de los organismos vivos, con objeto de explicar estos sistemas y hallar nuevas ideas para nuevos experimentos biológicos. La Biónica tiene un objetivo muy concreto: estudia e investiga en modelos de organismos vivos y en muchas ocasiones con métodos totalmente cibernéticos, pero su misión última es extraer de estos experimentos, de estos modelos, ideas para nuevas máquinas, para nuevas aplicaciones técnicas.

Podríamos considerar la Cibernética y la Biónica como las dos caras de una misma moneda, en lo que se refiere a los sistemas biológicos; una trata de explicarlos y otra de plasmarlos en sistemas físicos.

RECONOCIMIENTO DE FORMAS VISUALES

Bajo el punto de vista de los ordenadores electrónicos, es evidente la importancia que presenta el estudio del reconocimiento de formas (*), ya que constituye el proceso de percepción mediante el cual la máquina puede *ver* y *oir*. Una de las dificultades con que se tropieza para lograr ordenadores electrónicos potentes estriba, precisamente, en obtener medios más perfeccionados de este proceso de percepción. En efecto, actualmente los datos, sea cual fuere su procedencia (tablas estadísticas, columnas de cifras, etc.), deben introducirse en forma de cintas o tarjetas perforadas y cintas magnéticas. Si se compara la velocidad con que se realiza esta entrada de datos con la operativa de la propia máquina, se observa que aquélla es muy pequeña respecto a ésta. Ello influye en la rapidez global de la máquina cuando el número de datos es muy grande, como ocurre por ejemplo en los procedimientos de traducción automática o de recuperación de información. En estos casos y en otros muchos se comprende el interés de que la máquina lea los textos que constituyen la entrada de la calculadora, reconociendo directamente los caracteres impresos en lugar de transcribirlos primeramente a la cinta perforada o magnética.

Recordemos ahora el proceso de percepción en el hombre. Para clasificar los objetos que se presentan a nuestra vista, observamos la forma de cada uno de ellos en conjunto, globalmente y además apreciamos ciertos detalles característicos. Esto se verifica casi ins-

(*) El vocablo «forma» o también «escena» que utilizamos en el texto, lo mismo que el inglés «pattern» (modelo, muestra), no expresa (en una y otra lengua) toda la riqueza y amplitud de su contenido. Con la frase «reconocimiento de formas (o escenas)» se debe entender, en el caso más general, «reconocimiento de configuraciones» que pueden ser complejas y que pueden proceder de estímulos acústicos o visuales, pero también se puede referir a otro tipo de símbolos, como por ejemplo los que integran el lenguaje.

tantáneamente pero, en realidad, es un proceso en el que interviene, primero el ojo y después el cerebro. De esta manera distinguimos dos objetos diferentes y por otro lado clasificamos dentro de la misma categoría dos objetos, aunque ofrezcan ligeras diferencias, mientras sus elementos esenciales permanezcan idénticos.

¿Cómo reconocemos una figura, por ejemplo una letra determinada, sea cual fuere su posición, sus dimensiones, aunque esté sometida a rotaciones o desplazamientos? ¿Cómo identificamos una cara conocida aunque haga tiempo que no la hemos visto? ¿Cómo percibimos la escritura a mano e incluso llegamos a determinar el autor de la misma? Esta aptitud de percibir lo que se denomina «universales» no sólo es prerrogativa del hombre, sino que muchos animales la poseen. Entre sus cualidades para percibir formas visuales podemos señalar las siguientes:

- a) Los animales pueden reconocer formas independientemente de su localización en el campo visual.
- b) El reconocimiento se produce independientemente del ángulo de inclinación de la figura en el campo visual.
- c) Las dimensiones de la figura no interfieren en el reconocimiento de la forma.
- d) Tanto los seres humanos como los animales tienden a confundir las imágenes reales con las reflejadas en un espejo.
- e) Organismos primitivos, en lo que se refiere a la percepción visual, como la rata y el octopo, tienen dificultad en distinguir entre el cuadrado y el círculo [4].

Estos problemas de «universales» son los más difíciles de explicar y, por consiguiente, el obstáculo más grande al progreso o perfeccionamiento de los sistemas o artificios de reconocimiento de escenas.

Pitts y McCulloch publicaron un interesante trabajo [5] en el que proponen un posible modelo neuronal que permitiría explicar estas propiedades de los organismos vivos de reconocer universales.

«Estas estructuras neuronales detectan la equivalencia de formas relacionadas por semejanza y congruencia, como por ejemplo un mismo objeto físico visto desde planos distintos.» Estas formas equivalentes comparten, en todos los casos, una figura común y

definen un grupo de transformaciones que permiten pasar de una forma equivalente a otra, conservando la figura invariante.

En el trabajo de McCulloch se proponen métodos generales para lograr redes neuronales que puedan reconocer figuras que den lugar a la misma salida para cualquiera de las entradas pertenecientes a la misma figura. Es decir, si por ejemplo consideramos una figura sencilla, la letra A, ésta se presenta invariante para varios grupos de transformaciones a que se someta: rotaciones, desplazamientos, variación de dimensiones, etc.

La organización neuronal —en realidad un autómata finito— descrita en dicho trabajo muestra algunas analogías con la estructura del cerebro. Sin embargo no se ha podido probar neurofisiológicamente que dicha estructura sea utilizada por el cerebro humano. De todas formas, el parecido superficial es tan grande que dio lugar a la siguiente anécdota: cuando Wiener mostró un modelo de esta estructura al neurofisiólogo Von Bonin, éste preguntó ¿es el diagrama de la cuarta capa de la corteza visual?

En este trabajo vamos a exponer, en líneas generales, las tendencias actuales y resultados obtenidos en el estudio del proceso de datos visuales en los seres vivos y en las máquinas bajo el punto de vista cibernético. Por lo que respecta a los primeros, dedicaremos particular atención a aquellos modelos que, partiendo de resultados experimentales obtenidos por neurofisiólogos, han dado lugar a una interpretación matemática y física y han permitido explicar algunas características esenciales en la percepción de los datos visuales por los seres vivos. Algunos de estos modelos han influido, directa o indirectamente, en la realización de modelos de reconocimiento en máquinas y han marcado directrices de investigación en este campo.

Recíprocamente, en lo que se refiere a las máquinas, partiendo de los aspectos básicos del problema y los procedimientos generales para resolverlo, dedicaremos también nuestra atención a los modelos cuyo fundamento se halla en los prototipos biológicos o bien que operan utilizando procesos adaptivos y de aprendizaje.

Con este estudio en paralelo del proceso de datos ópticos en

los seres vivos y en las máquinas, desearía señalar, también, la íntima conexión entre ambas líneas de investigación y sus influencias en ambos sentidos, lo cual marca a estas investigaciones con el sello de la Biónica.

PROCESO DE DATOS VISUALES EN LOS ORGANISMOS VIVOS

Interacción entre elementos retinales

La interacción entre elementos retinales es de gran importancia en el mecanismo de la visión. Es un hecho conocido que las transiciones, los contrastes en intensidad luminosa o en color, tanto en espacio como en tiempo, contienen la información esencial de lo que vemos ; así, el contorno y la forma de una figura representa, muchas veces, la información más significativa sobre la misma. Los mecanismos que acentúan estos contrastes son los campos de influencias excitadoras o inhibitoras en las redes neuronales de la retina. La interacción lateral parece tener, en líneas generales, el mismo fundamento en los seres de organización sencilla que en los vertebrados. Por ello, este fenómeno ha sido estudiado exhaustivamente desde el punto de vista fisiológico en un organismo primario, como es el *Límulus* [6]. En él se ha investigado la inhibición lateral que es un caso particular de la interacción lateral.

El *Límulus* es un crustáceo que procede de la Era Primaria, habiendo evolucionado muy poco desde entonces. Su ojo contiene, aproximadamente, mil elementos denominados onmatidias, que actúan como unidades receptoras. Las fibras nerviosas salen de los onmatidias formando pequeños manojos que dan lugar al nervio óptico. Una capa horizontal de fibras nerviosas conecta entre sí estos manojos por su parte inferior (fig. 1). Cada onmatidia está constituido a su vez por diez o doce células retinulares en forma de cuña, distribuidas radialmente. Existe además, según puede apreciarse en la figura 2, la célula excéntrica. Tanto las células retinulares como la célula excéntrica poseen axones que, juntos, constituyen el pequeño manojito de fibras que salen de cada

onmatidia. Estas fibras se ramifican horizontalmente constituyendo el plexo bidimensional de interconexiones entre los axones de las células retinulares y excéntricas. En este plexo de interconexiones se verifica la inhibición lateral. Como han demostrado Hart-

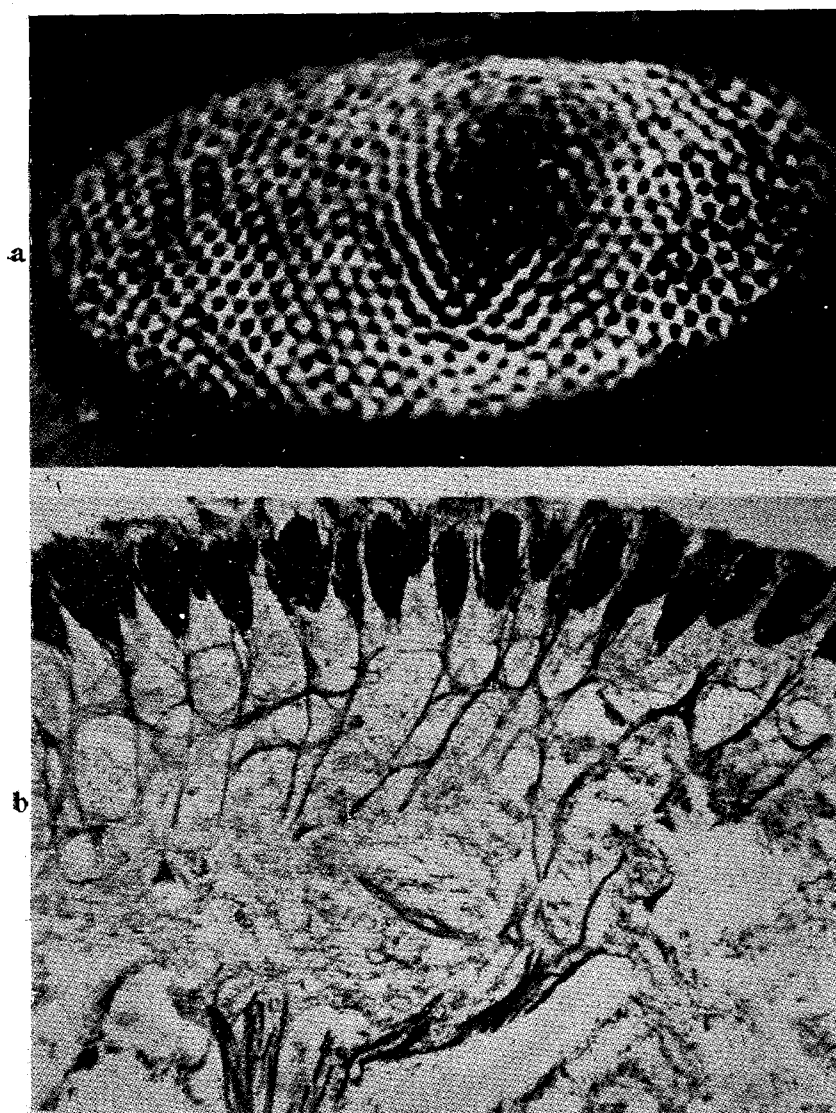


Fig. 1.

line y otros [7], si se cortan las ramificaciones laterales desaparece la inhibición. Debido a ésta, si se ilumina solamente un onmatidia.

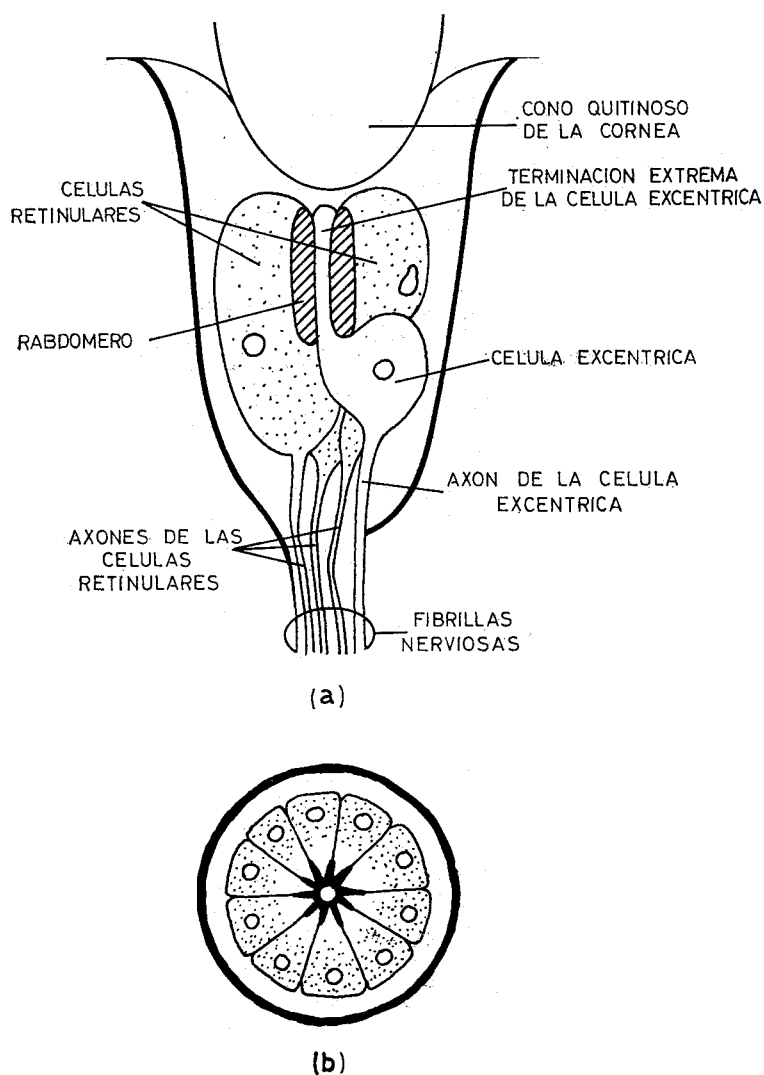


Fig. 2.

su respuesta es distinta de la que se obtendría si al propio tiempo se iluminaran otros próximos. La frecuencia de disparo de un onmatidia disminuye cuando se iluminan otros cercanos (fig. 3).

Las características más sobresalientes de la inhibición lateral en el ojo del Límulus [8] son:

1.º) La magnitud de la inhibición ejercida sobre cada elemento depende, en primera aproximación, linealmente de la respuesta del elemento que la ejerza, una vez superado cierto umbral.

