

Dirección y Administración:
Av. Martín García 653
(1268) Capital Federal
Tel. 362-0643/0747
Argentina



sumario

Mirador electrónico	601
Control con lógica difusa	
Interfaz oral para computadoras personales orientada al discurso continuo	602
Aldo Sigura, Daniel Zapata y Leonardo Rufiner	
Se presenta un sistema de reconocimiento automático de la voz humana, utilizando redes neuronales e inteligencia artificial.	
Usos de las redes de Petri generalizadas para modelar sistemas dinámicos a eventos discretos	610
Ing. Armando W. Colombo	
Se trata de una excelente herramienta de diseño para los sistemas complejos que se pueden separar en operaciones elementales.	
Hipertermia y campos electromagnéticos en el tratamiento de tumores	616
Ing. Victor H. Padula Pintos e Ing. Benito Alvarez Ovide	
Un resumen cronológico.	
Ruidos y perturbaciones eléctricas: Su importancia y control	622
Ing. Alberto M. Godel	
Se analizan las causas, forma de propagación, espectro de frecuencias y métodos de control de las interferencias radioeléctricas.	
La matemática en la formación del ingeniero	627
Ing. Ulises J. P. Cejas	
Una reflexión sobre el importante papel que juega el conocimiento matemático en el profesional de ingeniería.	
Desarrollo de servicios de interrupción para MS-DOS	629
Ings. Edgardo Ferro y Sergio Crocioni	
Criterios para el diseño de cargadores para pilas de Ni-Cd	633
Ing. Jorge A. Alberto	
Se estudian los métodos de recarga de baterías y se describe un circuito cargador de corriente constante y control por tensión máxima.	
Generación de números aleatorios por computadora	636
Ing. Rubén D. Maza	
No es tarea tan simple como parece obtener números al azar que tengan buenas propiedades estadísticas.	
Apuntes de proyectistas: Generador de pulsos	641
Ing. J. W. Zaninetti y C. A. Marqués	
Notas prácticas sobre DX y RF	643
Temas de radio y teledifusión: Televisión en la banda de UHF: Prejuicios que ya no son	644
M.O.S.	
Consejo Profesional de Ingeniería de Telecomunicaciones, Electrónica y Computación	688
Instituto de Ingenieros en Electricidad y Electrónica	689
Centro Argentino de Televisión	690
Asociación de la Pequeña y Mediana Industria Electrónica	691
Hojeando revistas	692
Nuevos productos	693
Noticias	695
Indice de anunciantes	702

H. L. Rufiner, D. Zapata & A. Sigura: "Interfaz oral para computadoras personales orientada al discurso continuo" *Telegráfica Electrónica*, No. 938, pp. 602-609, 1993.



Revista Telegráfica - Electrónica (ISSN 0035-5016) es una publicación mensual de Arbó S.A.C. e I. sobre Electrónica, Telecomunicaciones e Informática. La más antigua y prestigiosa escrita en castellano. Impresa en Argentina. Registro Nacional de la Propiedad Intelectual N° 209.082. Miembro de la Asociación de la Prensa Técnica Argentina (APTA) y de la Asociación Argentina de Editores de Revista (AAER). Queda prohibida la reproducción total o parcial de cualquier artículo sin la autorización previa y por escrito de los editores. La responsabilidad de los artículos corresponde únicamente a los autores y la de los avisos a los anunciantes. Precio: \$ 4.-. Precio de la suscripción anual en el exterior: 45 dólares (USA). Impresión Benavent Hnos. S.A.I.C., Mompox 1654 (1134) Buenos Aires, Argentina. Distribuidor en Capital Federal: Brihet e Hijos, Viamonte 1465, 7° Piso, Dpto. "B" (1055) Buenos Aires. Distribuidor en el interior y exterior: DIPU S.A., Azara 225/235 (1267) Buenos Aires, República Argentina.

Interfaz oral para computadoras personales orientada al discurso continuo

Se presenta un sistema de reconocimiento automático de la voz humana, utilizando redes neuronales o inteligencia artificial

Aldo Sigura*, Daniel Zapata* y Leonardo Rufiner*

Resumen

Se presenta el desarrollo de un sistema de reconocimiento automático del habla, capaz de traducir a texto escrito el discurso continuo efectuado en idioma español.

El sistema puede ser dividido en dos grandes bloques: un módulo basado en redes neuronales para el reconocimiento de fonemas y traducción a un código de símbolos pseudo-fonéticos; y un segundo módulo desarrollado en Prolog, capaz de discriminar las palabras contenidas en las secuencias de símbolos que envía el primer subsistema.

El primer bloque acondiciona la señal de habla y la lleva al plano espectral. Luego, una etapa de computación neuronal procesa la secuencia de espectros para devolver una cadena de símbolos que corresponden a la secuencia de fonemas emitidos. Se utilizan perceptrones multicapa para reconocer aquellos fonemas cuya espectro es temporalmente estable, y otras arquitecturas neuronales (Time-Delay Neural Networks) para manejar los aspectos espectrales dinámicos de la señal de voz.

El segundo módulo consiste en un "parser" desarrollado mediante DCG's (Definite Clause Grammars) en Prolog cuyo objetivo es descifrar las palabras y frases que se hallan contenidas en las cadenas de símbolos pseudo fonéticos que son emitidas por el primer módulo. El alcance del análisis sintáctico está limitado sólo a la cadena considerada, haciendo posible que también puedan reconocerse palabras aisladas.

Introducción

Uno de los aspectos que mayor importancia ha cobrado en el desarrollo de equipos y programas de computación es el relacionado con la comunicación con el usuario. Los periféricos de entrada disponibles hasta el presente —si bien se ha avanzado mucho en la materia— implican ciertos condicionamientos a los que se debe adaptar el usuario cuando quiere interactuar con la tecnología de hoy, ya que debe utilizar modalidades de ingreso de información que distan mucho de brindar la sencillez y flexibilidad propias de la comunicación verbal directa entre las personas.

El hecho de eliminar las barreras entre la gente y sus máquinas, acerca el poder de las computadoras a muchas más personas y ayuda a que aquellas que las utilizan cotidianamente incrementen su eficiencia.

Aparejadas a la facilidad de su utilización, el habla ofrece muchas otras ventajas para la comunicación con la computadora. Una es la velocidad: la mayor parte de la gente puede

pronunciar más de 200 palabras por minuto, mientras que muy pocas pueden escribir más de 60 palabras en el mismo lapso. Por otra parte, el reconocimiento automático del habla puede permitirle a los usuarios utilizar una computadora en lugares o situaciones en los que sería imposible o peligroso hacerlo del modo convencional.

Aunque el habla no es el modo de entrada óptimo y definitivo, ofrece —sin embargo— ventajas únicas no disponibles en otro tipo de interfaces.

La dificultad para resolver los problemas asociados al reconocimiento del habla mediante técnicas de procesamiento convencionales, está dada por la complejidad de las señales implicadas [1, 11, 13, 16], ya que las mismas presentan funciones estadísticas de densidad superpuestas, tienen formas complicadas en espacios de varias dimensiones y/o son no estacionarias. El castellano —y en particular la variante hablada en la Argentina— presenta ventajas comparativas frente a los idiomas anglosajones, y aun otras lenguas latinas, a los efectos de ser reconocido por una máquina [13, 14]. Su menor número de fonemas básicos, la mayor separabilidad de sus características acústicas y la correspondencia entre las regiones de decisión determinadas por estas últimas con los respectivos fonemas, facilitan la tarea. Sin embargo, el problema más grave está centrado en aceptar las señales provenientes de cualquier hablante, con independencia del tono de voz, velocidad de pronunciación y discontinuidad en el modo de hablar.

