PROYECTO ROARS: ROBUST ANALYTICAL SPEECH RECOGNITION SYSTEM*

José Miguel Benedí, Francisco Casacuberta y Enrique Vidal, Inmaculada Benlloch, Antonio Castellanos, María José Castro, Jon Ander Gómez, Alfons Juan y Juan Antonio Puchol.

Departamento de Sistemas Informáticos y Computación
Universidad Politécnica de Valencia

Abstract

The goal of the project is to increase the robustness of an existing analytical speech recognition system and to use it as part of a speech understanding system with connected words and dialogue capability. This system will be evaluated for one specific application in two European Languages.

The main objectives of the project are: the first is to have at our disposal a robust analytical (i.e. using word, syllable, phoneme and feature levels) speech recognition for Spanish and French languages with the following features: speaker independent, real time processing, medium size vocabulary (up to 250 words) and sentence error rate ≤2%. And the second objective is to demonstrate the usefulness of the system in an application domain of speech understanding (Air Traffic Control). In this application an elaborated multimedia dialogue will be used: feedback, error correction, elision, use of context, help, etc.

1.- Introducción y Objetivos

El proyecto lleva por título "Robust Analytical Speech Recognition System" (ROARS). Ha sido presentado y aceptado en la subárea I.5.1. "Robust Speech Understanding" de ESPRIT II con el número (5516), siendo su duración estimada en 36 meses.

El consorcio creado para la realización de ROARS está formado por 4 participantes, con un esfuerzo total de 37 años hombre (4.1 correspondientes a nuestro grupo), y un coste total previsto de 4592899 ECUs (401280 correspondientes a nuestro grupo). Los grupos participantes son:

THOMSON SINTRA ASM (Francia) (Coordinador principal
CRIN, Univ. de Nancy I (Francia)
ENA Telecomunicaciones (España)
Univ. Politécnica de Valencia (DSIC) (España)

El proyecto toma como punto de partida un sistema realizado inicialmente para la lengua Francesa [Alinat et al., 1987], que ha demostrado su utilidad en aplicaciones independientes del locutor, operando en tiempo real y con unos resultados satisfactorios en palabra conectada.

* Trabajo parcialmente subvencionado por ESPRIT II, proyecto 5516 (ROARS)

En base a este sistema actual, los objetivos principales de ROARS se pueden sintetizar:

El *primer objetivo* consiste en obtener un sistema de Reconocimiento Analítico del Habla para las lenguas Francés y Castellano, robusto y fiable. Se propone incrementar la robustez del sistema actual en los siguientes aspectos:

1) Incorporación de conocimiento acerca a de los cambios articulatorios producidos en los locutores y entre locutores, incluyendo la posibilidad de una progresiva adaptación automática.

2) Análisis de la degradación producida en el sistema (sobre todo a nivel acústico fonético) en ambientes ruidosos y a través de las líneas telefónicas.

El *segundo objetivo* consiste en demostrar la utilidad del sistema en el dominio de aplicaciones de comprensión del habla. En particular se ha pensado en una aplicación sobre una estación de trabajo, que incluya la integración de la entrada vocal a otros dispositivos y que emplee un *diálogo "multimedia"* elaborado. La aplicación concreta (de la parte española del consorcio) es: *Control de tráfico aéreo.*

2.- Descripción del Proyecto

Para la consecución de los objetivos planteados se proponen actuaciones en diversos aspectos. A continuación se detallan cada uno de dichos aspectos y la investigación a desarrollar en cada uno de ellos.

ORGANIZACION DEL SISTEMA ANALITICO DE RECONOCIMIENTO DEL HABLA

La arquitectura propuesta para el sistema se muestra en la figura 1, y consta de los siguientes elementos:

Analizador.- La adecuada elección y funcionamiento del analizador es muy importante y especialmente crítico para el resto del sistema. El análisis de la señal vocal está basado en el comportamiento perceptual del oído humano. Por esta razón se utiliza un banco de 48 filtros pasa banda [Rabiner y Schafer, 1978] cuya frecuencia central y función de transferencia representan la división natural de frecuencias audibles [Zwicker y Terhardt, 1980]. Otras características del analizador son: el ancho de banda es de 70Hz-10KHz y se obtiene una muestra del "espectro" de la señal cada 8 ms.