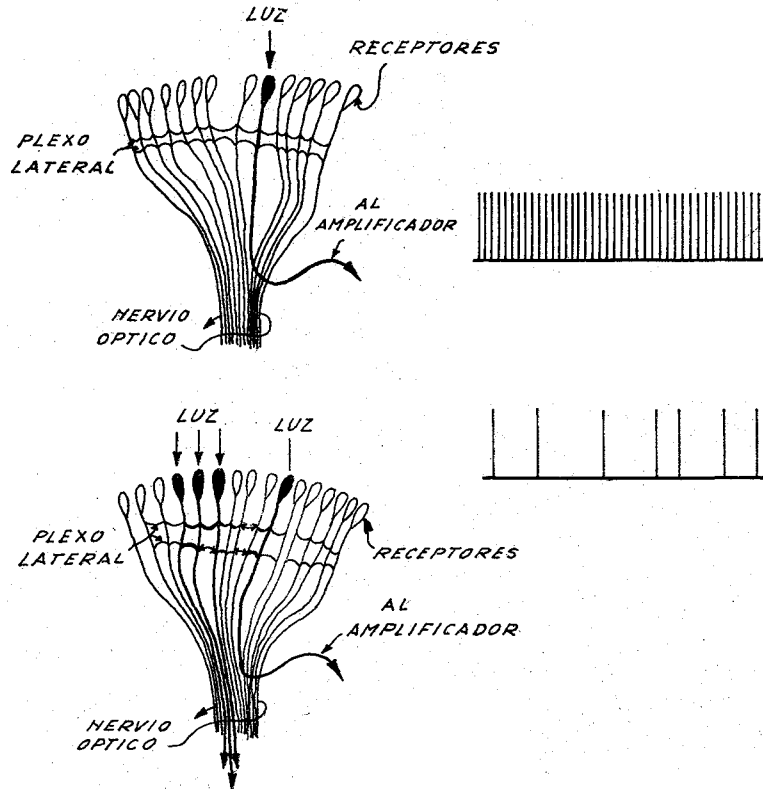


Fig. 3.

2.º) La acción inhibidora sólo se presenta sobre un elemento sometido a excitación.

3.º) La acción inhibidora de un elemento sobre otro disminuye al aumentar la distancia entre ellos de tal forma que se puede suponer una ley inversa de las distancias.

4.º) La acción inhibidora depende linealmente del número de onmatidias iluminados.

5.º) La respuesta de un onmatidia sólo tiene sentido para valores positivos de la «excitación neta», es decir, únicamente existe respuesta cuando la excitación es mayor que la inhibición.

6.º) En estado estacionario, la autoinhibición equivale a una disminución de la excitación.

Partiendo de las características de la inhibición lateral del ojo del *Límulus* [9], suponiendo la superficie del mismo reticulada y que en cada celda exista un onmatidia definido por un par de números (i, j) , la respuesta del onmatidia situado en la celda i, j viene dada por

$$\begin{aligned} y_{ij} &= k_{ij} x_{ij} - k_{ij} \sum_m \sum_n k_{mn,ij} y_{mn} & \text{para } y_{ij} > 0 \\ y_{ij} &= 0 & \text{para } y_{ij} \leq 0 \end{aligned} \quad (1)$$

teniendo presente que x_{ij} es la respuesta en frecuencia de disparo del onmatidia situado en la celda i, j sin considerar ningún tipo de inhibición sobre él. La relación entre la respuesta y la intensidad luminosa que incide sobre el elemento viene dada por

$$x_{ij} = K \log I_{ij}/I_{ij}^0$$

donde I_{ij}^0 representa la intensidad umbral; y_{ij} representa la respuesta en frecuencia de disparo del onmatidia teniendo en cuenta el efecto inhibitorio; $k_{mn,ij}$ representa la acción del onmatidia situado en la posición m, n sobre el de la i, j ; k_{ij} representa el factor de autoinhibición del elemento i, j ($k_{ij} < 1$). Las sumas de la expresión (1) se extienden para todo valor de m, n distinto del i, j .

Hay que tener presente que en esta expresión se parte de la hipótesis de que la inhibición es *recurrente*, es decir, que depende de la respuesta del elemento que ejerce la inhibición. En la figura 4-a se indica el esquema de interacción para dos elementos. En el caso simplificado de un modelo unidimensional la expresión (1) quedaría:

$$y_i = k_i x_i - k_i \sum_j a_{ij} y_j$$

en la que el significado de cada parámetro se deduce inmediatamente de (1).

Algunos autores [10] interpretan los resultados obtenidos suponiendo que la inhibición en el Límulus es *no recurrente*, es de-

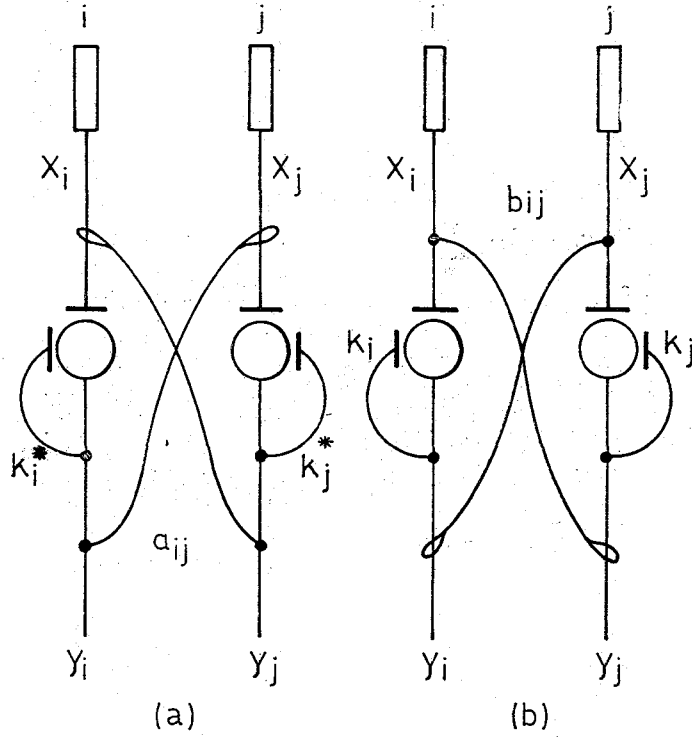


Fig. 4.

cir, que depende de la excitación del elemento que la ejerce (figura 4-b). En este caso, la expresión matemática queda:

$$y_i = k_i x_i - \sum b_{ij} x_j.$$

Bajo ciertas condiciones en el modelo estacionario, ambos tipos de inhibición son equivalentes, es decir, que dada una inhibición *recurrente* es posible, en algunos casos, encontrar una *no recurrente* que sea idéntica en sus efectos a la primera [9].

Juntamente con mi colaborador F. Rubio hemos propuesto un modelo del ojo del Límulus basado en la interacción lateral entre los elementos retinales y que da lugar a un sistema de ecuaciones lineales que permiten definir su funcionamiento generalizando los resultados a una retina continua, es decir, teniendo un gran número de elementos [9].

La generalización a retinas más complejas, como puede ser la de los vertebrados, se puede formular mediante ecuaciones integrales en algunos casos, mientras que en otros basta con simples integrales.

Sea $I(\alpha, \beta)$ la señal de entrada sobre un elemento situado en el punto de coordenadas (α, β) respecto de un sistema adecuadamente elegido en la capa transversal donde se encuentra dicho elemento genérico; sea $\Phi(x, y)$ la señal de salida de un elemento en posición (x, y) suponiendo que la interacción es recurrente, la expresión más general para expresar ésta, bajo ciertas hipótesis, es:

$$\Phi(x, y) = I(x, y) + \int \int_{\mathbb{R}} K(x, y; \alpha, \beta) \Phi(\alpha, \beta) d\alpha d\beta \quad (2)$$

$K(x, y; \alpha, \beta)$ es el núcleo de la transformación integral (factor de peso en la formulación discreta); su forma analítica viene determinada por la morfología de la capa de conexiones transversales que se desee modelar: pies de fotorreceptores, amacrinas de la capa plexiforme externa, células horizontales, ... [11]. La interacción no recurrente se formula de modo análogo a la anterior sin más que sustituir la función $\Phi(\alpha, \beta)$ por $I(\alpha, \beta)$

Como veremos más adelante, una formulación de este tipo es según creemos de las más generales que se pueden plantear, puesto que permiten englobar no sólo los fenómenos de interacción lateral sino también los de detección de fenómenos específicos, tales como detección de contrastes, movimientos, velocidades..., sobre todo si el modelo se hace no lineal (núcleos no lineales). Ahora bien, para utilizar dichos modelos se hace imprescindible simular-

los en calculadora, puesto que el tratamiento analítico resulta imposible por su complejidad y laboriosidad.

Con posterioridad se ha simulado electrónicamente el modelo discreto [12]. La unidad funcional de inhibición lateral está constituida por un receptor y por un inhibidor. El primero es un astable en el que la tensión de base del primer transistor está controlada por un fotodiodo, consiguiéndose de esta forma un tren de impulsos cuya frecuencia es proporcional al logaritmo de la intensidad luminosa incidente sobre el fotodiodo (fig. 5). La salida del astable se lleva al inhibidor, cuya estructura es análoga al receptor pero funcionando como monoestable. En este monoestable se puede incluir el fenómeno de inhibición que presenta el Límulus. La inhibición entre unidades elementales se realiza mediante el acoplo resistivo entre los elementos inhibidores de la misma. Este acoplo se encuentra entre los emisores del transistor T (fig. 6), siendo regulable, con lo que se puede dar un cierto peso a la acción inhibidora.

También se ha construido un modelo lineal constituido por diez unidades elementales, habiéndose encontrado una gran analogía entre los resultados experimentales obtenidos y los hallados por Hartline y Ratliff, partiendo de medidas directas en el ojo del Límulus (fig. 7).

Como ya se ha indicado, la inhibición lateral acentúa los contrastes de la imagen retiniana. Los bordes y contornos aparecen vigorosamente definidos en la representación neuronal. Esto no es una copia directa de la figura que aparece en el mosaico de receptores. Se acentúa cierta información importante para el organismo a expensas de otra información menos significativa.

Lo esencial es la percepción de las variaciones que aparecen en el mundo exterior. Si éste aparece invariable, como por ejemplo una superficie igualmente iluminada, al no contener ninguna información el organismo no observa nada.

La inhibición lateral es el fundamento de fenómenos visuales, como el contraste en color y brillo. Es curioso observar que estos fenómenos de inhibición lateral fueron previstos hace más de un siglo por Hach [13] basándose exclusivamente en observaciones psi-

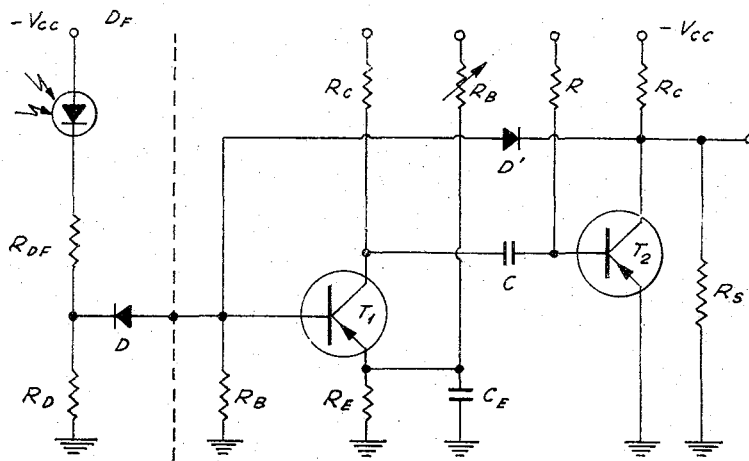


Fig. 5.

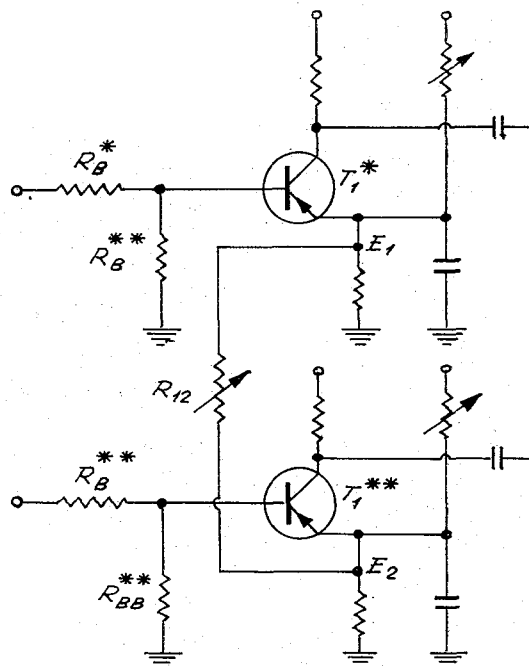
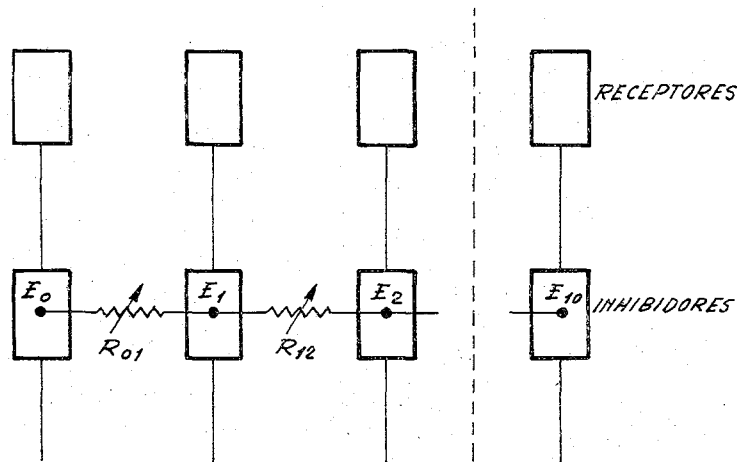


Fig. 6.

cofísicas. Reichardt y Mc Ginitie [14] han realizado un modelo interesante de reconocimiento de formas, basado en el modelo de Hartline y Ratliff de inhibición lateral (fig. 8). Cada receptor está conectado con los demás de forma tal que sus influencias inhibitoras disminuyen con la distancia. El modelo se simula con una calculadora analógica. Y_v representa la amplitud del potencial ge-



Red lineal de elementos sometidos a inhibición lateral. Los puntos E_i , corresponden a los puntos de la misma nomenclatura, de los cuales se han representado dos de ellos en la figura-6

$i = 0, 1, 2, \dots, 10$

Fig. 7

nerador en cada unidad y Z_0 la frecuencia de descarga en el nervio óptico. Las líneas diagonales representan el plexo de interconexiones a través de las cuales cada unidad ejerce influencias sobre los elementos vecinos. Los triángulos representan amplificadores operacionales cuyas salidas se conectan (realimentan) a las entradas que inhibirían sus salidas. Cada amplificador lleva conectado un diodo para prevenir cambios de polaridad en la salida, ya que las salidas equivalentes a frecuencias negativas no son permitidas. Al final todas las salidas se suman. No se pretende que esta suma

represente el proceso fisiológico pero permite la discriminación de cierto tipo de formas. Por ejemplo, si la figura proyectada sobre el mosaico de fotorreceptores es un rectángulo, cuanto más alargado sea, la acción inhibitoria de los diferentes elementos es menor

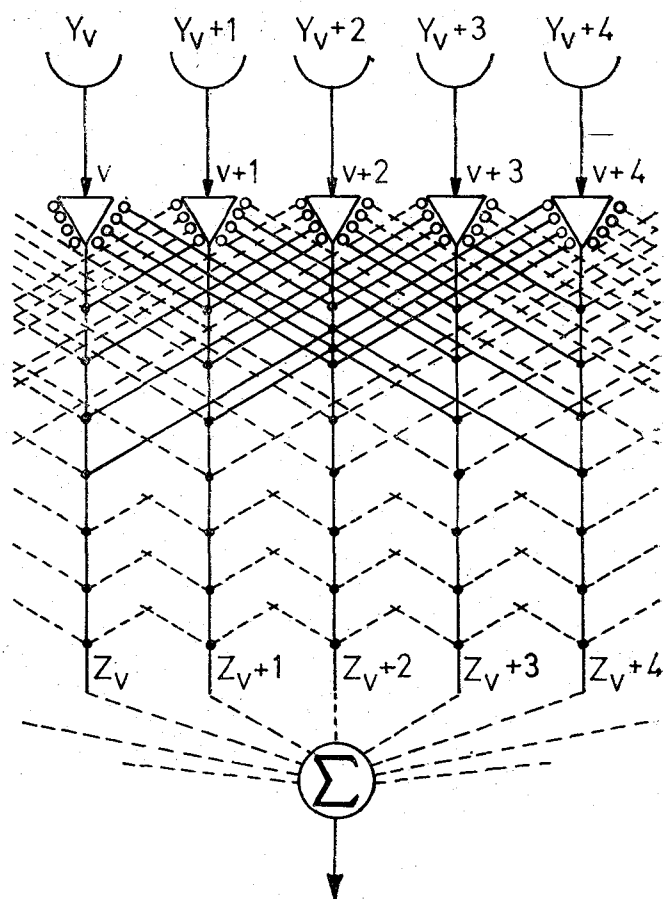


Fig. 8.

y, por consiguiente, mayor es la respuesta ; mientras que si la anchura del mismo se hace mayor hasta convertirse en un cuadrado, el efecto inhibitorio aumenta y la respuesta disminuye.