Existe virtualmente una gran cantidad de aplicaciones del sistema que se presenta: como periférico para controlar aplicaciones convencionales de software (v.g.: procesadores de texto), en sistemas para rehabilitación fonoaudiológica, como interfaz de control de instrumental, etc. Pero lo más interesante de este sistema es que puede representar la apertura de un gran espectro de posibilidades para las personas que padecen impedimentos motores para operar las interfaces convencionales de los productos tecnológicos actuales. El acceso a la utilización de máquinas, herramientas y software a través de la voz permitirá a los discapacitados desarrollar sus potencialidades en el campo laboral, estudiantil y recreativo.

Aspectos generales

Aunque las técnicas utilizadas por los sistemas de reconocimiento automático del habla han mejorado notablemente su desempeño en los últimos años, éstas distan mucho de ser adecuadas para algunas aplicaciones. La mayor parte de los sistemas se basan en la aplicación de modelos de Markov subyacentes (HMM), los que han sido útiles para tratar los aspectos secuenciales de la señal de habla [1, 6], pero han sido deficientes como clasificadores de fonemas y —si se pretende

* Facultad de Ingeniería - Bioingeniería - Universidad Nacional de Entre Ríos.

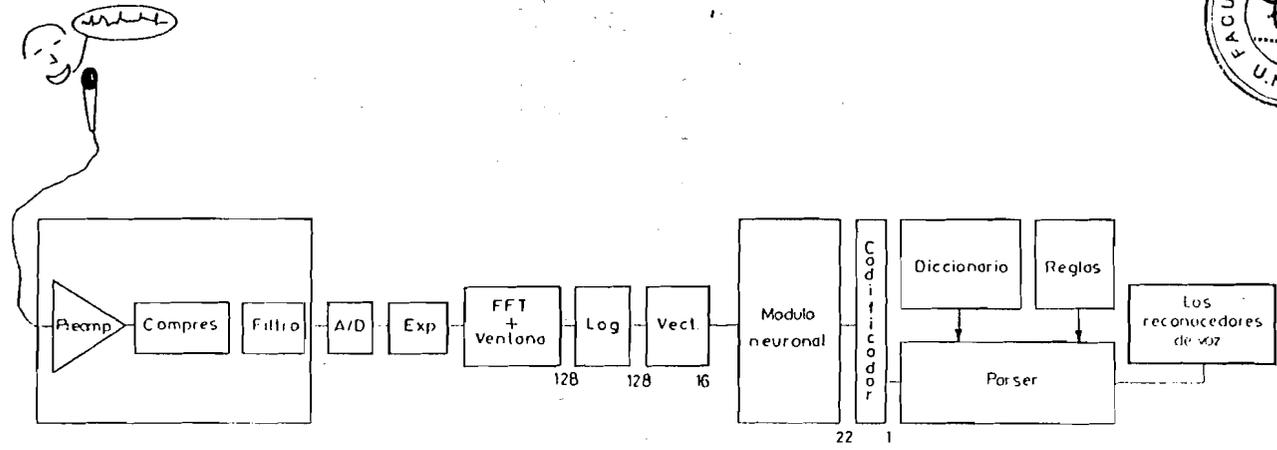


Fig. 1. Esquema básico del sistema de reconocimiento automático del habla.

que el cómputo tenga eficiencia— deben tratar algunos eventos como independientes, aunque en realidad no lo sean.

El sistema utilizado se divide en dos grandes etapas basadas en estrategias diferentes: Una está encargada del reconocimiento de los fonemas emitidos por un hablante cualquiera, y su traducción a un código de símbolos fonéticos (módulo M1, Red Neuronal). La otra etapa (M2, Parser) es un sistema inteligente capaz de extraer las palabras que están “escondidas” en las secuencias de símbolos que envía el primer subsistema.

Estos dos módulos son complementarios e independientes desde el punto de vista funcional; esto permitió el desarrollo de cada uno por separado, para lo cual se debieron tener presentes las características y restricciones del código fonético intermedio, que los liga.

Una representación esquemática del sistema completo puede verse en la Fig. 1.

Preprocesamiento acústico

La investigación fisiológica de la audición ha revelado una serie de detalles que pueden ser tenidos en cuenta para aplicarlos al reconocimiento automático del habla. Entre ellos se halla el preprocesamiento que se lleva a cabo a nivel coclear, el cual consiste fundamentalmente en un análisis frecuencial basado en las resonancias de la membrana basilar del oído interno. La descomposición espectral de la señal es transmitida al cerebro a través de los nervios auditivos [4]. Para reproducir ese preproceso se decidió aplicar técnicas de análisis frecuencial convencionales, ya que el análisis de Fourier digital es preciso y rápido (FFT) [12] y el empleo de una red que realizara tal análisis hubiera incrementado el tiempo de computación y la complejidad del sistema [1].

Básicamente, el preprocesador acústico de este sistema está compuesto por las siguientes etapas:

- Micrófono.
- Preamplificador con filtro pasabajo con frecuencia de corte en 4,5 kHz.
- Compresor de rango dinámico.
- conversor A-D de 8 bits con 10 kHz de frecuencia de muestreo.
- Expansor digital de rango dinámico.
- Transformada rápida de Fourier de 256 puntos calculada cada 10 msec utilizando una ventana de Hamming de 256 puntos.
- Logaritmación de potencias espectrales.

- Vectorización de canales espectrales en un arreglo de 16 componentes reales en escala logarítmica.

Cualquier sistema de tratamiento de sonido puede modificar el habla durante el proceso. Por lo tanto, deben encontrarse los componentes adecuados para cada una de las etapas, a fin de que el sistema específico no modifique las características del hablante.

La elección del micrófono juega, desde este punto de vista, un papel importante en el diseño del equipo. Durante el desarrollo del proyecto se llevaron a cabo diversas pruebas con micrófonos de tipo electrodinámico, capacitivo y electret.

Luego de analizar las tres clases de micrófonos, se optó por uno de tipo electret, con similar fidelidad a la del capacitivo, debido a su semejanza de principios y a la simplicidad de la electrónica empleada en la amplificación de la señal obtenida. Además posee una gran sensibilidad debido al TEC incorporado en la cápsula (-50 dB referidos a 1 V/dina/cm²) y su margen dinámico es 120 dB con una distorsión inferior al 3%, con una respuesta casi plana de 50 a 10 kHz. Es un micrófono insensible a la humedad, y el uso masivo de este tipo de micrófonos ha disminuido drásticamente sus costos. Todas las cápsulas ensayadas eran omnidireccionales. Sin embargo, la inclusión de la cápsula en un conjunto del tipo “headset” con auriculares proporciona una respuesta del tipo cardioide. Esta disposición, además, elimina la variación de la distancia a la fuente de sonido durante la operación del equipo y libera las manos del usuario. Para ambientes muy ruidosos, podría utilizarse una cápsula acústica, supresora de ruido.

En el circuito (Fig. 4) la cápsula X_1 , está conectada a través de R_1 a la alimentación de 5V. Para evitar problemas de variación del nivel medio se utilizó el regulador U_1 . R_1 se eligió de manera de conseguir máxima sensibilidad. El capacitor C_1 actúa como desacople de la corriente continua de alimentación.

El preamplificador debe poseer una respuesta plana en la banda de frecuencias considerada y proveer adecuada adaptación de impedancias a la cápsula microfónica (~100kΩ), por lo que se utilizan amplificadores operacionales con entrada TECJ tipo TL081 que superan los requerimientos necesarios.

La amplificación a los niveles deseados (5V) se consiguió con dos etapas inversoras (40dB) para distribuir la ganancia y evitar alinealidades o inestabilidad.