Estimación de los parámetros acústicos.- La salida del analizador es procesada para estimar un conjunto de parámetros acústicos. La obtención de estos parámetros se realiza de diferentes formas: a) Parámetros extraídos para cada muestra espectral: Energía, forma del espectro y posición de los "formantes" [Flanagan, 1972]. b) Parámetros extraídos teniendo en cuenta variaciones en espectros sucesivos: "burst" en diferentes zonas del espectro y nivel de ruido ambiente estacionario. Y c) un canal especial para la estimación de la naturaleza sonora-sorda del habla, el ruido de fricción y la frecuencia fundamental ("pitch").

Localización de los "fonos".- Los parámetros acústicos son utilizados para la determinación de las características acústico-fonéticas. En adelante se denominará "fonos" a las categorías acústico-fonéticas reconocidas, en contraposición a los "fonemas" como categorías lingüísticas para la descripción de las palabras del vocabulario. La localización de los fonos en la frase pronunciada se realiza en base a la definición de unas categorías fonéticas ("toscas")

generales [Cole y Hon,1988] y mediante la elaboración de reglas específicas, en términos de los parámetros acústicos, para cada una de dichas categorías o clases.

Todas las clases son analizadas en paralelo, y aquellos fonos detectados con mejor precisión serán marcados con una prioridad mayor, seguidos por el resto. Para resolver conflictos se usan reglas especiales. Se debe enfatizar el hecho de que este proceso de localización es diferente del clásico proceso de segmentación. Aquí el concepto de frontera entre segmentos adyacentes simplemente no existe.

Estimación de los parámetros del fono.- Para cada una de las clases se determina un conjunto de parámetros con las características esenciales atribuibles a dicha clase. Estos parámetros están relacionados con las propiedades que se estima que los fonemas deben poseer para pertenecer a esta clase.