También hay que citar una máquina de reconocimiento de formas, basada en la interacción lateral, construida en la General Electric Co. por H. Hildebrand.

Detección de figuras en movimiento. Sistema visual de la rana

El sistema visual de la rana es más complejo que el del *Límulus*, pero mucho menos que el del hombre. El estudio del ojo de la rana ofrece un gran interés, bajo el punto de vista de la estructura cerebral y bajo el punto de vista del sistema de reconocimiento de formas, particularmente en lo que se refiere a la detección de figuras en movimiento.

En efecto, el movimiento de pequeños seres, como insectos, son detectados por el ojo de la rana y esta característica de su retina le sirve para alimentarse. Hay que señalar que no reconoce a los insectos como tales, sino que cualquier objeto, con las dimensiones y con el tipo de movimiento de aquéllos, le excita igualmente y salta para capturarlo.

Si el cuerpo que se mueve es grande también es detectado, pero entonces la reacción de la rana es de huir. Es decir, dos de las funciones importantes de este animal —buscar el alimento para su sustento y evitar los peligros del mundo exterior— las realiza mediante características peculiares de su ojo, particularmente adaptado a observar el movimiento de los cuerpos. Si éstos no se mueven no dan lugar a ninguna reacción por parte de la rana que permanece indiferente al mundo estacionario que le rodea.

Los ojos de la rana no se mueven, como ocurre con los humanos, para observar el mundo exterior, sino que se mantienen fijos en una determinada dirección. Si mueve la cabeza, los ojos se mueven en dirección contraria con objeto de compensar aquel movimiento y mantener su mirada en la dirección anterior. Es decir, podemos considerar los ojos de la rana estabilizados. Esta falta de movimiento es una de las ventajas que ofrece su estudio. Otra es la simplicidad de las conexiones entre los ojos y el cerebro. No tiene área de sensibilidad (la fovea), como ocurre en el ojo humano. Su retina es uniforme.

¿Cuáles son las propiedades, las características del sistema visual de la rana que permiten explicar sus reacciones ante la posible caza de los seres que le sirven de alimento y su defensa ante un posible peligro?

Lettwin, Maturana, McCulloch y Pitts —del Instituto de Tecnología de Massachussets (USA)— emprendieron una investigación con este objeto [15] cuyos resultados vamos a exponer en líneas generales:

La estructura de la retina de la rana es conocida desde hace mucho tiempo. En la figura 9 se indica un esquema de la misma debido a Cajal. Aunque existen gran número de neuronas, se puede simplificar su descripción considerando que las células sensibles a la luz (conos y bastones) están conectadas a una capa intermedia de neuronas constituida por las células bipolares, las amacrinas y las horizontales. Los axones de las neuronas de esta capa están conectados a su vez con las células ganglionares cuyos axones constituyen el nervio óptico que sale del ojo y se dirige al cerebro.

El número de receptores en la retina de la rana es del orden de un millón; el de neuronas en la capa intermedia de dos y medio a tres y medio millones y el de células multipolares o ganglionares de medio millón.

El tratamiento de la información se realiza, pues, a dos niveles, teniendo presente que cada receptor está conectado a través de la capa intermedia con gran número de ganglionares y cada ganglionar a gran número de receptores. Esta complejidad en las interconexiones de los diferentes tipos de neuronas y los números distintos de éstas que existen en cada capa sugieren que la función de la retina no es suministrar una imagen exacta del mundo exterior del cerebro sino más bien analizar la información recibida y transmitir lo esencial de la misma a aquél.

El hecho de que cada ganglionar esté conectada a gran número de receptores —que pueden ser centenares o miles— explica que cada una de aquéllas responda a las excitaciones, a las variaciones de brillo, por ejemplo, sobre una pequeña área de la retina (que puede ser de varios grados). Se denomina *campo receptivo* de la ganglionar a esta zona de la retina que comprende todos los receptores cuya actividad afecta a aquélla. Estas observaciones, así como la definición de campo receptivo, son debidas a Hartline, que investigó estas materias y llegó a la conclusión de que exis-

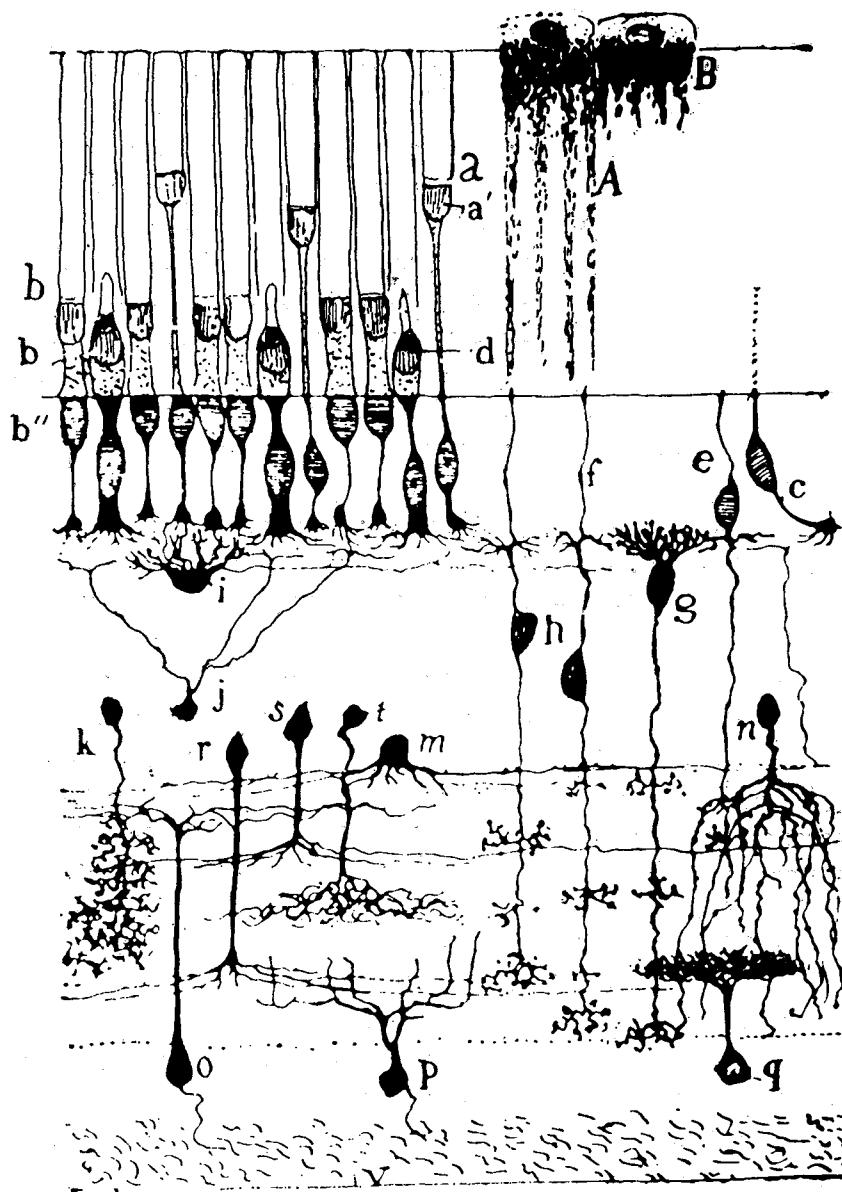


Fig. 9.—Los elementos de la parte superior izquierda de la fotografía son los conos y bastones; *g*, *h*, células bipolares; *i*, célula horizontal; *k*, *r*, *s*, *t*, *m*, células amacrinas; *o*, *p*, *q*, células ganglionares.

tían tres clases distintas de campos receptivos en la rana, que corresponden a los siguientes tipos de ganglionares:

1.º Células ON (denominadas así por Hartline). Si una pequeña señal luminosa aparece bruscamente en el campo receptivo de esta célula, su descarga empieza pronto, aumenta en frecuencia hasta un límite que depende del área e intensidad de la señal y después disminuye lentamente. Si se interrumpe la señal luminosa, la descarga desaparece.

2.º Ganglionares ON-OFF. Si la señal luminosa aparece y desaparece bruscamente en el campo receptivo, se produce una corta descarga, de frecuencia elevada, tanto en un caso como en otro (al pasar de señal luminosa —ON— a ausencia de señal —OFF— y recíprocamente).

3.º Ganglionares OFF. Cuando desaparece una señal luminosa del campo receptivo de estas células, se produce una descarga prolongada e inmediata. Esta descarga desaparece inmediatamente si se presenta nuevamente la señal luminosa.

Este trabajo experimental de Hartline, ampliado más adelante por Barlow, ha dado lugar a un conjunto de conclusiones muy interesantes por lo que hace referencia al comportamiento del ojo de la rana a estímulos producidos por pequeñas señales luminosas. Establecieron que los axones de las ganglionares (que constituyen el nervio óptico) no transmiten información solamente sobre la intensidad luminosa de simples puntos de la retina, sino más bien sobre la distribución de la iluminación en el área del campo receptivo.

Ahora bien, Lettwin y sus colaboradores adoptaron otro tipo de investigación. Partiendo de la base de que el sistema del ojo, su estructura, tiene por misión detectar ciertas configuraciones luminosas y las modificaciones que sufren, el estudio —en general— de las acciones producidas por pequeñas señales luminosas, si bien permite llegar a leyes de acuerdo con la realidad, éstas no son adecuadas para llegar al fondo del problema de cómo el ojo de la rana recibe sus *universales*, es decir, cómo reconoce a sus posibles víctimas y a sus posibles enemigos.

Por ello, las investigaciones de Lettwin y colaboradores se di-

rigieron a presentar a la rana una amplia gama de estímulos visibles como, por ejemplo, objetos que ella estuviera dispuesta a comer, objetos que pudieran volar análogamente a los insectos, figuras geométricas estacionarias y móviles, es decir, adoptaron una línea de investigación más realista que les permitiera llegar a obtener ciertas invariantes funcionales.

No haremos una detallada descripción de los dispositivos experimentales de estos investigadores, que pueden leerse en sus comunicaciones originales [15], [16]. Los estímulos consistían en siluetas de diferentes formas y dimensiones y de colores blanco, gris o negro, montadas sobre piezas de hierro. Estos objetos se movían mediante imanes. Las respuestas de las fibras nerviosas se medían mediante un electrodo inserto convenientemente detrás del ojo de una rana y que suministraba datos sobre la frecuencia y duración del disparo. Hay que señalar que estas experiencias tenían por objeto más que obtener valores cuantitativos «lograr resultados relativos, es decir, qué configuración o escena visual produce mayor respuesta y cuál la produce menor y qué aspecto de la imagen puede variarse sin que cambie la respuesta».

Los resultados de estas experiencias permitieron establecer los siguientes grupos de células ganglionares:

Grupo I. Detector de contrastes.

Estas ganglionares tienen campos receptores de 2 a 4 grados de diámetro. Cualquier límite o borde de separación entre grises es captado por estas células siempre que sea abrupto. Más que el contraste entre grises lo que se mide es la nitidez, lo pronunciado del límite de separación. La respuesta se incrementa si este límite se mueve y permanece invariable si se altera la iluminación aunque sea ampliamente.

Si no existe borde, aunque cambie de forma brusca la iluminación, no hay respuesta.

Por otra parte, si una línea se lleva al campo receptivo en ple-

na oscuridad y se ilumina de pronto, se obtiene una respuesta continua después de un corto tiempo. Sin duda alguna, estas ganglionares coinciden con las observadas por Hartline, denominadas células ON.

Grupo II. Detector de límites convexos.

Estas ganglionares tienen campos receptivos de 3 a 5 grados. Responden solamente a límites abruptos entre dos grises pero sólo si estos límites son curvos, siendo convexa la zona más oscura. Además estos límites deben moverse o se han movido. La célula, pues, no responde a límites constituidos por segmentos rectilíneos, tanto si se mueven como si están inmóviles. Cuando la curvatura es tan grande que el objeto es simplemente una pequeña área —con diámetro aproximadamente igual a la mitad del campo receptivo— si aquél se mueve centrípetamente y se para antes de llegar al centro de dicho campo, se produce una descarga de larga duración.

Las respuestas son invariantes a un amplio margen de iluminación. No pertenecen a ninguno de los tipos de ganglionares observados por Hartline.

Grupo III. Detectores de movimiento o de variaciones en contraste.

Estas ganglionares tienen campos receptivos de 7 a 11 grados. Responden de forma invariante al movimiento de la misma figura sobre el mismo fondo, si su velocidad es constante. Esta invariante se mantiene aunque existan cambios en la iluminación. La respuesta es tanto mejor cuanto mayor sea la velocidad o el límite sea más abrupto. No existe respuesta duradera sino simplemente disparo en el caso de que el contraste cambie o se mueva. Estas ganglionares se identifican con las OFF-ON de Hartline.

Grupo IV. Detectores de oscurecimiento.

Responden a cualquier oscurecimiento en el campo receptivo, dependiendo la respuesta de la distancia al centro del campo. Estas ganglionares no responden a límites o bordes. La respuesta depende del grado relativo de oscurecimiento, siendo prácticamente independiente del grado de iluminación inicial. Si se interrumpe la luz, se produce una respuesta duradera. Estas ganglionares coinciden con las OFF de Hartline.