La primera etapa es una configuración simple de sumador inversor y está constituida por $U_2, R_2, R_3, R_4, R_5, R_6, R_7$. De esta manera se adiciona una componente de continua necesaria para el ajuste de nivel al valor medio del conversor

sinc(f) Research Center for Signals, Systems and Computational Intelligence (rich.unl.edu.ar/sinc)
 H. L. Rufiner, D. Zapata & A. Sigara. "Interfaz oral para computadores personales orientada al discurso continuo"
 Telegrafía Electrónica, No. 938, pp. 602-609, 1993.

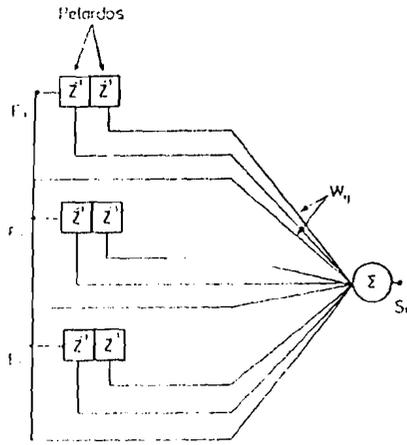


Fig. 2 — Neurona con 3 entradas y 2 retardos c/u. (9 entradas comunes).

A/D mediante el potenciómetro R_3 filtrado por C_4 . Como puede verse hay una ganancia de unos -12dB para el nivel de tensión y una de 20dB para la entrada de sonido. El potenciómetro R_2 permite una atenuación variable de dicha entrada.

La segunda etapa está constituida por R_8, R_9, R_{10} y el U_3 , con una ganancia de 20dB . El prototipo se muestra en la Fig. 6.

Un aspecto vital para mantener cualquier señal electrónicamente amplificada tan parecida a la original como sea posible es la conservación de una apropiada relación de nivel. Si el nivel del hablante baja mucho, la señal se pierde en el ruido de fondo. Si ésta es muy alta, la señal puede causar distintos tipos de distorsiones en los componentes del sistema, y puede ser dificultoso reconocerla.*

La señal de voz, para su digitalización, debe estar ubicada en una ventana de rango dinámico específica; aun cuando aquella provenga de hablantes de voz baja muy alejados del micrófono, o de aquellos que gritan encima del mismo.

Otro problema es que para no enmascarar bajo el ruido de cuantificación las señales débiles (especialmente las fricativas cuya diferencia de energía con el resto de los fonemas es muy grande), se debe aprovechar toda la ventana de amplitud provista por el conversor A/D del sistema.

Todo esto obliga a que un sistema práctico de reconocimiento del habla independiente del hablante, utilice algún tipo de control automático de ganancia CAG de manera de normalizar las distintas señales en cuanto a su amplitud.

Una de las razones que dificulta el procesamiento correcto del habla es que determinados sonidos, típicamente consonantes, tienen una relación pico/valor medio extremadamente alta. Dado que las consonantes son críticas para la inteligibilidad, el sistema debe ser capaz de mantener la integridad de estos transitorios y sus relaciones dinámicas. En la mayoría de los patrones de voz normalizados estos transitorios son seguidos por vocales de baja energía o silencio. Por consiguiente, el procesamiento debe ser capaz de reaccionar a esta nueva condición de entrada sin que las vocales sean "enterradas", o el silencio sea "inhalado".

Otra dificultad en el procesamiento del habla es que hay a menudo largos momentos de silencio entre frases y oraciones, así como cortos momentos entre sílabas y palabras. Cualquier

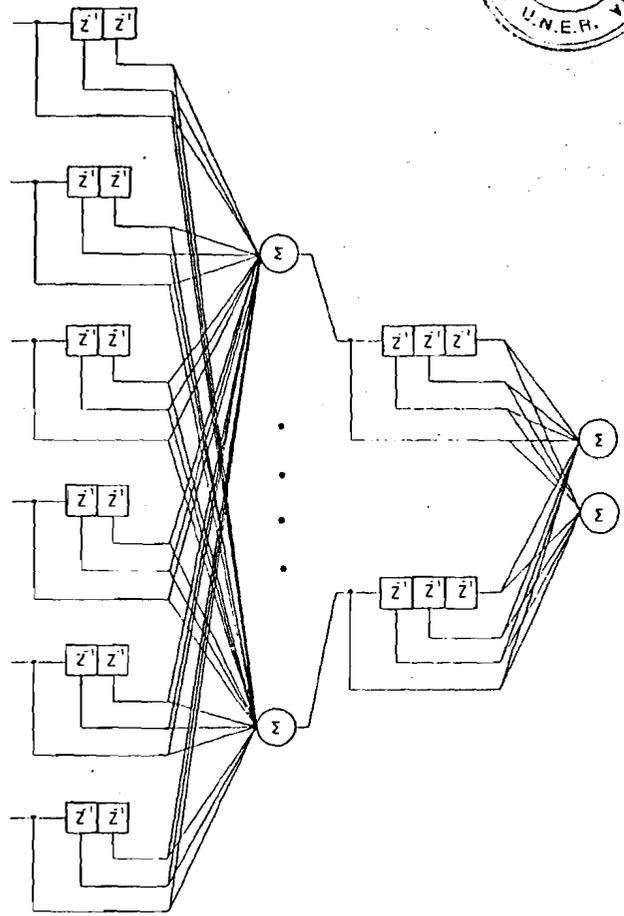


Fig. 3 — Estructura básica de una TDNN.

CAG con bajo umbral debe tener un mecanismo para congelar el aumento de ganancia durante estos silencios de manera que el ruido de fondo/sistema no aumente.

En el prototipo no se utilizó un CAG, pero se mantuvieron los niveles de entrada y se entrenó a la red neuronal con pequeñas variaciones alrededor de dichos niveles, lo que con el CAG no hubiese sido necesario.

La frecuencia de muestreo del conversor es de 10 kHz , por lo que se debe limitar el ancho de banda de la señal de entrada a aproximadamente $4,5\text{ kHz}$. A los efectos de eliminar los problemas de alias se utilizó un filtro Butterworth de 4° orden con frecuencia de corte de $4,5\text{ kHz}$. El filtro es del tipo activo con estructura Rauch y logra la transferencia antedicha con sólo dos componentes activos U_4 y U_5 , además de $R_{11}:R_{16}, C_5:C_8$. La frecuencia de corte y los valores de los componentes se calculan mediante las fórmulas recursivas clásicas.

A la salida del filtro y antes de la entrada al conversor se provee una protección para evitar tensiones excesivas sobre el A/D formada por R_{17} y el zener D_1 de $4,7\text{ V}$.

La placa A/D utilizada fue diseñada para propósitos generales (Fig. 5), y está basada en el ADC0808, con una frecuencia de muestreo de 10 kHz —programable por soft—. El mismo está almacenado como I/O y leído por "pooling" verificando una bandera de EOF y leyendo el dato convertido desde la dirección I/O 30011. Las rutinas de conversión están realizadas en MASM e incorporadas a un programa en Turbo Pascal 6.0 a través de sentencias InLine (). Este conversor de 8 bits posee un multiplexador analógico de 8 canales, de los cuales se utiliza uno para la señal de voz. La placa permite también el trabajo por interrupciones, lo que dejaría tiempo

* El problema no reside tanto en los amplificadores sino en la producción del fonema: el hablante cambia las características del generador y la forma de los resonadores cuando cambia el nivel.



libre durante la conversión para otras tareas. En la misma placa del A/D se provee de un D/A que permite reproducir señales de voz grabadas en el entrenamiento a través del "headset" o utilizarla para síntesis de voz en una interfaz oral completa, para lo cual se incluye en la placa analógica un amplificador de salida ($U_6 + R_{19}; R_{24}$) con el correspondiente filtro ($R_{18}; C_9$) de primer orden, eliminando la componente continua mediante la configuración como sumador inversor, para excitar los auriculares y una salida de potencia con un TDA 2002 para excitar un parlante de 4W.