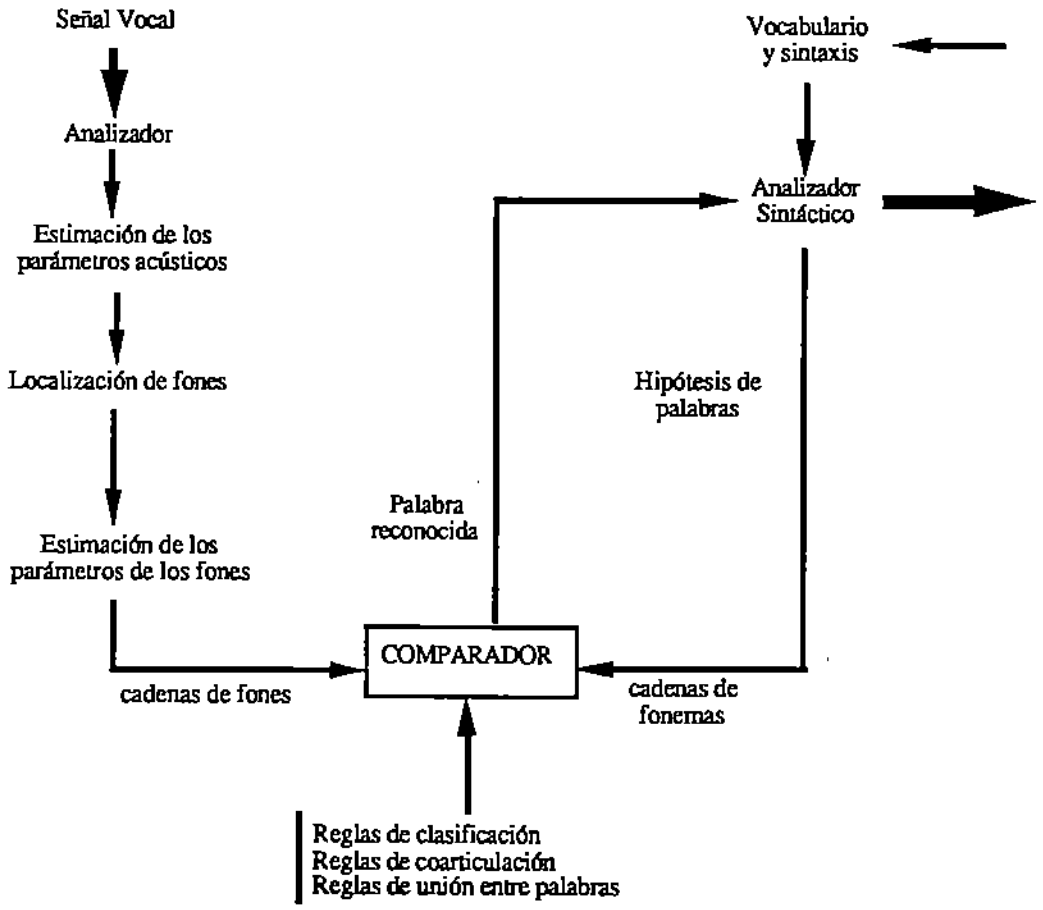

Figura 1.- Arquitectura general del Sistema Analítico de Reconocimiento del Habla.

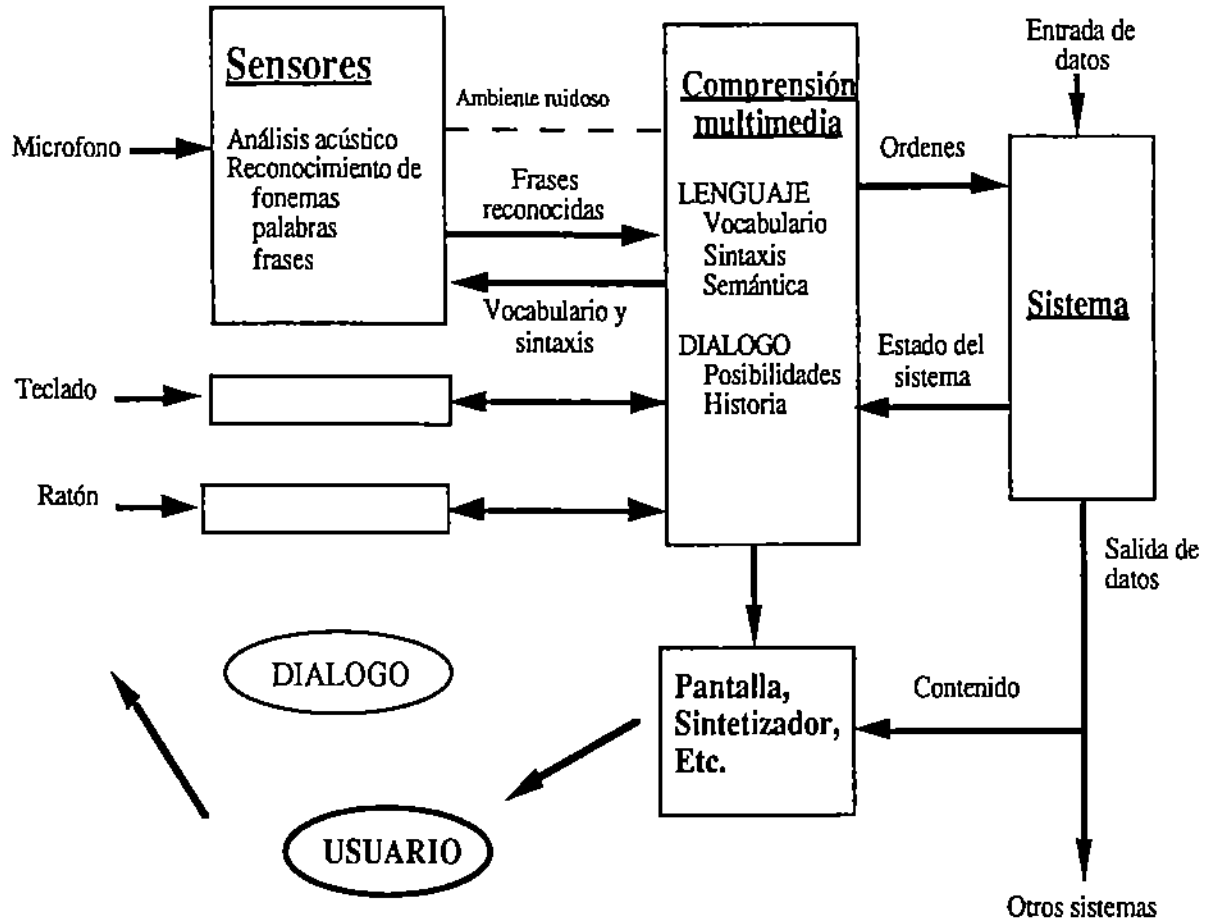

Figura 2.- Sistema de Diálogo y Comprensión "Multimedia"