Grupo V. Pequeño grupo de ganglionares que no parece tengan campo receptivo.

Miden la mayor o menor oscuridad en grandes áreas, de forma que la frecuencia de disparo es tanto mayor cuanto menor es la iluminación. Cuando ésta cambia, la frecuencia se modifica lentamente hasta alcanzar el nivel.

Las ganglionares del ojo de la rana pertenecen a uno u otro de los grupos indicados. Los axones de estas ganglionares que terminan en el *tectum* (fig. 10) están distribuidos de forma ordenada, de manera que representan un mapa continuo de la retina. Por pequeña que sea la zona, existen ganglionares de cada una de aquellas clases.

Estos axones de las ganglionares terminan en diferentes capas del *tectum*, correspondiendo cada una a un grupo, excepto la correspondiente al grupo III, en la cual terminan también los del último (V). Cada capa representa un mapa de la retina, por lo que hace referencia a las funciones asignadas a las correspondientes ganglionares. La capa más superficial corresponde a los terminales de las células detectoras de contrastes. Debajo, pero no muy separados, están los detectores de convexidad. Más profunda está la capa de los detectores de movimiento de bordes y el grupo V. En fin, la capa más profunda corresponde a los detectores de oscurecimiento.

Las conclusiones a que se llegan con estas investigaciones son

de que el ojo transmite al cerebro una información en un lenguaje organizado y que no es simplemente una copia más o menos exacta de la distribución de la luz en los receptores. La imagen de la

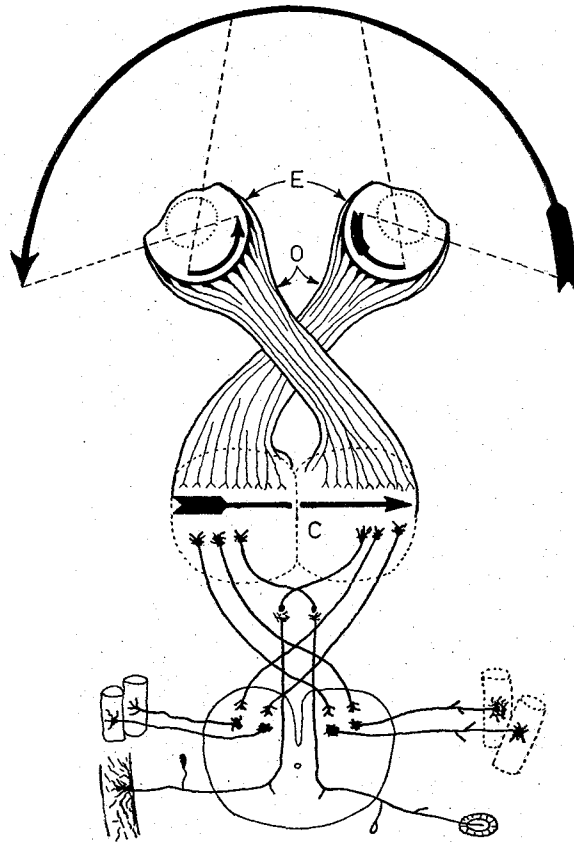


Fig. 10.

retina se analiza en cada zona en función de las operaciones que pueden realizar los cuatro grupos de ganglionares.

1.° Los detectores de *contrastes* locales dan información sobre la presencia de límites o bordes abruptos con más o menos contraste.

2.° Los detectores de *convexidad* informan sobre áreas más extensas, si el objeto tiene límites curvos, si es más oscuro que el

medio o se mueve en él. Observan si el objeto se mueve intermitentemente con respecto al medio o fondo de la imagen. Este grupo de ganglionares se denomina también *detector de insectos* porque las características que acabamos de indicar son las que definen su aptitud para localizar la presa.

3.° Los detectores de *movimiento de bordes* dan información sobre el movimiento de objetos mayores.

4.° Los *detectores de oscurecimiento* se refieren a grandes áreas, como por ejemplo la sombra producida por objetos o seres que puedan ser posibles enemigos y de los que hay que huir.

Así pues, la imagen sobre la retina se transforma en un conjunto de datos, en una información sobre cualidades destacadas de aquélla, a través de las diferentes neuronas y de las ganglionares que llegan a las cuatro capas del *tectum*.

Los investigadores citados han estudiado la anatomía de los diferentes grupos de ganglionares con objeto de lograr una correlación entre su estructura y las operaciones que realizan. Estas investigaciones se han seguido por los mismos autores y por otros tratando de explicar más concretamente las operaciones de las diferentes ganglionares. El grupo que resulta más complejo para proponer un modelo adecuado es el denominado «detectores de insectos» y por otro lado, el sistema de reconocimiento de configuraciones más especializado de la retina de la rana. Se han publicado varios trabajos sobre este tema de los cuales el último y, probablemente, hasta el momento, el mejor es debido a Moreno-Díaz [17].

En el laboratorio de Instrumentación del Instituto Tecnológico de Massachusetts, Luis Sutro y colaboradores han desarrollado un modelo electrónico que representa las células y conexiones neuronales de los distintos grupos de células ganglionares de la rana [18]. En la figura 11 se muestra esquemáticamente el sistema visual de la rana, con sus fotorreceptores, células bipolares, ganglionares, así como los grupos de terminales de estas últimas. La electrónica que emplean para simular el gran número de elementos existentes en la retina de la rana consiste en cámaras de televisión (vidicones) y calculadoras de propósitos especiales tra-

bajando en tiempo compartido. Han realizado dos diseños, uno empleando técnicas tanto analógicas como digitales (sistema alfa) ; y otro en el que se emplean únicamente técnicas analógicas (sistema beta). En las figuras 12 y 13 se muestran los diagramas de bloques de ambos sistemas.

Este modelo electrónico ha servido para cumplir dos objetivos: por un lado, para ayudar a la creación de un modelo conceptual

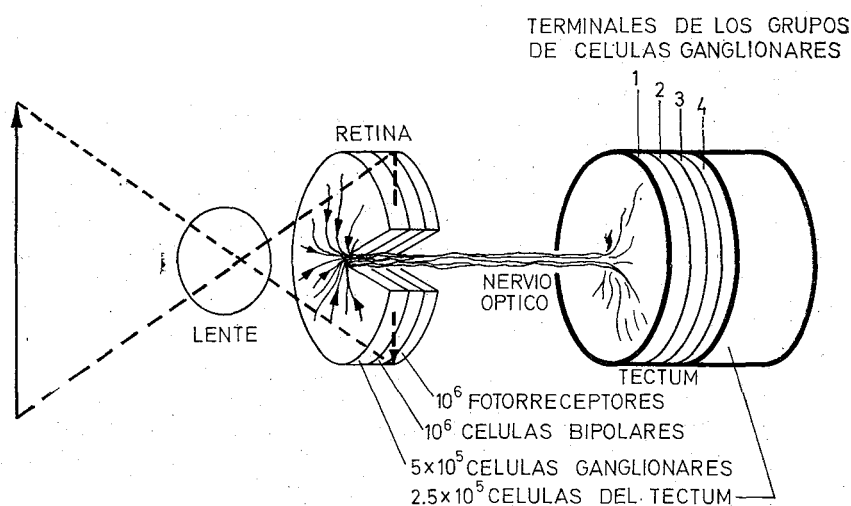


Fig. 11.

de la retina ; por otro lado, para ensayar sensores útiles para los viajes espaciales. Uno de estos sensores se ha proyectado para el primer viaje a Marte. No parece que actualmente puedan enviarse fotografías completas a la Tierra desde el vehículo espacial situado en aquel planeta. Por ello se considera que sería útil almacenar la información obtenida por un sistema que actúe como el ojo de la rana.

Trabajos análogos a los de Lettwin y colegas se han realizado con otros animales ; así es digno de citarse el llevado a cabo con el sistema visual del gato en su cuerpo lateral geniculado cuyas células se comportan de forma similar a las del *tectum* de la rana [19].

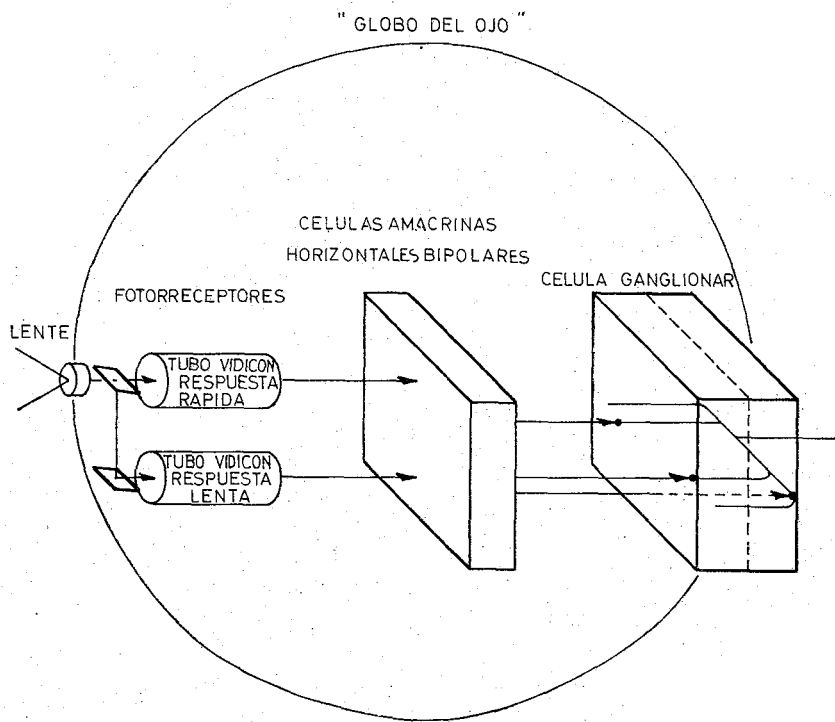


Fig. 12.

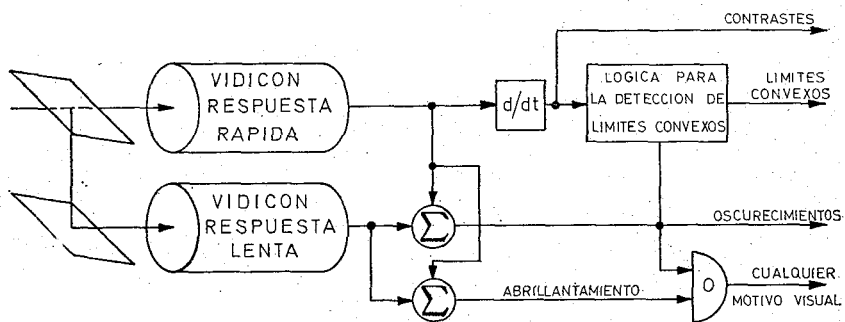


Fig. 13.

Se ha experimentado con otros animales y en todos ellos se observan resultados similares en lo referente a la detección del movimiento, su dirección, velocidad, etc.

Transmisión de la información espacial

Un camino de investigación interesante en el proceso de datos ópticos en los seres vivos es el que se refiere a la transmisión de la información espacial, puesta en evidencia por Campbell y otros investigadores. Ciertas propiedades puestas de manifiesto en algunos mamíferos (gato, mono, hombre) parecen indicar que en el camino visual de estos seres se lleva a cabo un análisis armónico espacial análogamente al realizado por el oído en el dominio temporal.

Helmholtz que, según es sabido, contribuyó de forma decisiva al conocimiento de la naturaleza física de la música y de la fisiología del oído trató también de aplicar el análisis armónico al ojo. Sin embargo, al comparar ambos órganos sensoriales, consideraba la visión de los colores como propiedad análoga o equivalente a la que tiene el oído al percibir un sonido compuesto. Pero si bien éste lo analiza el oído captando sus componentes, no ocurre así con la visión de los colores ya que el ojo no tiene aptitud para resolver un color compuesto en diferentes longitudes de onda. Por ello Helmholtz pensaba «que en el ojo no existía armonía en el mismo sentido que en el oído y que no existía música». Según Campbell *sí existe* música pero no en el dominio del color sino del espacio» [20].

F. W. Campbell considera que las propiedades de análisis y transmisión de datos ópticos del sistema visual humano y de los animales superiores pueden estudiarse considerando que el mismo está compuesto de dos partes: el sistema óptico del ojo y el sistema retina-cerebro.

Campbell y Green [21] hacen un estudio cuantitativo del papel relativo del sistema óptico y del sistema retina-cerebro en la detección de rendijas (fundamental para poner de manifiesto estos

fenómenos a diversas frecuencias espaciales), llegando a la conclusión de que el límite de visibilidad —mínimo contraste necesario para percibir una rendija sinusoidal— viene impuesto por el sistema nervioso más que por el óptico. Por lo cual estamos interesados fundamentalmente en el sistema retina-cerebro.

Las experiencias realizadas en los últimos años por Enroth-Cugell y Robson [22], Cooper y Robson [23], Campbell, Clelland, Cooper y Enroth-Cugell [24], Campbell y Kulikowski [25] nos permiten resumir para el camino visual las propiedades siguientes:

- a) En la retina de algunos mamíferos (gato, mono y hombre) existen caminos sintonizados a distintas frecuencias espaciales.
- b) Esta selectividad es mayor en el hombre que en el mono y en éste mayor que en el gato.
- c) La anchura de banda de estos canales es aproximadamente menor de un octavo de la frecuencia espacial a la cual resuenan.
- d) Esta selectividad se detecta a nivel de las ganglionares o centros superiores del camino visual.
- e) Existe una respuesta selectiva, a nivel de corteza, para una determinada orientación de una rendija.

Resumiendo se puede decir que a nivel ganglionar o a nivel de los centros superiores —dependiendo de la complejidad estructural del ser vivo— se han codificado dos propiedades importantes de una imagen: 1) la información acerca de la orientación de un borde; 2) la extracción del contenido en frecuencias espaciales de la imagen que se encuentra sobre la retina.

Partiendo de los resultados obtenidos por estos investigadores, se han propuesto modelos que explican el análisis armónico espacial en el camino visual. García Santesmases y colaboradores [26] suponen que el campo receptivo está constituido por capas concéntricas que actúan como excitadoras o inhibidoras y cuya acción decrece con la distancia. Partiendo de esta hipótesis y aplicando el proceso de convolución se llega a un resultado sobre la selectividad de frecuencias en las células ganglionares. La frecuencia propia de la ganglionar está relacionada estrechamente con las dimensiones del campo receptivo.

En esta última línea de investigación se pueden considerar los trabajos de Swigert y Vogelhut, que consideran que en el camino visual existe un filtraje espacial. Las tres funciones fundamentales de los procesos de filtraje: dispersión espacial de la imagen, modulación de la imagen por el filtro y reconstrucción de la imagen transformada, pueden ser realizadas en el camino visual puesto que éste presenta las estructuras adecuadas para ello. Este procedimiento de filtraje permite a la red neuronal poseer la propiedad de identificación de escenas. Este procedimiento ya ha sido realizado por procesadores ópticos aunque sin el grado de flexibilidad que incorporaría el llevar a cabo el modelo neuronal [27].