Una vez que la señal está digitalizada se procede a realizar su análisis STFFT (Short Time FFT) mediante una ventana de Hamming de 256 puntos c/10 ms. Una alternativa a este planteo consiste en utilizar un análisis del tipo Short time LPC, pero la idea de reemplazar la FFT por un banco de filtros en la versión definitiva —lo que evita el costoso tiempo de cálculo de la FFT— decidió la primera opción.

Luego se agrupan los 128 coeficientes obtenidos en 16 bandas de frecuencias en escala de Mel (logarítmica) un poco más "fisiológica". Se toma entonces el logaritmo de las potencias espectrales, dado que permite remarcar aún más las ca-

racterísticas distintivas al resaltar los fonemas de menor amplitud. Se prevé el uso de otra componente en el vector o "pattern", que refleje la energía total en la ventana temporal.

De esta manera se completa el proceso para llegar al "pattern" (el "pattern" completo en realidad consiste en tres espectros sucesivos que conforman un total de 48 entradas) y que es mostrado a la red para su clasificación según el entrenamiento previo.

Módulo neuronal

La aparición de técnicas de entrenamiento eficaces para redes neuronales —en particular las redes anteroalimentadas— permitió la aplicación de las mismas al procesamiento del habla, aunque hasta hace muy poco tiempo estuvieron orientadas a patrones estacionarios. Para evitar este escollo se diseñaron redes neuronales que —además de los patrones estáticos— incorporarán simultáneamente información que fuera generada en diferentes instantes. La utilización de estas redes con retardos (TDNN) permite descubrir características acústico-fonéticas y sus relaciones a lo largo del tiempo sin

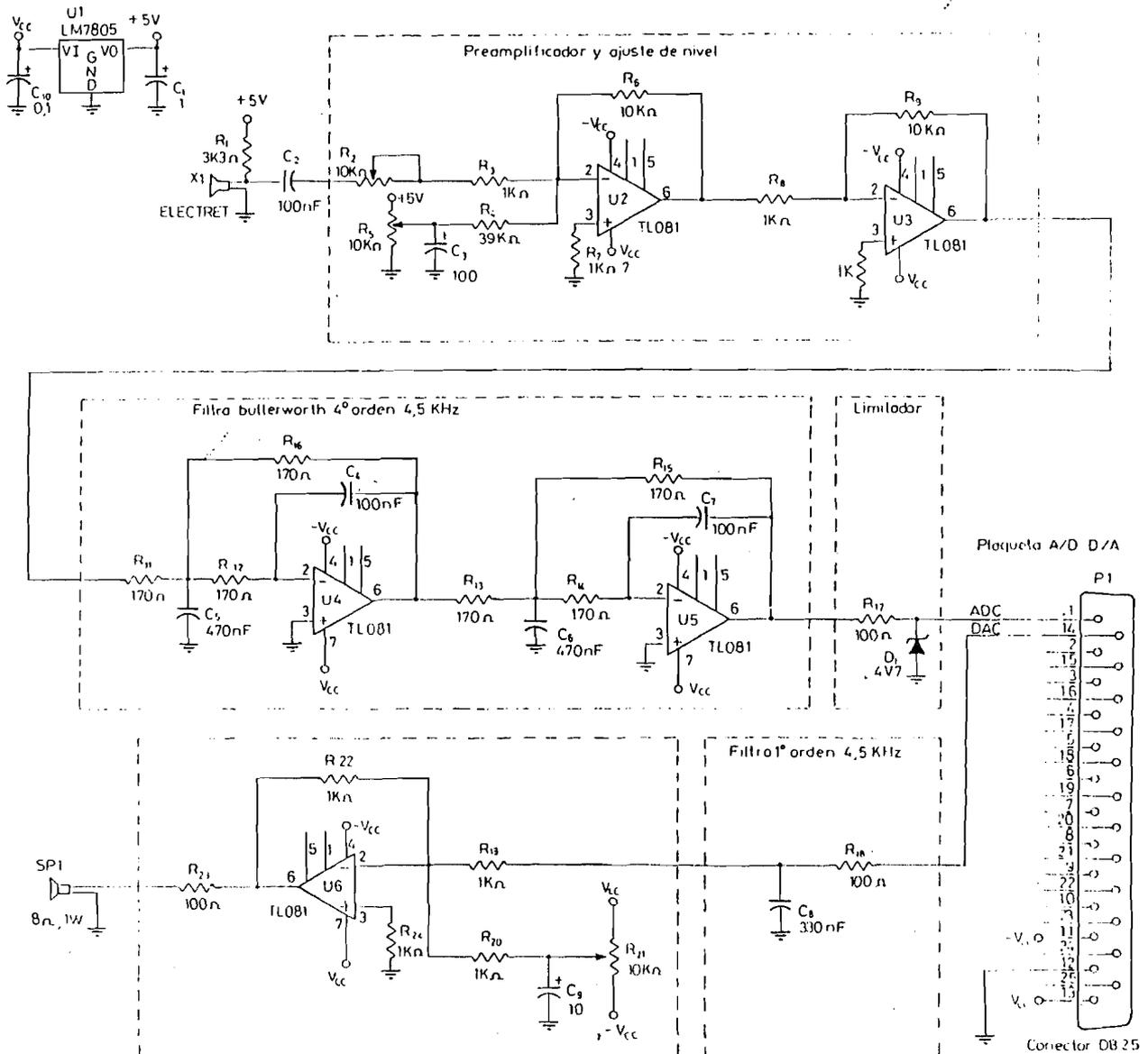
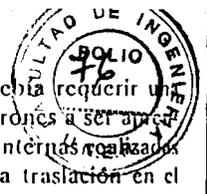


Fig 4 — Circuito preamplificador.

sine(0) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar) / Rufiner, D. Zapata & A. Sigurto, "Interfaz oral para computadores personales orientado a curso continuo" / SIGURTO, A. / 2006 / pp.062-079 / 179



verse afectadas por las traslaciones del patrón de entrada [1]. Una red neuronal que no es invariante a la traslación requiere una segmentación precisa para alinear el patrón de entrada en forma adecuada. Esta invariancia al desplazamiento ha sido considerada un factor determinante en algunos modelos neuronales que han sido propuestos para reconocimiento de secuencias fonéticas.

Aunque el objetivo de este trabajo no es utilizar necesariamente modelos neuronales que tengan un correlato biológico, se han tenido en cuenta algunos aspectos estructurales y fisiológicos del sistema auditivo en sus distintos niveles que se consideran significativos en cuanto a su influencia en el proceso de reconocimiento del habla.

Además, se usan técnicas de procesamiento digital de señales cuando han mostrado ser superiores a los circuitos basados en modelos biológicos, como en el caso del preprocesamiento. Para las tareas en las cuales se requiere la realización de una gran cantidad de análisis estadísticos, como en la detección óptima de fonemas, se utilizan redes neuronales. El entrenamiento de las redes se llevó a cabo con métodos que han demostrado ser eficaces para alcanzar una precisión en el reconocimiento cercano al máximo teórico de decisión y que, simultáneamente, no conllevan una implementación computacional compleja.

La serie de pasos seguida en el desarrollo del trabajo comenzó con la separación y caracterización de los fonemas, para lo cual se recogió información fonética grabando las emisiones de diez hablantes —5 varones y 5 mujeres— y siguiendo algunas pautas de selección de las secuencias.