De este modo se obtiene una secuencia o cadena de fonos, donde cada fono se describe en términos de la clase a la que pertenece y el valor de sus parámetros asociados. Hay que advertir que no se obtiene una cadena de fonemas etiquetados, ya que, debido a las fuentes de variabilidad del habla [Casacuberta y Vidal, 1987], se asume la imposibilidad de determinar con precisión la cadena fonética sin un conocimiento contextual (sintáctico-semántico). En otras palabras (figura 1), la relación entre unidades acústicas y unidades lingüísticas está gobernada por el conocimiento sintáctico-semántico de la tarea. Esta es la razón por la que se denomina a este sistema "Analítico."

Proceso de Comparación.- Para cada aplicación, el lenguaje de la tarea se define por medio de su sintaxis y su vocabulario. Cada palabra en el vocabulario se describe en términos de su cadena fonética. La evolución del sistema está gobernado por el Analizador Sintáctico, que en cada etapa de análisis propone al proceso de comparación (figura 1) ciertas hipótesis de palabras. La comparación entre una cadena de fonemas (que describe una palabra del vocabulario) y la cadena de fonos (proporcionada por las etapas de análisis fonético) se realiza utilizando técnicas de Programación Dinámica. Y para la comparación elemental fono-fonema, se toma simultáneamente en cuenta la identidad del fonema y la clase y parámetros del fono. La evaluación de dicha comparación depende de tres clases de conocimiento adicional, a saber: a) Reglas de clasificación, que determinan la definición de cada fonema; b) Reglas de coarticulación; c) Reglas de unión entre palabras.

ROBUSTEZ A NIVEL FONETICO

Se propone incrementar la robustez a nivel fonético frente a cambios articulatorios producidos en los locutores y entre locutores, y frente a las distorsiones producidas por diversos ruidos ambientales.

Robustez frente a cambios en, y entre, los locutores.- Los resultados de un sistema de reconocimiento analítico dependen de la calidad del conocimiento tanto acerca de las características acústico-fonéticas como a nivel fonémico. Existen diversos resultados experimentales obtenidos por los sistemas existentes, con diversos locutores, varios vocabularios (aplicaciones) y distintas condiciones experimentales [Mercier et al., 1989], [Zue et al., 1989], [Benedí, 1989], etc. Del análisis de estos resultados se han detectado limitaciones y posibles defectos. Este tipo de experiencias deben continuar (en los dos lenguajes del proyecto) con el propósito de obtener un conocimiento estadístico más completo de los fonemas y las características acústico-fonéticas.

Hasta el momento, se está asumiendo implícitamente que el conocimiento contenido por el sistema es absolutamente independiente del tiempo y del locutor. Sin embargo, para obtener mejores resultados, se ha previsto que se pueda establecer un método que adapte continua y gradualmente este conocimiento para optimizar los resultados alcanzados ante características de locutores muy diferentes, acentos de distintas regiones, etc.. Sin que para ello sea necesario ningún sistema explícito de aprendizaje dialogado. Existe muy poco trabajo realizado en este área, y debe verse este problema más como un objetivo especulativo que como una tarea prioritaria.

Robustez frente a ruidos ambientales.- Son diversos los tipos de ruido que pueden perturbar el habla, a saber:
- Ruido estacionario más o menos gausiano, ruidos de banda estrecha y de banda ancha.
- Ruidos impulsivos repetitivos.
- Ruidos explosivos aislados, etc.

Para el estudio de la degradación del sistema en ambientes ruidosos, se plantea abordar los siguientes puntos: En primer lugar, es necesario estimar los parámetros del ruido, y modificar el sistema de reconocimiento para que sea capaz de detectar periodos "mudos" de señal.

En segundo lugar, el sistema de reconocimiento debe tener en cuenta la influencia tanto del ruido estacionario (si la relación señal/ruido del habla no es suficiente) como de fuertes ruidos explosivos, sobre cada una de las características fonéticas, y poder así minimizar la degradación inducida por el ruido [Fink y Dalsgaard, 1989]. Esto es una tarea muy compleja porque, en el proceso de reconocimiento, la importancia y el papel de cada característica fonética debe adecuarse dependiendo del valor de los parámetros del ruido [Gómez, Sánchez y García, 1990].