El modelo de proceso de información visual, anteriormente expuesto, basado en admitir la existencia de campos receptivos concéntricos excitadores e inhibidores, no sólo da cuenta de los fenómenos descritos por Campbell y colaboradores, sino que incluye, como casos particulares, los de filtraje espacial e incluso los de detección de contrastes en movimiento y estacionarios. En un trabajo [28] presentado por mis colaboradores al Congreso de Automática celebrado en Madrid hace unos meses, se muestra que las propiedades de las células de los distintos grupos de la rana pueden explicarse por un modelo de este tipo. Este modelo ha sido simulado en una calculadora digital. En la actualidad se está trabajando en la extensión del mismo para explicar las propiedades detectadas en las retinas de otros animales. Creemos que modelos de este tipo son los más adecuados para explicar el proceso de datos visuales, debido a la estructura por capas que presenta la retina.

Entre los modelos puramente matemáticos sobre el proceso de datos visuales, cabe destacar los de K. N. Leibovic, el cual considera que los campos receptivos se encuentran distribuidos al azar sobre la retina y a partir de los métodos de la Geometría Integral extrae características: longitudes, áreas, curvaturas, detección de contrastes, etc. de los objetos utilizados como estímulos [29].

En cuanto a modelos puramente electrónicos cabe destacar el realizado por R. G. Runge, M. Uemura y S. S. Viglione, los

cuales incorporan en un dispositivo electrónico —que sigue con bastante fidelidad a la anatomía— las principales propiedades de la retina de la paloma: movimiento direccional de un punto o un eje oscuro sobre un campo iluminado y recíprocamente [30].

RECONOCIMIENTO DE FORMAS VISUALES EN LAS MÁQUINAS

Generalidades

Gran número de modelos de reconocimiento de formas se han realizado en estos últimos años. Muchos de ellos se han construido teniendo por objeto primordial el reconocimiento de símbolos en textos escritos. Ello ha llevado, sin duda, a una limitación del problema, ya que gran parte del esfuerzo se ha concentrado en el reconocimiento de las letras del alfabeto, sin tratar de relacionar estas investigaciones con las de un modelo *humano* de reconocimiento. Existe un divorcio entre los trabajos de los psicólogos que estudian los problemas de reconocimiento y los de los físicos, ingenieros, matemáticos, que tratan de problemas similares, si bien ambos grupos de científicos tratan de lograr modelos de la misma función. Por ello, antes de seguir adelante, vamos a situar el problema de reconocimiento de escenas lo más claramente posible. Primeramente ¿cómo podemos definir este proceso de reconocimiento? Varias son las definiciones dadas; una de ellas, muy sencilla, se puede enunciar así: *Es el proceso de clasificar la información sensorial en categorías mutuamente exclusivas.*

Una definición más amplia sería la siguiente: es un proceso mediante el cual se extrae suficiente información sensorial para contestar a una pregunta relativa a la escena. Para ello es necesario, según Minsky, una lista de los objetos que existan, la escena y una exposición de las relaciones entre ellos.

En fin, Selfridge señala que el reconocimiento de escenas debe basarse en extraer las características *esenciales*, *significativas*, partiendo de un conjunto de detalles sin importancia.

De la primera definición se deduce que el proceso de reconocimiento de escenas es un problema de clasificación. Para realizar

este proceso de clasificación debemos representar las escenas numéricamente, es decir, cada escena por un conjunto de números reales: $x_1, x_2, x_3, \dots, x_n$. Este conjunto puede representarse a su vez por un punto X en un espacio de n dimensiones, cuyas coordenadas son precisamente los números $x_1, x_2, \dots, x_n$, o bien, por un vector $\vec{X}$, cuyas componentes son estos números. Este proceso previo a la clasificación propiamente dicha, es decir, el proceso de convertir la escena en un punto X es lo que denominamos *preproceso o pre-elaboración*.

Según veremos existen diferentes métodos para lograr este preproceso. Una vez obtenido X se puede proceder a clasificar este punto y, por consiguiente, la escena. Supongamos que la escena de que se trata puede clasificarse en una cualquiera de P categorías, es decir, que puede pertenecer a cualquiera de éstas. Si ahora suponemos que todas las escenas que pertenecen a una categoría dan lugar al mismo vector $\vec{X}_i$, es evidente que el problema de clasificación sería muy simple, bastaría para clasificar una nueva escena, observar el vector $\vec{X}$ correspondiente y proceder a su identificación con uno de los P vectores representativos de las P categorías.

Sin embargo, el problema no es tan sencillo ya que, en general, las escenas correspondientes a la misma categoría no dan lugar al mismo vector $\vec{X}$. Existe una dispersión de los vectores representativos de las escenas de una determinada categoría. Los métodos de pre-elaboración no son suficientemente perfeccionados para obtener un solo vector para cada categoría. Esto lleva consigo el estudio e investigación de los métodos óptimos de clasificación, teniendo presente esta dispersión de los puntos representativos de cada categoría. Para ello se emplea la noción estadística de densidad de probabilidad con objeto de describir la distribución de los puntos representativos de la escena. En efecto, hay que tener en cuenta que se parte de la hipótesis básica de que los puntos X próximos entre sí tienden a pertenecer a la misma categoría, si bien el recíproco no es necesariamente cierto, es decir,

que no todos los puntos pertenecientes a una misma categoría están próximos entre sí.

Partiendo de la función de densidad probabilística y utilizando el análisis estadístico, se puede llegar a métodos óptimos para decidir a qué categoría pertenece una determinada configuración X .

En el caso de que la distribución de los puntos representativos de las escenas tenga una forma Gaussiana conocida, el proceso de clasificación resulta relativamente fácil. Sin embargo, no siempre es tarea fácil conocer la forma de la función. Debe realizarse un trabajo exhaustivo, observando gran número de escenas pertenecientes a una determinada categoría, con objeto de determinar los vectores correspondientes y, en consecuencia, precisar la naturaleza de la distribución.

Pero, en realidad, raramente ocurre que la distribución sea gaussiana. Normalmente la forma de la curva de la función de densidad probabilística se aparta de la gaussiana, ya que ofrece en algunos casos varios máximos. Por consiguiente, las técnicas de clasificación deben ser tales que puedan utilizarse para funciones complejas de densidad probabilística.

Nos encontramos, pues, con que el problema del reconocimiento de escenas comprende dos fases perfectamente definidas; por un lado el preproceso mediante el cual se representa la escena y por otro la clasificación propiamente dicha que asigna esta escena a una determinada categoría.

La primera fase, el preproceso, dependerá evidentemente de la escena en cuestión. Si este preproceso da lugar a una pequeña dispersión de los vectores $\vec{X}$ o bien da lugar a un único vector, evidentemente, entonces la fase segunda, la de clasificación, resulta fácil. Pero si el preproceso logrado es deficiente, es decir, la distribución de los vectores $\vec{X}$ es muy complicada, por muy elaborados que sean los métodos de clasificación no se llegará a resultados satisfactorios.

Por otro lado, si la escena no es simple, como ocurriría en un paisaje en que existan casas, árboles, ríos, etc., ya se comprende que el proceso de clasificación en categorías mutuamente exclusivas resultaría prácticamente imposible de aplicar, requeriría un

número prohibitivo de categorías distintas. Sin embargo, para casos relativamente sencillos, resulta apropiado el método de clasificación.

Cuando se trata de escenas acústicas, lo que se hace en general es convertir las señales acústicas en señales eléctricas y, partiendo de éstas, obtener el conjunto de números que definen la escena. Cuando se trata de una escena visual, ésta se divide en gran número de cuadrados formando un enrejado de celdillas. La intensidad luminosa de cada una de éstas se transforma en una señal eléctrica.

En estas condiciones, una representación satisfactoria de la escena lleva en general a un conjunto de números prohibitivo, por lo cual se trata de *extraer de la escena las características fundamentales*, lo esencial que puede distinguirlas de las demás y que en el caso de escenas visuales, por ejemplo, reconocimiento de letras, pueden ser segmentos rectilíneos de determinada pendiente, de determinada longitud, finales de líneas, número de intersecciones y su posición y orden, número y posición de las esquinas, determinadas curvas, etc.

En casos muy complejos el problema no es fácil. No existen reglas que permitan establecer métodos generales para la extracción de las características de una determinada escena. Depende del tipo de ésta y hay que adaptarse a la misma buscando las reglas adecuadas para cada caso. Es pues un procedimiento más bien empírico el que debe seguirse. Por ello resulta tan interesante el estudio de los prototipos biológicos a los cuales se dedican muchos investigadores.

Extracción de características

Antes de proceder a la extracción de las características de la escena, se realizan unas operaciones auxiliares que se refieren a eliminar lo que podemos llamar *ruido de la escena*. Si suponemos como caso más sencillo que esta escena es una letra, se deben eliminar los puntos negros o líneas delgadas que no pertenecen a

aquella ; pequeñas señales blancas situadas en las líneas que la forman, irregularidades en los bordes de estas líneas, pequeñas discontinuidades en éstas, etc.

Para eliminar este *ruido* se puede hacer lo siguiente : La figu-

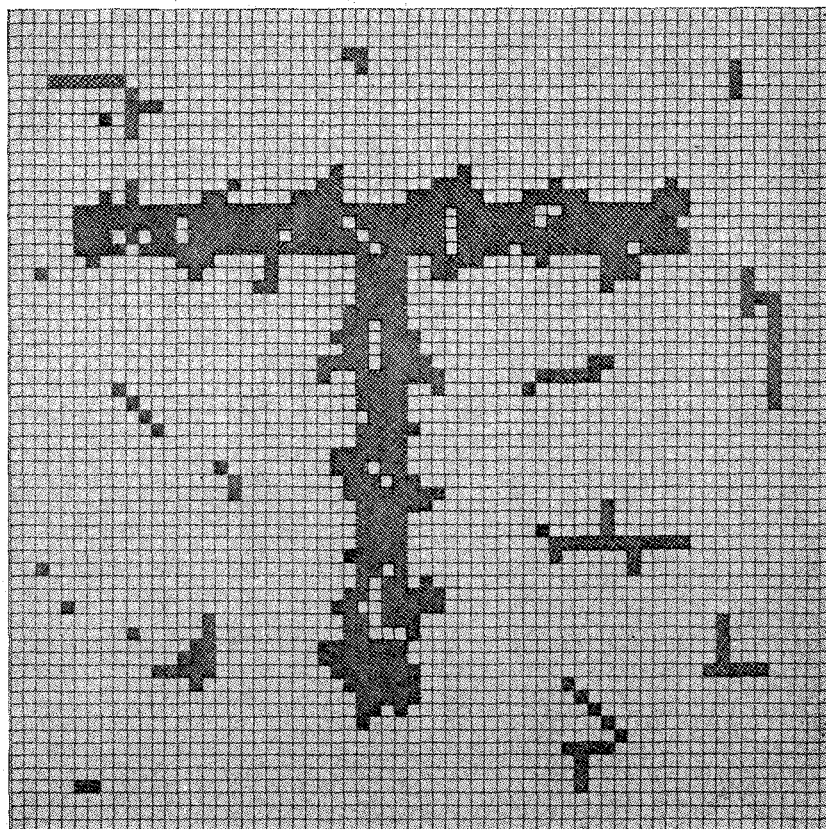


Fig. 14.

ra que se desea reconocer se halla sobre un enrejado constituido por gran número de celdillas cuadradas (fig. 14). Estas celdillas son negras o blancas ; a las primeras les asignamos el dígito 1 y a las segundas el 0. Cada celdilla, sea blanca o negra, se puede considerar como el centro de 9 celdillas que constituyen, a su vez, un cuadrado (fig. 15a). Independientemente de su color, la celda

central se sustituye por una blanca o negra, según que $n < \theta$ o $n > \theta$, siendo n el número de celdillas negras y θ el umbral. Por este procedimiento realizamos una transformación de la figura representativa de la escena. En el caso indicado en la figura 15a, si $\theta = 3$, la celda negra central permanece de este color ($n = 4$), mientras que en la figura 15b la celdilla central se transforma en blanca ($n = 2$).

Este método tiende a hacer desaparecer los puntos blancos rodeados de celdas negras (señales blancas situadas en líneas pertenecientes a la figura) y, recíprocamente, las manchas negras aisladas también desaparecen. Variando convenientemente el umbral,

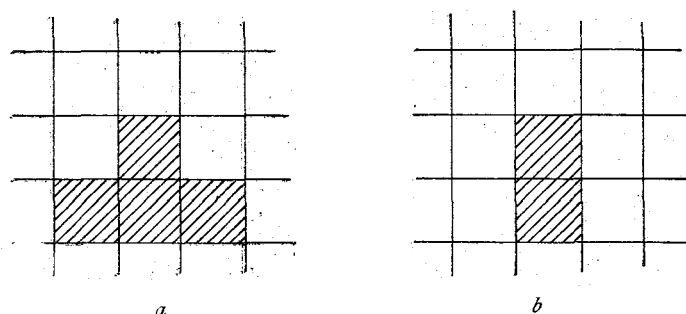


Fig. 15.

se puede lograr aumentar o disminuir el grueso de las líneas negras en fondo blanco o, recíprocamente, el de las líneas blancas en fondo negro.

¿Cómo se realiza la extracción de características? Limitándonos a las escenas visuales y más concretamente a las letras, las características importantes que deben extraerse han sido indicadas anteriormente. Número de rectas verticales o casi verticales, así como su posición y longitud. Análogamente número de rectas horizontales.