A partir del estudio realizado se decidió tomar los siguientes agrupamientos, que se consideraron adecuados para ser tratados mediante clasificadores neuronales, a fin de asignar una red diferente para cada grupo. Esta estrategia posibilita un ajuste en la discriminación de los fonemas en base a diferencias finas en las secuencias espectrales:

Vocales	Oclusivas sonoras
Fricativas	Oclusivas sordas
Laterales	Vibrantes
Africadas	Nasales

A partir de estos grupos se definieron las características generales de la red neuronal a diseñar. Esta debía ser una red anteroalimentada de varias capas donde una de ellas estuviera formada — a su vez— por subredes, cada una de las cuales se encargara de discriminar los fonemas dentro de un grupo específico. Como modelo de referencia se tomaron los perceptrones multicapa, dado que han demostrado ser eficaces como clasificadores [2, 3, 17]. Los criterios seguidos para el diseño de la red neuronal con respecto a su estructura y propiedades fueron los siguientes:

- La red debía poseer una cantidad suficiente de capas y de interconexiones entre las unidades correspondientes a cada una de esas capas. Esto permitió que la misma discriminara regiones de decisión no lineales complejas.
- El número de pesos ajustables de la red debía ser suficientemente pequeño comparado con la cantidad de datos de entrenamiento, de tal modo que la red estuviera obligada a extraer regularidades modificando los pesos de las interconexiones.
- La red debía poder representar internamente las relaciones entre eventos no simultáneos, aunque próximos en el tiempo. Esos eventos fueron coeficientes espectrales y también las salidas correspondientes a detectores de características de más alto nivel.

- La estrategia de entrenamiento no debía requerir una alineación temporal precisa de los patrones a ser aprendidos. Para ello, las representaciones internas realizadas por la red, debían ser invariantes a la traslación en el tiempo.

Estructura de la redes neuronales utilizadas

Otro de los datos importantes que se tuvieron en cuenta fue la arquitectura neuronal de la corteza auditiva. La corteza está formada por varias capas de células nerviosas, cada una de las cuales está constituida por tipos específicos de neuronas. Las capas corticales superiores fueron desarrolladas en las últimas etapas evolutivas del cerebro de los mamíferos y, en el caso del hombre, poseen una gran proporción de la totalidad de las neuronas [7, 10]. La actividad neuronal sigue, en general, un patrón vertical que da lugar a la formación de columnas que a su vez están relacionadas lateralmente entre sí. Dentro de cada columna, una neurona perteneciente a una capa hace sinapsis directas sobre neuronas de la siguiente capa, o bien indirectamente, a través de interneuronas [9]. Esto da lugar —teniendo en cuenta los retardos sinápticos— a que una neurona cualquiera de las capas más altas reciba simultáneamente información que fue generada en instantes distintos en la periferia.

Las unidades elementales de cómputo utilizadas en muchas arquitecturas neuronales realizan una suma ponderada de sus entradas, para luego aplicarle una función no lineal. Las propiedades matemáticas de esta función —monótona estrictamente creciente, continuamente diferenciable y con dos asíntotas horizontales que la acotan inferior y superiormente— la hacen muy adecuada para esta tarea, ya que permite que las regiones de decisión poscan contornos suaves.

En la arquitectura utilizada en este trabajo, estas unidades elementales fueron modificadas —utilizando los conceptos anteriores a fin de que pudieran procesar información generada en distintos instantes (Figs. 2 y 3). A las M entradas sin retardos de cada neurona, se les agregaron las $M \times N$ entradas correspondientes a N instantes distintos. La cantidad de pesos que influyen directamente sobre cada neurona es, entonces, $M \times (N + 1)$. Para $N = 2$ y $M = 16$, por ejemplo, serán necesarios 48 pesos para calcular la suma ponderada de las 16 entradas, donde cada una de ellas estará tomada en tres instantes diferentes. De este modo, una unidad neuronal de estas características es capaz de relacionar y procesar conjuntamente la entrada actual con eventos anteriores. Las subredes encargadas de discriminar los fonemas dentro de cada grupo, poseen tres capas de neuronas con sus correspondientes retardos. Ese número de capas garantiza que la red pueda determinar regiones de decisión de forma arbitraria pertenecientes a espacios $[(N + 1) \times M]$ dimensionales. Las regiones de decisión no lineales complejas, son necesarias para llevar a cabo una clasificación adecuada, teniendo en cuenta la variabilidad de las señales acústicas.

Cada unidad neuronal con retardos posee la capacidad de extraer relaciones temporales del patrón de frecuencias dinámico de la entrada. Las transiciones locales de corta duración son tratadas por las capas más bajas, mientras que las capas más altas relacionan información que involucra a períodos más largos. En los detectores-discriminadores de fonemas, las entradas a cada neurona son los coeficientes espectrales actuales y los que corresponden a instantes anteriores, aunque no necesariamente sucesivos, ya que —según sea el grupo fonético que se considere— se pueden excluir muestras intermedias redundantes a fin de ingresar conjuntamente muestras significativas más separadas sin aumentar el costo computa-

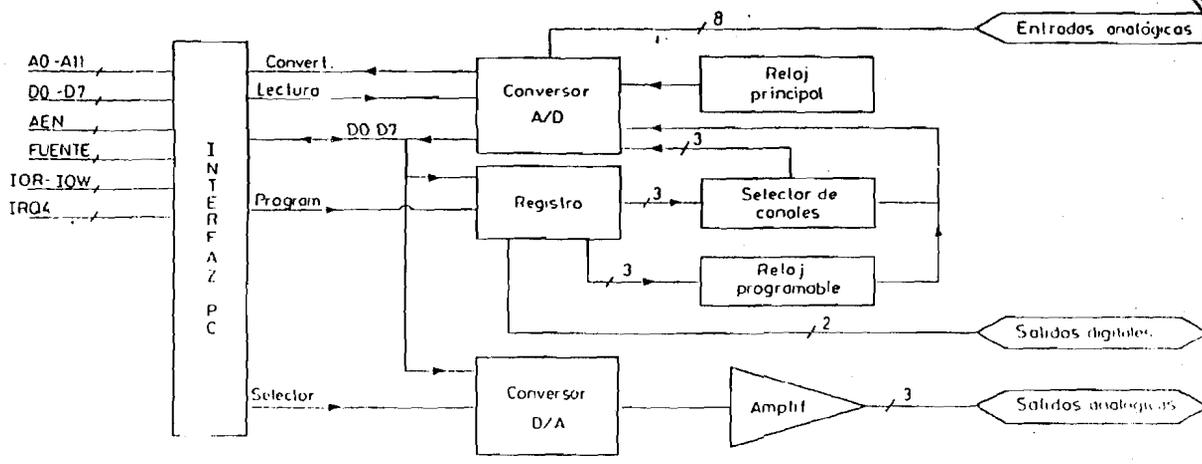


Fig. 5 — Diagrama en bloques del conversor AD/DA

cional. Más aún, el número de entradas por neurona puede variar para distintas subredes.

Entrenamiento de las redes

Las subredes se entrenaron individualmente a fin de que cada una discriminase las características fonéticas determinadas a priori. Las subredes son, por lo tanto, detectores-identificadores de vocales, fricativas, nasales, etc.