Finalmente, en condiciones ruidosas, los locutores adaptan sus características articulatorias para permitir una mejor comprensión por parte del interlocutor(es) (efecto Lombard) [Hanson y Applebaum, 1990]. Este problema también debe considerarse en el desarrollo del sistema y debe usarse adecuadamente su conocimiento correspondiente.

COMPRENSION Y DIALOGO

El propósito principal de un sistema de comprensión es determinar el mensaje asociado a la pronunciación reconocida y ejecutar las acciones prescritas. En el marco de un sistema de

comprensión una componente de diálogo es muy útil y necesaria para eliminar ambigüedades, informar al operador, confirma y/o corregir la sentencia, etc.. Las propiedades relevantes de un sistema de diálogo en la comunicación hombre-máquina pueden ser tan distintas como diferentes sean los tipos de aplicaciones [Pierrel, 1987]. Esto hace que no sea posible en general diseñar una componente de diálogo común y general a todas las aplicaciones. Aunque los problemas de diálogo son tratados dependientemente de la aplicación en esta propuesta de proyecto, se pueden considerar algunos principios generales.

La utilización de una entrada vocal a un sistema dista mucho de ser simplemente la mera implantación de un micrófono. Es importante y extremadamente útil implementar un *sistema de comprensión "multimedia"*; es decir, una integración completa de la entrada vocal con otros medios, principalmente el visual (figura 2). Esta integración sería necesaria para:

-Validar los resultados del sistema de reconocimiento, ya que diversos errores y ambigüedades pueden producirse.
-Ser capaz de utilizar comprensión contextual.
-Utilizar realimentación visual o acústica, y en algunas circunstancias, emplear explícita o implícitamente procedimientos de confirmación.
-Posibilidad de una función de ayuda ("help") para la utilización en su caso por un usuario no enseñado.
-Establecer una sucesión de acciones tendentes a alcanzar el objetivo del usuario y consecuentemente conocer como determinar este objetivo.

Para mantener este proyecto en posiciones realistas, el diálogo utilizado para demostración estará limitado a *diálogos terminales*; es decir, orientados a la tarea y cooperativos, Debido a esta característica del diálogo y las restricciones en la complejidad de la tarea, los lenguajes de definición de la aplicación serán de vocabulario limitado y sus gramáticas generativas muy simples. Usualmente serán gramáticas de palabras conectadas con algún grado de flexibilidad; en otras palabras, se trata de implementar la versión hablada de un menú jerarquizado.

DEMOSTRACION

Para validar el sistema globalmente (sistema de reconocimiento del habla y sistema de comprensión multimedia) es necesario comprobarlo en el contexto de una aplicación real. Se implementarán dos sistemas de test, teniendo en cuenta las dos lenguas (Francés y Castellano). La evaluación del sistema se deberá realizar de forma tan cuantitativa como sea posible, comparándolo con la misma tarea sin entrada vocal (medida de la calidad ergonómica). Para la parte española del consorcio se propone el siguiente dominio de la aplicación:

Consola para Control de Tráfico Aéreo.- La tarea del controlador de tráfico aéreo es muy pesada ya que se debe responsabilizar de muchas trayectorias de vuelos, además de poder dar ordenes a los pilotos directamente, en caso necesario, y emplear un gran número de dispositivos: Una pantalla geográfica principal con líneas de tráfico, secciones, etc; una pantalla secundaria con informaciones alfanuméricas; "bandas" de papel para cada vuelo; etc.

Las características más importantes de la aplicación propuesta son: integración con otros dispositivos de entrada, talla del vocabulario reducida (100 a 150 palabras), palabras conectadas y frases cortas (máximo 6 palabras por frase), tiempo real, y diálogo multimedia elaborado: uso de contexto, multiplexado voz-teclado, realimentación, corrección de errores, ayuda, etc.

Esta demostración no será utilizada en una situación real pero sí en unas condiciones de laboratorio realistas; principalmente desde el punto de vista de ruido ambiente, integración con otros medios de intercomunicación, usuarios entrenados y no entrenados, multilocutor y procesamiento en tiempo real.