Los métodos utilizados son varios. Como ejemplo indicaremos el método del par de ventanas para detectar bordes [31]. Supongamos que se desea detectar un borde de la figura 16 situado entre la tercera y cuarta columna. En estas condiciones se sitúa un

par de ventanas (con el mismo significado que los cuadros), tal como se muestra en dicha figura. Comprende 8 celdas blancas y 8 negras; la ventana de la derecha limitada por la columna 3 se considera como positiva y la ventana de la izquierda, también limi-

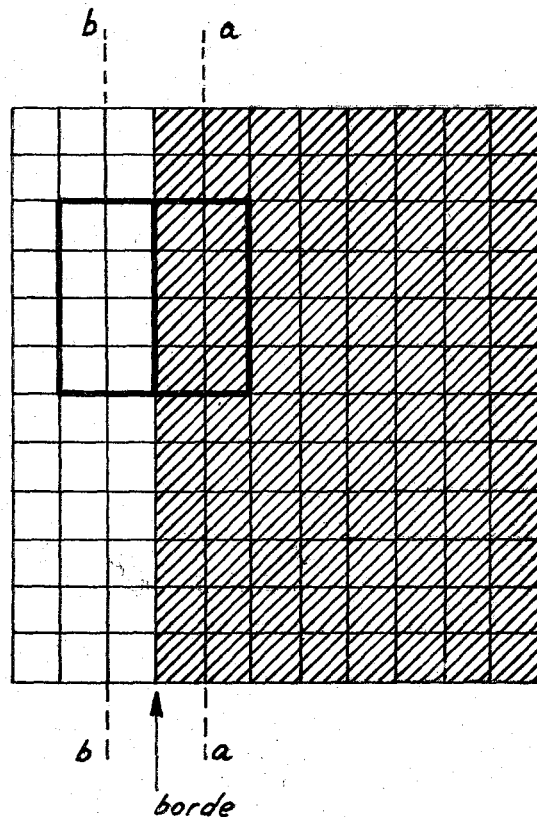


Fig. 16.

tada por dicha columna, como negativa. Las celdas negras dentro de la ventana positiva serán consideradas como positivas y las celdas negras dentro de la ventana negativa como negativas. Se suman algebraicamente todas las celdas negras que existen en ambas ventanas teniendo presente el signo antes convenido. Por ejemplo, en la posición indicada en la figura, existen ocho celdas positivas y ninguna negativa, ya que en la ventana de la izquierda no exis-

te ninguna celda negra: por lo tanto su suma es 8. Se elige ahora, arbitrariamente, un umbral, por ejemplo $\theta = 6$. Si la suma algebraica de las celdas negras es mayor —como en este caso— que el umbral ($8 > 6$), consideraremos que existe un borde en la figura. Hay que tener en cuenta que este procedimiento detecta un borde vertical y en una posición determinada. En efecto, si permaneciendo constante la posición de las ventanas el borde estuviera situado en la línea de puntos $a - a$, es decir, desplazado una columna hacia la derecha, la suma algebraica antes indicada sería 4 y por consiguiente al ser $4 < 6$ este procedimiento no detectaría ningún borde. Análogamente, si desplazamos el borde hacia la izquierda (columna $b - b$) —permaneciendo inalterable la posición de las ventanas— tampoco en este caso se detectaría el borde ya que la suma algebraica de las celdas negras también sería 4 ($4 < 6$). Pero no sólo este procedimiento indica concretamente la posición del borde, sino también su situación respecto a la transición entre blanco y negro. Define un borde en que la parte negra se halla precisamente a la derecha del mismo. Si ocurriera el caso inverso, es decir, que la parte negra se hallara a la izquierda, tampoco sería detectado el borde como puede verse fácilmente en la figura, ya que en este caso la suma algebraica antes indicada sería —8.

Otro ejemplo sencillo es la detección de segmentos rectilíneos de cuya presencia o ausencia deseamos estar seguros (fig. 17). Supongamos que la longitud del segmento es de doce celdas. Se sitúa una ventana en torno a la columna de celdas donde se supone debe estar situado el segmento. Esta ventana rectangular contiene tres columnas, es decir, 36 celdas. En estas condiciones se puede fácilmente adoptar un criterio sobre el número de celdas negras que deben existir para que pueda aceptarse la existencia del segmento rectilíneo [32].

Existen otros métodos similares para la obtención de características, todos ellos comprendidos en el procedimiento general denominado de la *máscara*. Creo interesante señalar aquí que las funciones realizadas por estos detectores de características son simi-

lares a las que verifican algunos organismos vivos y que hemos destacado en la primera parte del trabajo. La detección de bordes y contrastes en el *Límulus*, los muy importantes que hemos mostrado en la rana, los que se han observado en el sistema visual del gato —en el que la estructura neuronal de su cuerpo lateral geni-

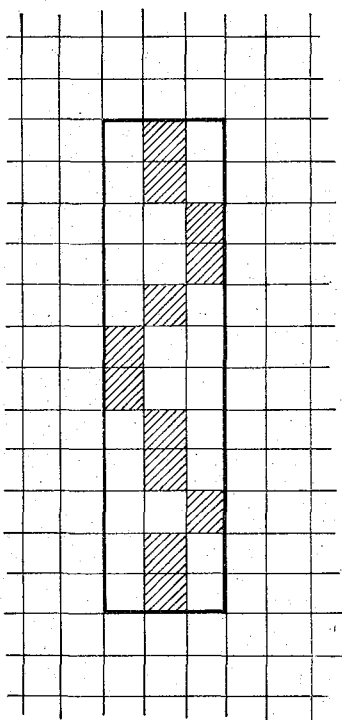


Fig. 17.

culado extrae señales y su corteza visual detecta segmentos lineales y bordes [33]—, así como otros ejemplos que hemos indicado, son semejantes en sus objetivos últimos (aunque no siempre en los procedimientos utilizados) a los detectores de características empleados en las máquinas. Conviene llamar la atención sobre este paralelismo que hasta ahora sólo se ha logrado a nivel de animales inferiores.

Realizaciones técnicas

Hasta ahora hemos explicado brevemente los principios generales en que se basa el reconocimiento de formas, destacando particularmente los procedimientos básicos de extracción de características.

¿Cómo se realiza técnicamente una máquina que siga estos principios? Uno de los procedimientos consiste en almacenar o registrar la figura que se desea leer (por ejemplo la letra indicada en

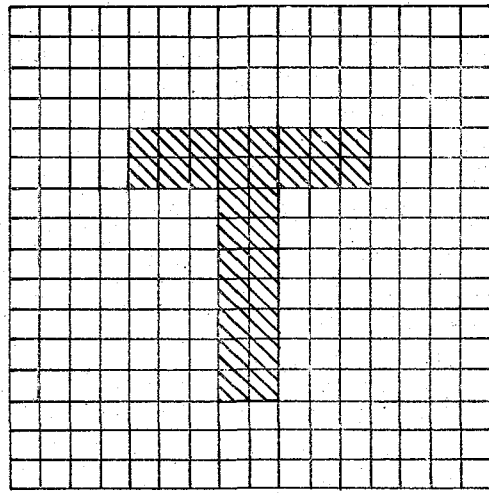


Fig. 18.

la fig. 18) en un mosaico de fotorreceptores que corresponden a las celdillas de la figura. Las celdas negras no excitan a los fotorreceptores, pero sí las blancas. Las corrientes producidas por los fotorreceptores correspondientes a las celdillas blancas —juntamente con la ausencia de corriente de las celdas negras— constituyen una configuración que representa a la figura mostrada y que puede almacenarse en un sistema de elementos biestables. De esta suerte, la configuración definida por los estados en que se encuentran estos biestables es una representación física, en determinado código, de la figura que se pretende leer.

Una variante de este método consiste en proyectar la escena en la pantalla de un tubo de rayos catódicos ; la imagen proyectada se transforma, mediante una técnica de barrido, en una configuración almacenada en un sistema de elementos biestables o de núcleos magnéticos y que en un determinado código representa la escena.

La extracción de las características, así como la eliminación del ruido inherente a la escena, puede realizarse partiendo del mosaico de elementos biestables donde aquélla se ha almacenado y utilizando los métodos antes indicados.

Por ejemplo, el sistema de nueve celdillas indicado en la figura 15 puede tener su equivalente en nueve biestables que representarán a la configuración de celdillas blancas y negras. Por consiguiente, la misma transformación que allí se mostró puede realizarse también con el sistema de biestables. Así, las salidas de éstos se conectan a un circuito de conmutación con umbral, el cual suministrará en la salida una señal que, por ejemplo, será positiva si el número de biestables con estados correspondientes a las celdas negras es mayor que el umbral y negativo en caso contrario. La señal de este circuito de conmutación se utiliza para que el biestable correspondiente a la celda central pase al estado debido. De esta forma se logra la transformación de la escena, eliminando el ruido, y, análogamente, se puede obtener la detección de características.

Partiendo de estos principios se construyen aparatos que eliminan el ruido y detectan las características a gran velocidad.

Otro procedimiento consiste en almacenar la configuración que representa la escena directamente en la memoria de una calculadora en lugar de hacerlo en un sistema de biestables. Entonces, la eliminación del ruido y la detección de características se verifica mediante un programa adecuado. El inconveniente que presenta, en algunos casos, es que este programa es excesivamente largo.

Los procedimientos descritos adolecen del defecto general de que la detección de características se verifica después de almacenar la escena en la memoria de la calculadora o bien en un sistema de biestables. En muchas ocasiones, el número de dígitos binarios

que se deben almacenar es tan grande que el procedimiento no resulta apropiado. Por ello se ha tratado de extraer las características por métodos que no exijan el almacenamiento previo de la representación numérica de la escena. Así, por ejemplo, el sistema óptico utilizado por Brain [34] y colaboradores se sirve de un sistema óptico y hace uso del par de ventanas para detectar bordes, como se ha indicado precedentemente. El sistema está constituido por 1.000 lentes que suministran 1.000 réplicas completas de la escena. Cada par de réplicas se observan a través de una ventana positiva y de una negativa por medio de fotocélulas que miden la luz procedente de cada ventana. Esta está constituida por una región transparente rectangular realizada en una máscara opaca. Las corrientes producidas por las fotocélulas correspondientes a las dos ventanas se comparan y de esta comparación se puede deducir si existe o no un borde en la posición observada por el par de ventanas. Las observaciones obtenidas de los bordes se agrupan en familias.

Por el mismo procedimiento se puede también detectar otras características además de los bordes, sustituyendo las ventanas rectangulares por otros tipos de ventanas.

Otros métodos de realizar físicamente la detección de características utilizan un «flying-spot scanner» para leer la escena y un conjunto de circuitos que operan, partiendo de la información procedente de aquél. Mediante estos circuitos se puede detectar eficientemente características de la escena, como por ejemplo bordes, longitud de segmentos rectilíneos, variación de brillo, etc.

RECONOCIMIENTO DE FORMAS MEDIANTE APRENDIZAJE

Cuando se trata de reconocimiento de escenas relativamente complejas, no simplemente letras, el problema se presenta difícil. Hemos visto que el reconocimiento de escenas se reduce a un problema de clasificación. Pero si se trata, por ejemplo, de reconocer una cara con expresiones diferentes o en diferentes épocas, la clasificación no es sencilla, ya que entre los distintos aspectos de la misma pueden existir notables diferencias.

Si se desea establecer el reconocimiento de una cosa tan vulgar como una silla, nos encontramos con la dificultad de describir todos los tipos de sillas. Esto nos indica que en el fondo el problema del reconocimiento de escenas se halla dentro del aprendizaje de conceptos generales. Este aprendizaje se debe realizar basándose en ejemplos concretos, que dan lugar a un proceso de inducción y a formar hipótesis partiendo de este proceso [35].

Como ejemplo de sistemas de reconocimiento de formas, en los que las máquinas aprenden a realizar aquél, es decir, se les adiestra para ello, explicaremos brevemente dos que pueden considerarse como clásicos: el Pandemonium de Selfridge [36] y el Perceptron de Rosenblatt [37].

Pandemonium

Las características esenciales de este modelo es que puede llegar a reconocer ciertas configuraciones por un proceso adaptivo capaz de progresar por sí mismo. Se puede aplicar a escenas no especificadas previamente. En definitiva, utiliza una forma de aprendizaje. Se presentan al modelo ejemplos de escenas tomadas de un conjunto de ellas, informando cada vez al sistema la escena particular que se le ha presentado. Después de cierto tiempo de aprendizaje, el modelo reconoce correctamente la escena, sin necesidad de información alguna. Se trata de imitar de una forma primaria el comportamiento del hombre en los problemas de reconocimiento.

Otra característica interesante de este modelo es que el proceso de la información se realiza en paralelo. Según el propio Selfridge señala, ello tiene dos ventajas: a) El tratamiento de datos en paralelo es más fácil, en general y además es la forma más natural de manejarlos. b) Es más sencillo modificar un conjunto de módulos casi independientes, que una máquina en que todas sus partes interaccionan inmediatamente y de una manera compleja.

El principio general del modelo consiste en un conjunto de *demonios cognoscitivos* (denominación caprichosa derivada, posiblemente, de que sus atributos son análogos a los del diablillo de

Maxwell), cada uno de los cuales representa cada una de las posibles escenas del conjunto cuyo reconocimiento se quiere establecer. Estos demonios determinan, mediante los cálculos necesarios, su analogía, el grado de semejanza con la escena procedente de la entrada (la escena que se desea reconocer), que es observada *simultáneamente* por todos ellos. El grado de similitud representa la salida de cada demonio. Existe otro de jerarquía superior, el *demonio de decisión*, cuya misión es discriminar entre aquellas salidas la que ofrece una mayor semejanza. La escena representada por el demonio cuya salida es mayor es elegida por el de decisión, como correspondiente a la clase a que pertenece la escena representada en la entrada. Se logra, pues, de este modo, una identificación de ésta.

Es evidente que cuanto mayor sea el número de clases distintas en que puedan catalogarse las escenas, el proceso de reconocimiento será más complicado. En algunos casos, la escena presentada equivale a una función de un conjunto de características, siendo cada una de ellas común a varias escenas y, análogamente, su ausencia también puede serlo de otras.

En la figura 19 se indica esquemáticamente la estructura del Pandemonium. En realidad existen cuatro niveles o capas de elementos: los que almacenan datos, los que calculan, los cognoscitivos y el que decide. Los primeros convierten la escena de entrada en forma digital y transmiten la información al segundo nivel, donde se realizan los cálculos convenientes y se extraen ciertas características. Esta información pasa, a su vez, al tercer nivel, donde cada elemento (demonio cognoscitivo) entra en conocimiento del conjunto de características que le llegan y da un peso a las mismas, calculando el valor que se debe suministrar al último nivel (demonio de decisión), para que éste pueda realizar la selección.

El proceso de aprendizaje se realiza dando pesos variables a las salidas de los elementos del 2.º nivel (subdemonios). Cada elemento cognoscitivo suma de forma ponderada (es decir, teniendo en cuenta aquellos pesos) las salidas de todos los elementos del 2.º nivel. Estas sumas ponderadas diferirán de un elemento cog-

noscitivo a otro, debido únicamente a los pesos variables asignados, ya que si éstos no existieran o fueran los mismos, aquellas sumas serían iguales para todos los elementos cognoscitivos. Los pasos asignados a los elementos calculadores *se varían por los pro-*

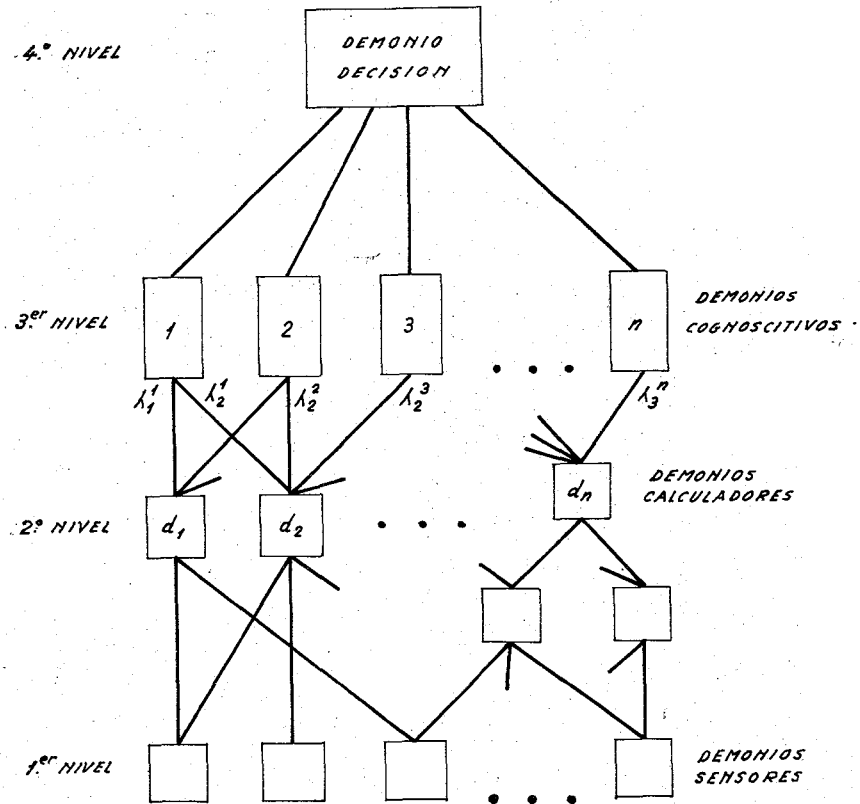


Fig. 19.

pios elementos cognoscitivos, con objeto de lograr un proceso operativo óptimo.