Existen varias técnicas para el entrenamiento de las redes neuronales anteroalimentadas. En nuestro caso se adoptó el algoritmo de retropropagación [2, 4, 15] (backpropagation) el cual es una generalización del algoritmo de cuadrados medios mínimos (LMS). Este usa una técnica de búsqueda por gradiente para minimizar una función costo igual a la diferencia cuadrática media entre la salida deseada y la salida real. Cada subred se entrenó asignando inicialmente valores aleatorios pequeños a los pesos de interconexión y a los umbrales internos de las neuronas, y luego se presentaron todos los datos de entrenamiento en forma repetida. Esta información consistió en las secuencias de patrones espectrales. Los pesos se ajustaron después de cada intento brindando la información que especificara la clase correcta, hasta que los pesos convergieron o su variación con el número de entrenamiento tomó un valor aceptablemente pequeño, alcanzando el comportamiento adecuado.

Una vez logrado el comportamiento deseado de cada subred se procedió a su interconexión configurándose de este modo la red global. Luego se continuó el entrenamiento en forma análoga a los anteriores, sólo que el coeficiente de ajuste de los pesos de interconexión que se tomó fue más pequeño que el utilizado con las subredes. Se eligió esta estrategia a fin de que la red total fuera capaz de descartar los "falsos positivos" alcanzados por una o más subredes sin que el nuevo entrenamiento afectara sustancialmente lo aprendido por cada una de ellas.

Software de entrenamiento

La realización del entrenamiento de la red neuronal se lleva a cabo en base a los datos de frecuencia-tiempo etiquetados con el programa anterior. La red utilizada es del tipo perceptrón TDNN y el entrenamiento se realiza por el algoritmo de "backpropagation" con pequeñas modificaciones para que trabaje de acuerdo con la filosofía de estas redes. El programa permite variar la estrategia de entrenamiento para leer los datos del archivo en forma secuencial o aleatoria, corrigiéndolos con cada entrada, cada cierto número de

ellas (para que la red no intente responder a patrones espurios o incompletos), o según una promediación de los valores correspondientes a un número fijo de espectros sucesivos (token). Este último método ha sido aplicado con buenos resultados en algunos trabajos [4].

El software permite un entrenamiento supervisado con intervención del usuario a fin de poder variar el coeficiente de aprendizaje, el cual es variable por capas, o el de afectación del número según evolucione el comportamiento de la red. Esta técnica permite el "congelamiento" de una capa cuando parece haber extraído las características necesarias de manera de aumentar la eficacia del algoritmo.

En la pantalla se proporciona información permanente del error cuadrático medio sobre las salidas con respecto a la clasificación deseada. Se puede observar la evolución de los pesos y umbrales de la red en forma gráfica, representados según una escala de colores. Esta representación permite apreciar la dinámica de la red y decidir sobre el cambio de éstas, número de capas, neuronas por capa, retardos, etc. Por ejemplo si el coeficiente de ajuste es muy grande, la red oscila y tiende a aprender constantemente el último "pattern" mostrado. Esta situación puede apreciarse sin necesidad de que la red termine su entrenamiento, ya que se manifiesta como un "calentamiento" sucesivo (en color rojo) de los nodos de salida correspondientes a los fonemas mostrados en ese momento de la entrada, mientras se "enfrian" (en azul) los nodos restantes.

Durante la fase de entrenamiento, se puede efectuar una alteración porcentual aleatoria de los valores de los pesos y umbrales de la red de manera de escapar de algún mínimo local (simulated annealing). Otra forma de visualización de la red utilizada muestra las neuronas según colores que representan su actividad o nivel de salida y los pesos de acuerdo al producto de la salida afectada por el valor de dicho peso. Esta representación permite "seguir" el camino de la señal de entrada y evaluar si la activación se corresponde con los "caminos" favorecidos por cada "pattern".

La finalización del entrenamiento tiene lugar cuando el ECM de las salidas es menor que un valor prefijado. El tiempo que insume un ciclo de entrenamiento completo es del orden de 16 horas en una AT 486 a 33MHz.

Implementación del módulo conversor fonemas-texto

El reconocedor de voz utiliza los modelos de redes para clasificar las señales de voz en tiempo real, basándose en los archivos de pesos correspondientes a dichas redes ya entrenadas. Para ello se mide la energía de una ventana pequeña

(c) Research Center for Signals, Systems, and Computational Intelligence (fich.unl.edu.ar)
 L. Kullmer, D. Zapata & A. Sigara: "Interfaz para computadoras personales orientada a discurso continuo"
 Telegráfica Electrónica, No. 938, pp. 602-609, 1993.

sobre la señal (30 ó 40 ms) de manera de detectar los silencios en la señal y utilizar ese tiempo muerto para calcular los espectros y la salida de la red. La utilización de interrupciones permitiría aumentar la eficiencia del proceso.

Se emplea una técnica de optimización del tiempo de cálculo utilizando un "buffer" circular como registro de desplazamiento de manera de computar la energía instantánea en la ventana con sólo una resta y una suma, lo que evita la sumatoria sobre toda la ventana a cada instante. Mientras se detecta energía por encima de un umbral se llena un "buffer" con la señal a tratar hasta que se detecta un silencio y entonces comienza el proceso de conversión. La información de salida se presenta en pantalla como cadenas separadas por retornos de carro en correspondencia con las pausas, y se graba en un archivo para su tratamiento por el módulo de análisis sintáctico.

El programa posee varios mensajes grabados en forma digital tomados de emisiones reales de voz, que se presentan al usuario a fin de hacer más "amistosa" la interfaz del sistema.

Módulo sintáctico

El módulo M2 de este sistema está basado en una serie de bloques realizados en lenguaje Prolog y puede describirse como un sistema inteligente orientado a detectar las palabras y frases en español que están contenidas en cadenas de símbolos pseudo-fonéticos. Estas cadenas son generadas por el sistema neuronal M1 y corresponden a las secuencias de fonemas emitidas por el hablante que han sido reconocidas. Este módulo no está orientado al tratamiento del lenguaje natural en su sentido más amplio, sino que toma alguna de las herramientas del campo para resolver el problema de la segmentación de palabras asociado al discurso continuo, mediante un análisis sintáctico local.

El módulo M2 está dividido en tres partes: la base de conocimientos, la base de datos y un programa integrador que gestiona la consulta a los anteriores y se encarga de establecer la interfaz con el usuario.

La base de conocimientos está formada por un conjunto de reglas que describen las estructuras gramaticales válidas. Estas reglas fueron implementadas utilizando el formalismo DCG y constituyen la médula del sistema. El alcance del análisis sintáctico que se lleva a cabo está limitado a una sola cadena continua, acotada por las necesarias pausas respiratorias del hablante, lo que hace posible también el reconocimiento de palabras aisladas. Asimismo cada cadena no debe corresponder necesariamente a una frase completa, sino que también se comprueba la validez de secuencias de palabras que no constituyan una oración.

La base de datos está constituida por predicados que relacionan las secuencias de símbolos de entrada con las correspondientes secuencias de caracteres del alfabeto español; y además, estos predicados incluyen la función gramatical, género y número de la palabra en cuestión.

La versión del lenguaje de programación que se utilizó fue Arity Prolog V. 5.00. Se eligió esta implementación Prolog —que constituye una versión muy ampliada del lenguaje Prolog básico— ya que incluye potentes predicados predefinidos para el manejo de texto, gestión de base de datos, facilidades para el desarrollo de la interfaz con el usuario, etc. [20, 21]. Además posibilitó trabajar en modo intérprete durante la etapa de desarrollo, para luego compilar las aplicaciones y enlazarlas con módulos desarrollados en otros lenguajes (Assembler, Pascal, etc.). Pero —fundamentalmente se eligió



Fig. 6 — Foto del prototipo.

esta versión de Prolog porque permitía usar notación DC de suma utilidad para este trabajo.