3.- Planificación y estado actual

En el plan de trabajo propuesto para el proyecto ROARS, el grupo de la Universidad Politécnica de Valencia se encuentra involucrado en todas las tareas de adquisición y representación del conocimiento de la versión en Castellano para su posterior implantación en el sistema (versión base, análisis de los cambios articulatorios en y entre locutores, y análisis de los cambios articulatorios debido al ruido ambiente), así como en las tareas de especificación, análisis de las variaciones del diálogo y evaluación de la versión española del sistema de demostración.

El proyecto comenzó el 1 de Enero de 1991 y en el momento actual (Mayo 91) se encuentra en fase de construcción de un primer prototipo del sistema de Reconocimiento Analítico del Habla, que deberá estar terminado en Marzo de 1992. En paralelo se están realizando los estudios de la influencia de la variabilidad articulatoria en y entre locutores, y de la degradación del ruido en el conocimiento acústico-fonético que se posee.

4.- Referencias Bibliográficas

Alinat,P., E.Gallais, J.P.Haton, J.M.Pierrel and P.Richard(1987): "A Continuous Speech Dialog System for Oral Control of a Sonar Console". **IEEE ICASSP, Dallas.** Vol.1, pp.368-371.

Benedí, J.M.(1989): **Un sistema de Reconocimiento Automático del Habla: TABARCA.** Tesis Doctoral. Universidad Politécnica de Valencia.

Casacuberta, F., and E. Vidal (1987): "Reconocimiento Automático del Habla.". Marcombo.

Cole,R.A. and L.Hon(1988): "Segmentation and Broad Classification of Continuous Speech". **IEEE ICASSP, New York.** Vol.1, pp.453-456.

Flanagan, J.L.(1972): **Speech Analysis Synthesis and Perception.** Springer Verlag.

Fink,F.K. and P.Dalsgaard (1989): "Estimation of Formants in Noise Corrupted Speech Using Auditory Models". **Eurospeech.** Vol.2, pp.677-689.

Gómez,J., L.Sánchez and R.García(1990): "A Comparative Study of Feature Extraction Methods for Noisy Speech Recognition". In **Signal Processing: Theories and Applications.** Eds. L.Torres, E.Masgrau and M.A.Lagunas. Elsevier Science Publishers. Vol.2, pp.1191-1194.

Hanson,B.A. and T.H.Applebaum(1990): "Robust Speaker-Independent Word Recognition Using Static, Dynamic and Acceleration Features: Experiments with Lombard and Noisy Speech". **IEEE ICASSP, Albuquerque.** Vol.2, pp.857-860.

Mercier,G. et al. (1989): "Recognition of Speaker-Dependent Continuous Speech with KEAL". **Proceedings of IEEE.** Vol.36, No.2, pp.145-154.

Pierrel,J.M.(1987): "Aspects of Man-Machine Voice Dialog" in **Fundamentals in Computer Understanding** Ed. J.P.Haton Cambridge University Press, pp.249-274.

Rabiner, L.R., and R.W. Schafer(1978): **Digital Processing of Speech Signals.** Prentice Hall.

Zue,V., J.Glass, M.Phillips and S.Seneff(1989): "Acoustic Segmentation and Phonetic Classification in the SUMMIT system". **IEEE ICASSP, Glasgow.** Vol.1, pp.389-392.

Zwicker, E., and E. Terhardt(1980): "Analytical Expressions for Critical-Band Rate and Critical-Bandwidth as a Function of Frequency". **J.Acoust.Soc.Am.** Vol.68, pp.1523-1525.

ASELA: Análisis, Síntesis y Evaluación del Lenguaje y la Audición.

Victoria Marrero*, Andrés de Santos+, Santiago Aguilera+

* Dpto Lengua Española
Facultad de Filología. U.N.E.D.

+ Dpto. Ingeniería Electrónica.
E.T.S.I.Telecomunicación. U.P.M.

El equipo que presenta este trabajo es un grupo interdisciplinar constituido, fundamentalmente, por ingenieros expertos en proceso digital del habla (Departamento de Ingeniería Electrónica de la U.P.M.) y lingüistas especializados en fonética y fonología (Departamento de Lengua Española de la U.N.E.D.), en colaboración con audiólogos (Hospital Ruber Internacional) y logopedas.