Se asignan, también, pesos variables y con el mismo objeto entre los elementos del nivel calculador.

Con este proceso de aprendizaje realizado en estos dos niveles, optimizando los pesos variables asignados, se logra que el sis-

tema vaya aprendiendo á reconocer escenas y progresando en su comportamiento.

Según Selfridge su modelo puede evolucionar. En efecto, se puede realizar también una selección entre los elementos del 2.º nivel. Hay que tener presente que, inicialmente, se escoge un conjunto de estos elementos, que suponemos serán útiles y a los cuales se les asigna pesos arbitrariamente. En el primer proceso adaptivo, al variar los pesos se optimiza el sistema, pero ello no implica que todos los elementos seleccionados sean necesariamente buenos. Es necesario eliminar los elementos cuyo funcionamiento es deficiente, es decir, que no ayudan mucho al progreso en el comportamiento del sistema. Hay que generar nuevos elementos que se pueden obtener combinando adecuadamente y de diferentes maneras las componentes de los antiguos. Estos nuevos elementos sobrevivirán o no según sus méritos.

Partiendo de este modelo, se ha realizado una simulación en calculadora, aplicada particularmente al reconocimiento de letras escritas a mano [38].

Perceptron

Este modelo de reconocimiento de formas debido a Rosenblatt y colaboradores en la Universidad de Cornell (USA) sigue, en cierto modo, el modelo propuesto por Selfridge, si bien se precisa más el principio de aprendizaje, premiando y castigando a las componentes del sistema según sus méritos y el trabajo que realizan. Pero sus objetivos son análogos, es decir, se trata de un sistema que *aprende* a reconocer escenas de un conjunto, después de un número finito de ensayos.

En la figura 20 se muestra un esquema del Perceptrón elemental de Rosenblatt. Consta de tres tipos de unidades distintas. Las unidades sensoriales S_i , las de asociaciones A_i y las de respuesta R_i . Las primeras constituyen la retina, formada por fotocélulas que reciben los estímulos luminosos y los transforman en señales eléctricas. El número de conexiones que entran en cada unidad A_i , procedente de las S_i , está especificado, pero no así los puntos de

la retina donde tienen su origen, que están seleccionados estocásticamente. Es decir, las conexiones están realizadas al azar. Las señales eléctricas procedentes de las unidades sensoriales pueden ser excitadoras (peso + 1) o bien inhibitorias (peso - 1).

Si llamamos θ_i al umbral (umbral fijo) correspondiente a la

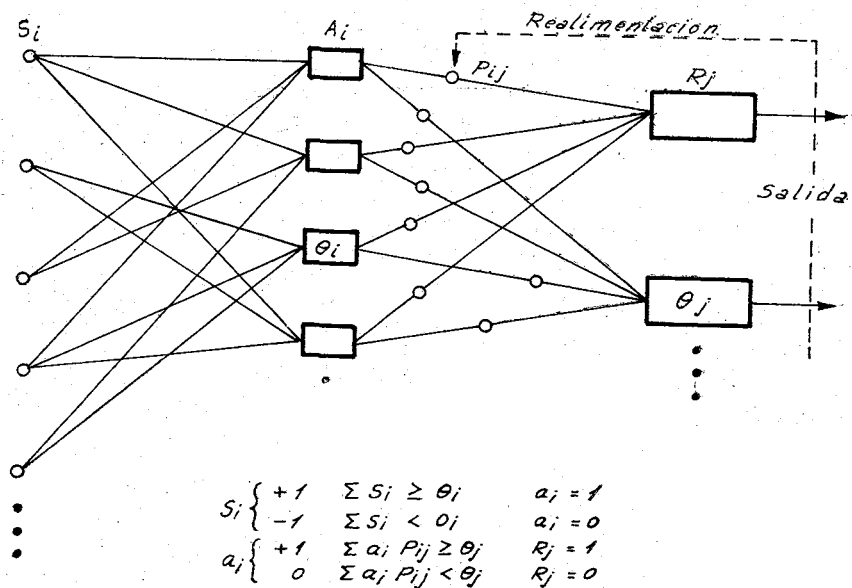


Fig. 20.

unidad asociativa A_i para que dispare será preciso que la suma algebraica de los pesos de las entradas sea igual o mayor que θ_i , o sea :

$$\begin{aligned} \sum S_i &\geq \theta_i & a_i &= 1, \\ \sum S_i &< \theta_i & a_i &= 0. \end{aligned}$$

donde a_i representa la respuesta de la unidad de asociación A_i

Las unidades A_i no están conectadas estocásticamente con las R_i , sino que a cada unidad de éstas, están conectadas todas las sa-

lidas de A_i , según se ve en la figura 20. Ahora bien, cada unidad de asociación posee un elemento de memoria que está indicado por el círculo P_{ij} y realiza una función similar a la unión sináptica variable en el modelo de reflejo condicionado. Estos elementos de memoria almacenan una señal, que puede variar y que define el peso que se debe dar a la salida de la correspondiente unidad A_i . Es decir, según el valor de la señal almacenada en la memoria, la amplitud de la señal de salida de la correspondiente A_i es mayor o menor. En estas condiciones, la señal que llega a una unidad de respuesta será $a_i P_{ij}$ y para que esta unidad R_j dispare, será preciso :

$$\sum a_i P_{ij} \geq \theta_j \quad R_j = 1;$$

siendo θ_j el umbral de la citada unidad y naturalmente, si

$$\sum a_i P_{ij} < \theta_j,$$

se tiene $R_j = 0$.

Entre las unidades R_j y las A_i existen bucles de realimentación que permiten corregir los errores de las salidas de las unidades A_i , es decir, reforzar los valores almacenados en la memoria, de estas unidades. Si, por ejemplo, para la correcta clasificación de los estímulos de entrada, la salida de R_j debe ser 1, y realmente ocurre así, no se verifica ninguna modificación en la memoria de las A_i , pero si la salida es errónea, un cero, entonces todos los pesos P_{ij} de las entradas de la unidad R_j deben sufrir un aumento con objeto de aumentar la señal total que llega a la unidad de respuesta. En caso contrario, es decir, si en lugar de un 1, debía ser un cero, se deben disminuir los pesos con objeto de que la señal total que llega a la unidad R_j no sea suficiente para disparar.

Este proceso de realimentación o de refuerzo puede realizarse por medio del operador humano, siguiendo una determinada estrategia, o bien mediante una calculadora digital, siguiendo un programa adecuado.

Supongamos que con este perceptrón así bosquejado deseamos

reconocer todas las letras del alfabeto. Las unidades sensoriales S_i constituirán un mosaico de células fotoeléctricas en forma de retina. Sobre esta retina se proyecta la letra que se desea reconocer, por ejemplo, la A. De esta forma quedarán activadas, es decir, incidirá la luz sobre una parte de las fotocélulas, mientras que el resto quedará sin activar. Si llamamos n al número de fotocélulas de la retina, el número de configuraciones posibles, teniendo en cuenta las fotocélulas activadas y las no activadas es 2^n .

Hay que tener en cuenta que todas las fotocélulas constituyen un conjunto de dígitos binarios (fotocélulas activadas + 1, fotocélula no activada 0). Es decir, cuando se proyecta una letra sobre la retina, un determinado número de fotocélulas se activarán (incidirá la luz sobre ellas) mientras que las otras no se activarán. Podemos imaginar que todas las fotocélulas están numeradas, es decir, que su posición queda perfectamente determinada respecto a las otras. En estas condiciones podemos imaginar un número binario 1 0 1 1 1 0 1 ... representativo del estado general de todas las fotocélulas en un instante determinado, es decir, para un determinado conjunto de estímulos luminosos. Cada uno de estos conjuntos tendrá n dígitos, siendo n el número de fotocélulas. Por consiguiente, el número de configuraciones posibles será el número de variaciones con repetición de orden 2 de n elementos, es decir, $V_{2,n} = 2^n$.

Si pretendemos discriminar las 29 letras del alfabeto, se dividirán las 2^n configuraciones posibles en 29 clases. En el caso más general que se deseen reconocer las diferentes posiciones, corresponderán a un conjunto de estímulos de entrada, pero todos estos conjuntos corresponden a la misma clase.

Ahora bien, el reconocimiento de estas letras, en las unidades de respuesta, se verificará mediante un cierto código, establecido *a priori*, de forma que a cada letra le corresponda una determinada configuración de las salidas de las unidades. Cada una de estas configuraciones tendrá m dígitos, siendo m el número de unidades de salida. El número de configuraciones posibles será 2^m .

En estas condiciones, y colocada una letra en la posición adecuada, se activan las fotocélulas correspondientes y a su vez las A_i

apareciendo una respuesta en la salida. Si ésta es errónea, es decir, la configuración de salida no es igual a la prevista, se modifican los elementos de memoria P_{ij} , asociados con las unidades A_i conectadas con las redes R_j erróneas, hasta lograr la respuesta correcta. De esta forma la máquina ha aprendido a reconocer una letra determinada.

De forma análoga se puede lograr el aprendizaje para las otras letras del alfabeto. Cabe pensar que al modificar los pesos correspondientes, con objeto de lograr el reconocimiento de una determinada letra, la máquina olvide alguna de las ya aprendidas. Puede ocurrir, pero no es probable, ya que el número de A_i es muy grande (500) y no siempre se activan todas y las activadas no son las mismas para cada letra. Por consiguiente, la modificación de los pesos P_{ij} para una determinada letra no interfiere con los pesos previamente modificados para otra letra.

Rossmblatt ha estudiado la estructura matemática del perceptrón de forma rigurosa, encontrando que el problema del entrenamiento del sistema para lograr una respuesta correcta a todos los estímulos tiene solución, es decir, que después de una serie finita de ensayos se puede lograr una serie de pesos P_{ij} que realicen la clasificación correcta de las clases prefijadas.

En la figura 21 se muestra la forma de realizar el proceso de realimentación y, por consiguiente, la variación de los pesos P_{ij} en el perceptrón Mark I. Si la suma algebraica de las señales procedentes de los elementos sensoriales es superior o igual a θ_i , la unidad A_i se dispara y se cierran los relés indicados, apareciendo una señal en la entrada de R_i que dependerá del valor almacenado en la memoria P_{ij} . La señal almacenada en P_{ij} es simplemente una diferencia de potencial que se puede variar mediante un potenciómetro. Si es preciso incrementar esta señal para que dispare R_i , ello se verifica moviendo convenientemente el cursor del potenciómetro citado, realizando un proceso inverso si se desea disminuir la señal de entrada en R_i .

El Perceptrón ha sido muy discutido. Glushkov [39] señala que el proceso de aprendizaje del modelo que acabamos de describir y otros similares, difieren considerablemente del que realiza el

cerebro humano. Sin embargo, posteriormente, el propio Rosseblatt [40] ha propuesto otro tipo de Perceptrón más completo que consta de diferentes capas de neuronas ; las que se hallan en la primera se conectan con la retina y su salida constituye la retina para las neuronas de la segunda capa y así sucesivamente. Por otro lado se realizan acoplamientos entre las capas superiores y las inferiores. Se ha verificado una simulación de este modelo en calculadora y su comportamiento presenta un mayor parecido a lo que realizan los seres humanos al reconocer formas.

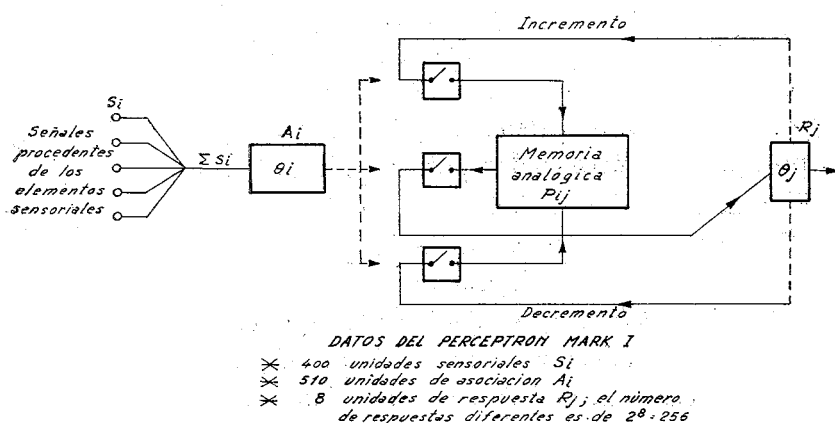


Fig. 21.

ESTADO ACTUAL DE LAS INVESTIGACIONES Y DIRECTRICES FUTURAS

Los progresos mayores en el reconocimiento de formas se han logrado, evidentemente, en el proceso más sencillo y, al mismo tiempo, de gran interés comercial, cuál es el reconocimiento de caracteres impresos. Su gran importancia en el problema comunicación hombre-máquina ha motivado que se esté trabajando con gran intensidad en este campo y se hayan publicado multitud de comunicaciones sobre estas materias.

En esta directriz, los primeros resultados se obtuvieron en sistemas de lectura de símbolos magnéticos sencillos, utilizados en cheques bancarios, con objeto de registrar el número de la cuenta

corriente y se continuó con signos menos estilizados, en cartas de crédito, por ejemplo, y en documentos escritos a máquina. En general, el tipo de símbolos (letras, números) está estrechamente controlado, tanto bajo el punto de vista de su forma, como de su posición, ya que cualquier variación de la prevista, en uno u otro caso, puede dar lugar a errores en la lectura.

Después de alcanzar estos primeros objetivos se fueron logrando perfeccionamientos en esta línea. Por ejemplo, se pueden señalar las investigaciones de Liu [41] sobre sistemas de reconocimiento de caracteres de imprenta de tipos diferentes, realizados con un conjunto elevado de datos (70.000 caracteres de máquina de escribir, en 13 diferentes formatos). Los resultados demostraron que el funcionamiento no es todo lo bueno que sería de desear, cuando la máquina de escribir está poco ajustada, cuando existen mutilaciones (aunque sean pequeñas) en los caracteres, etc.

Mayores dificultades aparecen en el reconocimiento de la escritura a mano. El problema se ha abordado imponiendo severas restricciones respecto a las dimensiones, formato y posición de las letras o números. Así se han hecho pruebas utilizando hojas para la programación en Fortran, escribiendo caracteres del alfabeto separados entre sí, arbitrariamente, por diferentes personas. Se ha utilizado una cámara de televisión para examinarlos, detectores de bordes y un clasificador lineal. A pesar de ello, los errores cometidos son del orden del 20 por 100, en el caso más favorable.