La tarea que desempeña este módulo consiste en la conversión de un archivo de texto de entrada con información representada con símbolos fonéticos, en otro archivo de salida con la información correspondiente, pero representada según símbolos y sintaxis del idioma castellano. El archivo de texto de entrada tiene una existencia solamente formal si el proceso de reconocimiento se efectúa en tiempo real; o una existencia física concreta si el proceso se realiza en tiempo diferido. La primera versión del sistema funciona según esta última modalidad.

Uno de los primeros pasos fue delimitar las características de los datos de entrada en cuanto a la longitud de las cadenas fonéticas y tipo de codificación a utilizar para los fonemas. También se tuvo en cuenta la posible existencia de fonemas compartidos por dos palabras sucesivas (ej.: la "n" en "bunivel") y fonemas redundantes que pudieran provenir de la primera etapa.

Las reglas gramaticales codificadas según el formalismo DCG, constituyeron el bloque sobre el cual se trabajó más intensamente. Previamente se consideró la posibilidad de implementar las reglas para el tratamiento de lenguaje natural [18, 19, 22], mediante RTN, ATN, etc.; pero se optó por utilizar DCG debido a la simplicidad y claridad de su sintaxis, y al hecho de que la versión de Prolog usada permite su utilización en forma directa.

La información de entrada consiste en secuencias de caracteres ASCII provenientes del primer módulo de reconocimiento de fonemas. Este subsistema envía la información según las siguientes pautas y restricciones:

- El número de palabras contenidas en una cadena continua de símbolos fonéticos es entero, ya que las pausas respiratorias o de puntuación dan origen a silencios, que son traducidos como señales de fin de cadena; tales silencios sólo tienen lugar cuando finaliza una palabra completa o antes del comienzo de otra.
- Los fonemas son representados por la letra del alfabeto

español que más se le ajuste; en su defecto se utiliza un carácter ASCII específico para representarlo.

- Los fonemas o secuencias de fonemas que den lugar a ambigüedades en su representación escrita son representados por una única forma simplificada con el fin de tender a reducir el número de símbolos fonéticos utilizados.
- Los casos en los que varios símbolos del alfabeto español están asociados a un solo símbolo pseudo-fonético, se presentan cuando las realizaciones acústicas de los primeros son idénticas o no presentan caracteres pertinentes diferentes (b = v, z = s, etc.).

Existen algunos casos especiales donde, por ser fonéticamente equivalentes, se utilizan secuencias de dos símbolos pseudo-fonéticos para representar a un solo símbolo del alfabeto español (ñ = NI, ll = LI, x = CS, etc.).

La salida del sistema se presenta según un archivo de texto escrito, el que incluye caracteres ASCII adecuados para el idioma español (á, é, í, ó, ú, ñ y Ñ). El texto presenta las líneas separadas por <CR> en correspondencia con las pausas de la entrada. La inclusión de un editor del texto de salida, permite modificarlo para una mejor presentación final.

Interfaz con el usuario

Un aspecto tratado cuidadosamente fue el desarrollo de la interfaz con el usuario, ya que, además de ser la interfaz para el módulo M1, lo es también para el sistema completo en el caso de que se lo utilice operando en tiempo real.

El programa de interfaz presenta menús tipo "pull-down" para activar las diferentes opciones organizadas como sigue:

Entrada fonética

Este menú permite ver primero la vía de ingreso del módulo M1 al módulo M2, con la información codificada en forma de un alfabeto pseudo-fonético.

Archivo ASCII:

Realiza la lectura de un archivo de texto que contiene la información fonética ya procesada. Esta es la opción por defecto.

Voz en tiempo real:

Pone en línea los módulos M1 y M2 de modo tal que el código intermedio sea tomado directamente de la RAM, sin que medie almacenamiento en disco posibilitando el procesamiento en tiempo real. En el prototipo desarrollado esta opción no se halla habilitada.

Teclado:

Abre una ventana en la que pueden ingresarse las cadenas continuas en código pseudo-fonético desde el teclado, de manera que se obtenga en forma inmediata la traducción al español en una ventana inferior.

Consulta datos de entrada

Esta opción pone a funcionar el "parser" según la modalidad escogida anteriormente. La información traducida va apareciendo en una ventana que puede ser usada para la posterior edición del texto.

Salida en español

Mediante esta opción puede escogerse la vía de salida de la información procesada, de acuerdo a las necesidades del usuario: archivo de texto en ASCII, impresora a puerta serie. En cualquier caso, la información es presentada simultáneamente en pantalla.

Diccionarios

Permite mantener actualizados los diccionarios del sistema, de acuerdo con las necesidades del usuario, posibilitando la

optimización de los mismos respecto de su extensión y su especificidad de vocabulario.

Con la información que se incorpora a través de la interfaz se amplía o corrige la base de datos, agregando nuevas palabras que pasarán a formar parte del vocabulario o modificando las ya existentes.

Ampliación de vocabulario:

Permite definir las nuevas palabras que se incorporan al diccionario activo. En una ventana se ingresa la cadena pseudo-fonética y a continuación su correspondiente traducción al español. Además pueden seleccionarse los atributos gramaticales de la palabra en cuestión: Función gramatical (sustantivo, adjetivo, verbo, adverbio, artículo, pronombre, preposición, conjunción, contracción y auxiliar), Género (masculino, femenino y neutro), Número (singular, plural e indiferente).

Modificación de vocabulario:

Permite la corrección de cualquiera de los datos correspondientes a una palabra que ya está presente en el diccionario activo.

Eliminación de palabras:

Solicita la palabra que se desea eliminar del diccionario activo.

Cambio de diccionario:

Permite activar un diccionario diferente al que está activo. Esta posibilidad optimiza los tiempos cuando se pretende operar con información cuyos vocabularios son muy diferentes (ej.: un contrato y un programa en Basic).

Las reglas que conforman la base de conocimientos no son accesibles al usuario, por lo tanto no es posible incorporar a la misma nuevas reglas, ni eliminar o modificar las existentes desde la interfaz. Esta particularidad no constituye una limitación importante de la interfaz, ya que las reglas gramaticales no se actualizan de la manera en que lo hacen otro tipo de reglas, tales como las de los sistemas expertos aplicados a otras áreas del conocimiento mucho más dinámicas.

Resultados y conclusiones

Durante el desarrollo del sistema, muchas de las premisas o especulaciones iniciales fueron modificándose por la fuerza de los hechos. Debido a que los dos grandes módulos que componen el sistema tienen identidad propia, pueden ser analizados por separado.

Las distintas subredes implementadas tuvieron un buen comportamiento durante el período de entrenamiento supervisado, antes de la interconexión entre ellas. La posibilidad de incidir sobre el aprendizaje en forma interactiva fue de gran utilidad para "guiar" a la red hacia la configuración más satisfactoria "evitando" mínimos locales.

La capacidad de discriminación de fonemas de la red varía de acuerdo al grupo de fonemas del que se trate. Es así que los fonemas pertenecientes a los grupos de las vocales, nasales y consonantes estables alcanzaron índices de reconocimiento superior al 95%, mientras que —por ejemplo— para las oclusivas el índice varió entre el 80 y 90%.

La velocidad del proceso de reconocimiento está fuertemente determinada por el tiempo que insume el preprocesamiento digital. La optimización de las rutinas implementadas, o el reemplazo de la FFT por un banco de filtros analógicos, puede permitir el trabajo del sistema en plataformas más populares que la configuración utilizada. Por otra parte, plataformas más potentes se hallan disponibles cada día, lo que permitiría difundir el sistema sin cambios apreciables en poco tiempo.

conecta a la compuerta de disparo del temporizador 1 de todos los MGP. Al programar los temporizadores en el modo 1, su salida permanece en "1" hasta que son disparados. Como se observa en las formas de onda, una vez disparado el MGP el temporizador 1 produce en su salida un valle, que al finalizar con flanco ascendente, dispara al temporizador 2 que suministra el pulso de salida del MGP. Se utilizan dos "drivers" para obtener el pulso y su complemento, mejorando las características de salida.