Nuestros estudios se han centrado en varios temas que comparten un objetivo prioritario: su utilidad para el tratamiento y ayuda a discapacitados visuales y auditivos.

De los trabajos presentados aquí, los dos primeros (*Conversión Texto-Habla* -con atención a las aplicaciones para invidentes- y *Sistemas para la Representación y el Análisis del Lenguaje Natural* -ayudas para sordos-) son implementables en una única arquitectura hardware adaptable a cualquier PC, y constituyen el proyecto SIPHAD (*Sistema Integral de Procesamiento del Habla para Ayuda a Discapacitados*). El tercero, (*TRD, Test de*

Rasgos Distintivos) está encaminado a proporcionar un conjunto de pruebas para la evaluación del lenguaje sintético, por un lado, y como complemento a las pruebas clínicas de audiometría verbal por otro.

1. Conversión Texto-Habla

Desde el año 1981 hemos trabajado en la elaboración de un conversor texto-habla en tiempo real para el castellano.

Se ha incidido especialmente es tres áreas:

- estudio de los rasgos prosódicos, fonológicos y segmentales de la lengua hablada (duraciones, entonación, transición de formantes, además de la caracterización general del sistema fónico);
- programación de algoritmos que realicen el proceso completo de conversión texto-habla (normalización del texto, conversión de letras en sonidos, asignación de patrones prosódicos, obtención de parámetros y síntesis del habla);
- diseño de sistemas hardware que soporten dichos algoritmos en tiempo real; entre otros se ha desarrollado una placa con un procesador de señales preparada para su conexión a un ordenador personal. Esta placa soporta además otras aplicaciones que se describirán más adelante.

Actualmente estamos trabajando en la mejora de la calidad del habla producida, así como en añadir nuevas posibilidades al sistema.

Según las pruebas de inteligibilidad realizadas, algunos

sonidos presentaban ciertas deficiencias debidas, entre otros factores, a la dirección y velocidad de las transiciones, distribución de áreas de frecuencia, asignación de intensidad, etc. Una vez analizados estos elementos, se está procediendo a su reajuste.

Otro factor que estamos mejorando es el modelo que asigna la duración a cada alófono, especialmente considerando diversas velocidades de producción de habla (*tempo* medio, rápido o lento).

En el momento actual llevamos a cabo también la modificación de determinadas características del sintetizador para permitir la simulación de distintos hablantes, utilizando para ello, tanto modificaciones en la propia estructura del sintetizador (en el modelo de excitación glotal) como cambios en la prosodia y en la caracterización de sonidos.

Entre las nuevas posibilidades del sistema, contamos también con su conexión a diversos programas de uso común (procesadores de textos, bases de datos, etc.), con el fin de proporcionarles una salida hablada, de gran utilidad para deficientes visuales. Las mejoras introducidas (variaciones del tipo de voz y de su velocidad de producción, mejora en la calidad del habla) contribuirán a dar más utilidad y mayores prestaciones a estas aplicaciones.

2. Sistemas para la representación y análisis de lenguaje natural.

Nuestro grupo dispone hoy en día de un sistema potente y versátil para la detección, análisis y posterior representación de diversos parámetros de la voz.

Sobre una única placa hardware, que se conecta al bus de un ordenador personal y se configura por software desde éste, disponemos de dos tipos fundamentales de aplicaciones, además de la ya mencionada de conversión texto-habla:

- Analizador de voz (PCvox). Nos porporciona una representación gráfica en color de los parámetros más representativos de la voz: forma de onda, energía, entonación, cruces por cero y sonogramas; sobre ellos se pueden llevar a cabo numerosas operaciones: segmentación temporal (con posibilidad de escuchar sólo el trozo segmentado, e incluso edición de voz), segmentación frecuencial, inserción de marcas (para la transcripción fónica), modificación del filtrado, modificación de la intensidad de entrada, etc.

El PCvox ofrece las prestaciones más importantes de los sonógrafos tradicionales, pero las integra en un sistema muy manejable, multifuncional y más asequible, por lo que comienza a ser solicitado desde diferentes campos técnicos y profesionales.

- Sistema para el entrenamiento de parámetros suprasegmentales del habla (ISOTON). Sistemas de Ayuda a Sordos

El sistema ISOTON, (acrónimo de *I*ntensidad, *S*onoridad y *Ton*o, parámetros que permite entrenar), está constituido por un módulo de proceso digital de señal, que extrae estos elementos a partir

de la señal de voz suministrada desde micrófono, y por un programa que los procesa y presenta sobre la pantalla de un ordenador personal.