Es interesante señalar que las investigaciones en reconocimiento de formas no solamente se han dirigido a la lectura de textos, sino también en otras direcciones. Podemos citar como ejemplos: Análisis automático de glóbulos rojos, cromosomas y preparaciones análogas, mediante microscopio, con un «flying spot scanner», un fotómetro, un conversor analógico digital y el correspondiente equipo de registro digital [42].

Se están haciendo ensayos para el reconocimiento automático de fotografías aéreas, como por ejemplo, lagos, bosques, huertos, ferrocarriles, etc.

Se ha logrado también automatizar el monótono y laborioso proceso de localizar y medir la trayectoria de las partículas que apa-

recen en las fotografías de la cámara de burbujas. Como es sabido, esta cámara se usa actualmente mucho para el estudio de la producción e interacción de partículas nucleares de elevada energía. Se ensayan procedimientos de reconocimiento automático de electrocardiogramas, en los que se trata de imitar la forma cómo realizan esta tarea los expertos en este campo.

Aunque no hemos tratado del reconocimiento del lenguaje hablado, también hay que señalar que se observan progresos en este campo [43]. Hasta ahora se han logrado resultados aceptables en el reconocimiento de palabras aisladas y sílabas, pero no así en lenguaje hablado continuo, es decir, en frases largas. Dado su enorme interés en el problema de la comunicación hombre-máquina, muchos investigadores dedican sus esfuerzos en esta dirección, pero creemos que una solución completa, es decir, el reconocimiento de palabras pronunciadas de forma continua, como se habla normalmente, no se halla próxima.

Muchos de los modelos realizados, tanto en reconocimiento de formas visuales como acústicas, se han simulado en una calculadora digital. Las posibilidades de estas máquinas, en este aspecto, son inmensas.

Las investigaciones en este campo se realizan con el curso de psicólogos, psiquiatras, neurofisiólogos, matemáticos, físicos, ingenieros. Sin embargo, a pesar del esfuerzo combinado de tantos científicos de orígenes distintos, estas investigaciones se han llevado a cabo, hasta ahora, reduciendo y simplificando el problema general. Se considera simplemente un sistema bidimensional de símbolos discretos y se pretende determinar la clase a que corresponde.

Pero aún dentro de esta limitación de objetivos, el obstáculo mayor se halla en la elaboración previa, es decir, en la extracción de las características esenciales de una escena, a cuyo efecto no se pueden establecer principios generales, desgraciadamente, sólo pueden utilizarse procesos empíricos que se refieren a cada tipo de reconocimiento.

Posiblemente una mayor compenetración entre investigadores

procedentes de campos distintos dará como resultado lograr métodos más generales de investigación.

Hay que tener presente que, incluso entre los psicólogos y los neurofisiólogos, el problema de la percepción se estudia por procedimientos distintos y es curioso hacer notar que la calculadora digital ha tenido la virtud de servir de lazo de unión entre ambos tipos de investigadores.

En efecto, debemos tener presente que la calculadora es básicamente una red de conmutación que podemos asimilar a un sistema nervioso primario. Evidentemente, el sistema nervioso es mucho más complejo que una máquina calculadora, pero el hecho de que muchos procesos complejos de tipo psicológico (reconocimiento de escenas) puedan programarse en la máquina (que es un modelo de sistema nervioso primario) permite interpretar este proceso a nivel neurofisiológico.

Hasta ahora poco se ha hecho para coordinar los esfuerzos de los psicólogos y de los neurofisiólogos para lograr modelos conjuntos; es de esperar que con la ayuda de los ordenadores se lograrán resultados más satisfactorios.

Hay que tener presente que la máquina, como en otros muchos casos, sólo actúa como herramienta en la que el modelo se escribe, utilizando su gran rapidez y flexibilidad. Hemos visto que el reconocimiento de escenas se convierte en un proceso complejo cuando se aparta de los casos más simples. Precisamente por ello son muy útiles las calculadoras, ya que son aptas para describir modelos complejos.

Sólo unas palabras para terminar. Hemos estudiado paralelamente el proceso de reconocimiento de formas por los seres vivos y por las máquinas, y hemos tratado de señalar las analogías existentes entre ambos, en particular cuando en las máquinas se utilizan los métodos de aprendizaje. Los detectores de características de escenas, que constituyen el órgano fundamental de todo sistema de percepción, tienen, actualmente, severas limitaciones, están diseñados para tareas muy restringidas. El único detector de características, que podríamos llamar universal, lo posee el hombre y estamos muy lejos de lograr un modelo, aunque sea sola-

mente aproximado, del proceso de la información desde los órganos sensoriales al sistema nervioso central. Y, sin embargo, a pesar de sus dificultades, éste es el camino. Los detectores de características desarrollados basándose en las teorías estadísticas de decisión o, en general, en otros procesos matemáticos, no pueden sustituir a la perfección de los procesos biológicos. El proceso de reconocimiento de formas, como todos los que se hallan englobados en los sistemas de inteligencia artificial, para progresar realmente necesitan basarse en el conocimiento de los sistemas biológicos que son, en definitiva, a los que se trata de imitar. Por ello, aunque las dificultades son muy grandes y los problemas que se presentan parecen insalvables, seguir otro camino no es ignorar la única directriz que tenemos.

Por otro lado, el estudio del sistema nervioso, bajo el punto de vista cibernético, es decir, tratando de investigar cómo las redes neuronales adquieren, procesan, almacenan y emiten la información, es posible que no exija, en principio, una explicación en términos moleculares, sino más bien (haciendo abstracción del detalle molecular) el desarrollo de una teoría matemática, de unas leyes que engloben los principios generales implicados. Nos hallamos en una situación en que se puede establecer un paralelo, como dice Reichart, «con el estado de la Física, cuando Maxwell estableció sus ecuaciones, que representan uno de los principios fundamentales de la Naturaleza y pasó a establecerlas, a pesar de que en aquella época las propiedades atómicas de la materia no se conocían aún. Desde este punto de vista se comprende que el sistema nervioso atraiga el interés de los físicos en número creciente».

Sin duda la complejidad de estos problemas exigirá para llegar a resultados razonables la contribución de gran número de científicos durante muchos años.

REFERENCIAS BIBLIOGRÁFICAS

- [1] WIENER, N.: *Cybernetics*. The M. I. T. Press, Cambridge, Mass., 1961.
- [2] *Survey of Cybernetics*, Edit. por ROSE, J. Iliffe Books, Londres, 1969.
- [3] Primer Simposium de Biónica organizado por la División de Medicina Aeroespacial y la de Investigación y Tecnología de las Fuerzas Aéreas. Dayton, Ohio, mayo 1960.
GARCÍA SANTESMASES, J.: *La Biónica, nueva frontera del conocimiento*. «Revista de Ciencia Aplicada», nov.-dic. 1965.
- [4] DEUTCH, J. A.: «British Journal of Psychology», VI, 46, 1955.
- [5] PITTS, W. & MCCULLOCH, W. S.: *How we know Universals?* «Bull. Math. Biophys.», vol. 7, 1945.
- [6] HARTLINE, H. K., WAGNER, H. G. & RATLIFF, F.: *Inhibition in the eye of Limulus*. «J. Gen. Physiol.», 39, 1956.
TOMITA, T.: *Mechanism of lateral inhibition in the eye of Limulus*. «J. Neurophysiol.», 21, 1958.
- [7] HARTLINE, H. K. & RATLIFF, F.: *Inhibitory interaction of receptor units in the eye of Limulus*. «J. Gen. Physiol.», 40, 1957.
TOMITA, T., KIKUCHI, R. y TANAKA, I.: *Excitation and inhibition in lateral eye of horseshoe crab*. «Electrical Activity of Single Cells», Igakushion, Hongo, Tokyo, 1960.
- [8] RATLIFF, F., HARTLINE, H. K. & LANGE, D.: *The Dynamics of lateral inhibition in the compound eye of Limulus*. I. Symposium on the compound eyes. Estocolmo, 1966.
- [9] GARCÍA SANTESMASES, J. y RUBIO ROYO, F.: *Modelo de inhibición lateral en el ojo compuesto del Limulus*. «Anal. de la Real Soc. Esp. de Fis. y Quím.», tomo LXIV, jul.-ag. 1968.
- [10] RATLIFF, F.: *Mach Bands: Quantitative studies on neural networks in the retina*. Holden-Day, Inc., 1965.
- [11] MORENO DÍAZ, R., RUBIO ROYO, F. y MIRA MIRA, J.: *Aplicación de las transformaciones integrales al proceso de datos en la retina*. «Revista de Automática», jul.-ag.-sep. 1969.
- [12] GARCÍA SANTESMASES, J., RUBIO ROYO, F. y MIRA MIRA, J.: *Modelo electrónico de inhibición lateral en el ojo del Limulus*. «Anal. de la Real Soc. Esp. de Fis. y Quím.», tomo LXV, jul.-ag. 1969.

- [13] MACH, E.: *Über die Wirkung der räumlichen Vertheilung des Lichtreizes auf die Netzhaut, I-V*. «Sitzungsberichte der mathematisch-naturwissenschaftlichen Classe der Kaiserlichen Akademie der Wissenschaften, 1865-68. En el libro citado en la ref. 10 se encuentran estos trabajos traducidos al inglés.
- [14] REICHARDT, W. & MC GINITIE, G.: *On the lateral inhibition*. «Kybernetik», 1, 1962.
- [15] LETTWIN, J. Y., MATURANA, H. R., McCULLOCH, W. S. & PITTS, W. H.: *What the Frog's eye tells the Frog's brain?* «Proc. of the IRE», vol. 47, nov. 1959.
- [16] LETTWIN, J. Y., MATURANA, H. R., PITTS, W. H. & McCULLOCH, W. S.: *Two remarks on the visual system of the Frog*. «Proc. of the Symposium on Principles of Sensory Communication», jul.-ag. M. I. T. 1959.
- [17] MORENO DÍAZ, R.: *An analytical model of the «Bug Detector» Ganglion Cell in the Frog's retina*. «Proc. of the Bionics Symp», Ohio, 1966.
- [18] SUTRO, L.: *Proposed electronics to represent the properties of a Frog's eye*. «Proc. of the Bionics Symposium», Ohio, 1966.
- [19] HUBEL, D. H. & WIESEL, T. N.: *Integrative action in the cat's lateral geniculate body*. «J. Physiol.», 155, 1961.
- [20] CAMPBELL, F. W.: *Trends in physiological optics*. Conferencia pronunciada en el Curso «Optical Information Processing by Organism and by Machines». Escuela Internacional «Enrico Fermi» de Física. Varenna, Italia, 1968.
- [21] CAMPBELL, F. W. & GREEN, D. G.: *Optical and retinal factors affecting visual resolution*. «J. Physiol.», 181, 1965.
- [22] ENROTH-CUGELL, CH. & ROBSON, J. G.: *The contrast sensitivity of retinal ganglion cells of the cat*. «J. Physiol.», 187, 1966.
- [23] COOPER, G. F. & ROBSON, J. G.: *Successive transformations of spatial information in the visual system*. «Symposium on Pattern Recognition», Teddington, Inglaterra, 1968.
- [24] CAMPBELL, F. W., CLEELAND, B. G., COOPER, G. F., ENROTH-CUGELL, CH.: *The angular selectivity of cortical cells to moving gratings*. «J. Physiol.», 198, 1968.
- [25] CAMPBELL, F. W. & KULIKOWSKI, J. J.: *Orientalional selectivity of the human visual system*. «J. Physiol.», 187, 1966.
- [26] GARCÍA SANTESMASES, J. G., RUBIO ROYO, F., ARES ESCOLAR, L., MIRA MIRA, J. & HERNÁNDEZ VERDUZCO, F.: *Harmonic Analysis in the Visual Pathway of the higher mammals*. «Proc. of the Int. Congress of Cybernetics», Londres, sep. 1969.
- [27] SWIGERT, C. J. & VOGELHUT, C. J.: *Spatial filtering in a Neuron Network Model*, 20th Annual Conference on Engineering in Medicine and Biology, nov. 1967, Mass.
- [28] RUBIO ROYO, F., ARES ESCOLAR, L., MORENO DÍAZ, R. y MIRA MIRA, J.: *Un modelo del proceso de datos visuales*. «Actas del Congreso de Automática», abril 1970, Madrid (en prensa).

- [29] LEIBOVIC, K. N.: *A model for information processing with reference to vision*. «Bull. Math. Bio-Phys.», 25, 1965.
- [30] RUNGE, R. G., UEMURA, M. & VIGLIONE, S. S.: *Electronic Synthesis of the Neural Network in the Pigeon Retina*. «Proc. of the Bionics Symposium», Ohio, 1966.
- [31] NILSSON, N. J.: *Adaptive Pattern Recognition: A Survey*. «Proc. of the Bionics Symposium», Ohio, 1966.
- [32] SCHUH, J. F.: *What a Robot can and cannot do*. En «Survey of Cybernetics». Ed. por J. Rose. Iliffe Books. Londres, 1969.
- [33] HUBEL, D. H. & WIESEL, T. N.: *Receptive Fields, Binocular Interaction, and Functional Architecture in the Cat's Visual Cortex*. En «Pattern Recognition». Ed. por L. Uhr, John Wiley, Nueva York, 1966.
- [34] BRAIN, A. E. & MUNSON, J. H.: *Graphical Data Processing Research Study and Experimental Investigation*. «Stanford Research Institute Report», núm. 22, 1966.
- [35] *Pattern Recognition*. Ed. por L. Uhr, John Wiley, Nueva York, 1966.
- [36] SELFRIDGE, O.: *Pandemonium. A Paradigm for learning*, en «The Mechanisation of Thought Processes», Her Maj. Stat. Office. Londres, 1959.
- [37] ROSENBLATT, F.: *The perceptron. A perceiving and recognizing automaton*. «Cornell Aeronautical Lab. Rept. 85-460-1». En 1957.
- [38] SELFRIDGE, O. & NEISSER, G.: *Pattern Recognition by Machine*. «Scientific American», vol. 203, ag. 1960.
- [39] GLUSHKOV, V. M.: *Contemporary Cybernetics*. En «Survey of Cybernetics». Ed. por J. Rose. Iliffe Books. Londres, 1969.
- [40] ROSENBLATT, F.: *A Comparison of several perceptron models*. En «Self-Organizing Systems», Spartan Books, Washington, 1962.
ROSENBLATT, F.: *Principles of Neuro-Dynamics*. Spartan Books, Washington, 1962.
- [41] LIU, C. N.: *A programmed algorithm for designing multifont character recognition logics*. «I. E. E. E. Trans. Electronic Computers», vol. EC-13, oct. 1964.
LIU, C. N. & SHELTON, G. L.: *An experimental investigation of a Mixed-font print recognition system*. «I. E. E. E. Trans. Electronic Computer», vol. EC-15, dic. 1966.
- [42] PREWITT, J. M., MAYALL, B. H. & MENDELSON, M. L.: *Pictorial data processing methods in microscopy*. «Proc. Soc. Photographic Instrumentation Engrs.», jun. 1966.
- [43] FLANAGAN, J. L.: *Speech Analysis, Synthesis and Perception*. Academic Press, Nueva York, 1965.