En resumen, al dispararse el temporizador 1 introduce una demora (valle) en el disparo del temporizador 2 (pulso), ambos programables y precisos.

El circuito desarrollado dispone de ocho MGP.

La entrada "reset" del FFD es comandada con un temporizador programable, lo que permite reproducir en forma secuencial un tren de pulsos sin tener que disparar, con la computadora, la secuencia cada vez.

Módulo sumador de pulsos (MSP)

Para poder formar un tren de pulsos se debe combinar uno o más pulsos provenientes de los MGP. A tal fin se incluyó un circuito implementado con compuertas "AND" y un regis-

tro de selección de pulsos ("latch").

Para poder seleccionar o no cada pulso, que conforma el tren, debe programarse el bit correspondiente del registro de selección con nivel lógico ("1"/"0"). Una compuerta "OR" de ocho entradas realiza la suma de los pulsos. Se incluyeron, en el GP, dos MSP para obtener simultáneamente dos trenes de pulsos distintos, el "A" y el "B". A su vez se dispone de la suma de ambos, "A + B". Todas las salidas del GP se disponen en forma directa y complementada. En la Fig. 4 se muestra un esquema del MSP.

Todos los dispositivos programables del GP se insertaron en un área disponible del "mapa" de puertos de la PC/XT/AT. La programación puede realizarse con cualquier lenguaje que permita el acceso a los puertos de la computadora (Pascal, C, QBASIC, etc.).

Conclusión

El generador de pulsos en cuestión está siendo utilizado con todo éxito en la Facultad de Matemática de la UNC, demostrando una versatilidad muy apropiada a los fines prácticos. Nótese además que el MSP puede ser implementado con PLA o PAL.

Interfaz oral para computadoras...

(de la pág. 609)

Los resultados obtenidos hasta el momento hacen promisoría la aplicación de este tipo de sistemas híbridos (redes neuronales + I.A.) al reconocimiento del habla en condiciones de más dificultad, tales como presencia de ruido, independencia del hablante, reducción del ancho de banda por uso de vía telefónica, reconocimiento del habla patológica, etc.

El comportamiento del "parser" es muy bueno y —para un diccionario de unas 1.000 palabras— responde con la velocidad necesaria para mantener una conversación rápida.

Uno de los aspectos que puede modificarse es la inclusión de símbolos en el alfabeto de entrada que indiquen signos de puntuación, de modo que el sistema pueda realizar el análisis de mayúsculas y el ajuste del texto de salida en forma automática.

Otra característica sobre la que se está trabajando es la posibilidad de que el sistema ofrezca respuestas alternativas frente a entradas ambiguas, cuya primera interpretación no sea satisfactoria. El agregado de otra fuente de conocimiento (semántica, pragmática, etc.) puede ayudar a resolver dichos casos, sin embargo el nivel sintáctico resuelve el problema de identificación de palabras satisfactoriamente para la mayoría de los casos prácticos. Esto se cumple absolutamente en el caso en la interpretación de sentencias de lenguajes de computadora donde el número de reglas ha sido disminuído al máximo y las ambigüedades eliminadas por completo.

Las restricciones del sistema en general están muy vinculadas a las limitaciones del primer módulo (M1), ya que de éste depende la fiabilidad del código fonético de entrada al módulo M2. Los errores en el archivo de entrada darán lugar —en esta primera versión— a errores en el archivo de salida, ya sea como frases erróneas o directamente no identificadas. Por otra parte, el alcance del análisis sintáctico que realiza el sistema está limitado a una sola cadena continua, sin tener en cuenta la coherencia sintáctica de dos frases o secuencias de palabras que ingresaron formando parte de cadenas fonéticas contiguas, lo que en un sentido es beneficioso, pero desaprovecha la información más global, que también podría utilizarse.

La posibilidad de utilizar también redes analógicas o híbridas sin demasiados cambios, permite un sistema de comunica-

ción oral totalmente abierto de muy bajo costo y gran velocidad, abriendo el campo a la implementación práctica de técnicas de tratamiento de lenguaje natural como interfaz de sistemas convencionales.

Referencias

- [1] Waibel, A.; Hanazawa, T.; Hinton, G.; Shikano, K. y Lang, K.: "Phoneme Recognition Using Time-Delay Neural Networks". IEEE Trans. ASSP Vol. 37, N° 3, 1989.
- [2] Lippmann, R.: "An Introduction to Computing with Neural Nets". IEEE ASSP Mag., Vol. 4, 1987.
- [3] Kohonen, T.: "The 'Neural' Phonetic Typewriter". Computer, Marzo 1988.
- [4] Rumelhart, D.; Hinton, G. y Williams, R.: "Learning Internal Representations by Error Propagation". Nature, Vol. 323, p. 533-536, Oct. 1986.
- [5] Travis, Bryan J.: "A Layered Neural Network Model Applied to the Auditory System", en Neural Networks for Computing, AIP Conference Proceedings 151, 1986.
- [6] Cohen, M.; Hurvitz, H.; Bernstein, J.; Price, P. y Weintraub, M.: "The Decipher Speech Recognition System" - Proc. IEEE ICASSP, 1990.
- [7] Sejnowsky, T.: "Open Questions About Computation in Cerebral Cortex", parallel Distributed Processing, Vol. 2, The MIT Press, 1986.
- [8] Kelly, J.: "Auditory System" en E. Kandel, J. Schwartz; Principles of Neural Science, Elsevier, 1985.
- [9] Martin, J.: "Cortical Neurons, the EEG and the Mechanisms of Epilepsy", en E. Kandel, J. Schwartz. Principles of Neural Science, Elsevier, 1985.
- [10] Lara y Zavala: "Cibernética del Cerebro". C.I.E.C.S.A., 1987.
- [11] Schafer, E. y Rabiner, L.: "Digital Representations of Speech Signals" IEEE Proc. Vol. 63 N° 4, 1975.
- [12] Rabiner, L. y Gold, B.: "Theory and Application of Digital Signal Processing", Prentice-Hall, 1975.
- [13] Rocha, L.: "Procesamiento de voz", I Escola Brasileiro-Argentina de Informática, Kapelus, 1987.
- [14] Guirao, M. y Borzone, A. M.: "Identification of Argentine Spanish Vowels", Journal of Psycholinguistic Research, Vol. 4 N° 1, 1975.
- [15] McClelland, J. y Rumelhart, D.: "Explorations in Parallel Distributed Processing", The MIT Press, 1987.
- [16] Rabiner, L. y Schafer, R.: "Digital Processing of Speech Signals", Prentice-Hall, 1978.
- [17] Chaves, J.; Rufiner, H.; Sigura A. y Zapata, D.: "Reconocimiento Automático de Fonemas Basado en Redes Neuronales", VII Reunión Científica de la Sociedad Argentina de Bioingeniería, 1990.
- [18] Gazdar & Mellish: "Natural Language Processing in Prolog", Addison-Wesley, 1989.
- [19] Simons, G.: "Introducción a la Inteligencia Artificial", 1987.
- [20] Arity Corp.: "The Arity/Prolog Language Reference Manual", 1987.
- [21] Arity Corp.: "Using The Arity/Prolog Interpreter and Compiler", 1987.
- [22] Fu, K. S.: "Syntactic Methods in Pattern Recognition", Academic Press, 1974.