Es un sistema destinado al entrenamiento del lenguaje que utiliza realimentación visual, permitiendo la retención y superposición de imágenes para analizar con posterioridad los resultados de la pronunciación y compararlos con patrones previamente establecidos. También dispone de diversos videojuegos controlables por la voz.

Permite tres modos de funcionamiento, en cada uno de los cuales se pueden entrenar distintos parámetros del habla:

a) *Modo Intensidad:* entrenamiento de ritmo, acentuación y duración
b) *Modo Sonoridad:* entrenamiento de los rasgos oclusión/fricación y sonoridad/sordez.
c) *Modo Tono:* entrenamiento del tono fundamental.

Actualmente estamos ampliando sus posibilidades para detectar nuevos parámetros del habla, como el rasgo de nasalidad.

Otras aplicaciones en las que estamos trabajando actualmente son un sistema de audiometría tonal informatizada y un compresor de frecuencias personalizado. El primero permitirá la simulación de un audiómetro básico de forma automática. El segundo utilizará técnicas de trasposición y compresión de frecuencias, adaptándolas a las características fónicas peculiares de nuestra lengua, con el fin de aprovechar al máximo los restos auditivos del hipoacúsico.

3. Pruebas para la Evaluación del Lenguaje Sintético.

Desde el comienzo de las investigaciones sobre el tratamiento automático del habla, los equipos que trabajaban en síntesis de voz en nuestro país se enfrentaban al problema de la elaboración de tests para probar la inteligibilidad de sus sistemas. Normalmente esta cuestión se ha solventado acudiendo a las pruebas elaboradas en Estados Unidos y adaptándolas, con mayor o menor fortuna, al castellano.

Nosotros hemos elaborado un Test que, conservando las normas propias de las pruebas más utilizadas internacionalmente, tiene muy en cuenta las peculiaridades de nuestra propia lengua, tanto en el plano fónico como en todos los niveles del sistema lingüístico: el TRD, Test de Rasgos Distintivos.

En su estructura presenta las características esenciales del DRT (*Diagnostic Rhyme Test* [1]): un test de respuesta cerrada (determinados pares de palabras) para la diferenciación de las consonantes, que persigue reducir en todo lo posible la información proporcionada por el contexto.

Para delimitar al máximo los factores de respuesta al estímulo, hasta reducirlos a una única variable pertinente (un único rasgo distintivo) se han seguido los siguientes pasos:

1) Selección de pares mínimos: parejas de palabras que sólo difieren entre sí en un fonema y gracias a un único rasgo distintivo: *peso/beso* (sonoridad/sordez).

[1] El primer trabajo publicdo a este respecto fue el de G.Fairbanks: "Test of Phonemic Differentiation: The Rhyme Test". *J.A.S.A.*30–7, 1958, pp.596–600; En los años posteriores, hasta la actualidad, ha ido apareciendo numerosa bibliografía al respecto, con variaciones sobre el material original, con la creación del MRT (*Modified Rhyme Tests)*, etc., pero no es éste el momento de entrar en ello.

2) Control de rasgos secundarios. En cada par mínimo, los fonemas que oponen las dos palabras presentan, además del rasgo distintivo diferenciador, otros varios que deben ser controlados para evitar que su presencia falsee los resultados del test.

3) Control del contexto: selección del patrón silábico de las palabras, de la posición del fonema en la sílaba, del entorno en el que aparece, etc.

4) Homogeneización morfológica, con el fin de evitar los desequilibrios consecuencia de la diferente frecuencia relativa de cada clase de palabras.

5) Control de frecuencia y familiaridad, para evitar las distorsiones que produciría la comparación de una palabra muy conocida con otra poco o nada usual para el oyente.

Actualmente estamos adaptando esta prueba para su utilización con lenguaje natural, adecuando su presentación y posibilitando el procesamiento automático de la información que nos proporciona, con el fin de emplearlo en la práctica clínica, dentro del campo de la logoaudiometría, como complemento a otras pruebas (de Umbral de Recepción Verbal, de discriminación, audiometría verbal infantil, etc.), elaboradas también por una parte de este equipo.