

# **SERVIDORES VOCALES INTERACTIVOS: DESARROLLO DE UN SERVICIO DE PÁGINAS BLANCAS POR TELÉFONO CON RECONOCIMIENTO DE VOZ PROYECTO IDAS (Interactive telephone-based Directory Assistance Service)**

R. San-Segundo, J. Colás, J.M. Montero, R. Córdoba, J. Ferreiros,  
J. Macías-Guarasa, A. Gallardo, J.M. Gutiérrez, J. Pastor, J.M. Pardo  
Grupo de Tecnología del Habla. Departamento de Ingeniería Electrónica. UPM  
E.T.S.I. Telecomunicación. Ciudad Universitaria s/n, 28040 Madrid  
Telf: 91 5495700 ext. 403 Fax: 91 3367323 [lapiz@die.upm.es](mailto:lapiz@die.upm.es)

**Resumen.-** En esta comunicación se presenta los desarrollos realizados en el marco del proyecto IDAS (LE4-8315). El objetivo del proyecto es desarrollar un demostrador capaz de dar un servicio de páginas blancas por teléfono que facilita números de teléfono (o fax) particulares y de empresas. El sistema de reconocimiento utilizado es de habla aislada para grandes vocabularios e independiente del locutor. En esta comunicación describimos la primera versión del demostrador con vocabularios de 1000 palabras. El servicio desarrollado funciona en tiempo real en un ordenador Pentium II-350Mhz permitiendo una tasa del 55% de llamadas procesadas automáticamente.

## **I.- INTRODUCCIÓN**

En los últimos años, las tecnologías del habla han consistuido un campo importante de investigación, pasando de ser un objetivo meramente científico a ser un objetivo claramente comercial. En la comercialización de estas tecnologías tienen un protagonismo relevante los *Servidores Vocales Interactivos* (SVIs). Un SVI no es más que un sistema capaz de proporcionar un servicio de adquisición o difusión de información a través de la línea telefónica, utilizando síntesis y reconocimiento de voz. Tradicionalmente, los operadores de telefonía han venido dando este tipo de servicios a través de operadores que atendían personalmente las llamadas. La automatización que se puede conseguir mediante la introducción de las tecnologías del habla, y la reducción de costes asociada, está despertando un gran interés comercial.

El proyecto IDAS, objeto de esta comunicación, es un proyecto a dos años (1998-2000), financiado por la Unión Europea en el que interviene la Universidad Politécnica de Madrid a través de nuestro grupo de investigación. El objetivo del proyecto es desarrollar un demostrador capaz de dar un servicio de páginas blancas por teléfono, proporcionando números de teléfono o fax, tanto de particulares como de empresas. Un número de teléfono particular queda determinado, para el caso de nuestro demostrador, al conocer la ciudad donde vive la persona, su nombre y sus dos apellidos. Por otro lado, un número de teléfono de una empresa se puede obtener conociendo la ciudad donde está localizada y su nombre. El sistema de reconocimiento desarrollado es de habla aislada para grandes vocabularios (10000 palabras) e independiente del locutor, siendo el que se describe en esta comunicación la primera versión del demostrador, con vocabularios más reducidos de hasta 1000 palabras.

## **II.- ANTECEDENTES**

### *II.1 Entorno de desarrollo de aplicaciones telefónicas.*

Desde 1992, el Grupo de Tecnología del Habla, demostrando una gran visión de futuro, apostó firmemente por la integración de las tecnologías del habla en los sistemas de información a través de la línea telefónica. Con esta orientación se comenzó un ambicioso proyecto cuyo objetivo fue desarrollar un entorno integrado de desarrollo de aplicaciones telefónicas TADE (Telephone Application Development Environment). Este entorno proporciona un nuevo lenguaje, con ciertas primitivas de alto nivel, para el diseño de aplicaciones telefónicas, principalmente SVIs, incluyendo igualmente utilidades para cubrir todo el ciclo de vida de una aplicación:

diseño, compilación y ejecución [1]. De forma general, una aplicación telefónica especificada en nuestro lenguaje consta de las siguientes partes:

- *Inicialización de variables*: donde se configuran las variables generales del sistema (las relacionadas con la línea telefónica, los tiempos máximos de espera en la interacción con el usuario, etc.).
- *Tratamiento de errores*: donde se indican las acciones a realizar por el sistema cuando se produzcan algunos de los errores considerados, (máximo tiempo excedido sin que el usuario conteste, número máximo de intentos fallidos por parte del usuario, colgado antes de finalizar la aplicación o error general del sistema).
- *Subrutinas*: donde se definen e implementan las subrutinas de las cuales se harán uso en la aplicación.
- *Aplicación*: que constituye el conjunto de instrucciones que se irán ejecutando de forma secuencial. El lenguaje diseñado ofrece sentencias de gestión de la línea telefónica (como colgar, descolgar, marcar o esperar llamada), sentencias de voz (reconocimiento, síntesis en Español [2], reproducción y grabación), de acceso a bases de datos (abrir/cerrar bases de datos y ejecutar consultas), de correo electrónico (enviar e-mail), de gestión de directorios y archivos, de manejo de cadenas y operaciones aritméticas básicas.

Además de las instrucciones comentadas anteriormente, el sistema ofrece análisis del progreso de llamada (*CPA: Call Progress Analysis*) configurable, redirección de llamadas dentro de una misma centralita (Alcatel o Ibercom) y la activación de un hilo musical en caso de accesos largos a bases de datos.

Como ejemplos de *SVIs* desarrollados con este entorno cabe destacar el sistema de atención al cliente de Hewlett Packard, que identifica al usuario a través del reconocimiento de su código de cliente (por voz o DTMF) y redirige la llamada hacia alguno de los ingenieros que esté libre en esos momentos, el servidor de notas que ofrece nuestro departamento a sus alumnos, para consultar sus calificaciones por teléfono sin más que identificarse con su DNI. Este mismo servicio se ofrece en el Rectorado de la Universidad Politécnica de Madrid para dar las calificaciones definitivas de todas las carreras de la Universidad Politécnica. Otro servicio es el de buzón vocal que está instalado en nuestro Departamento y permite atender las llamadas de cualquier despacho en los casos de teléfono comunicando o cuando no contesta. En este servicio, el sistema ofrece la opción de grabación de mensajes, enviando un e-mail de aviso al destinatario de la llamada.

## *II.2 Módulo de preselección para sistemas de reconocimiento de habla aislada de gran vocabulario.*

El nuevo sistema de reconocimiento a incorporar en el Entorno TADE está basado en TIDAISL[3] (proyecto realizado por nuestro grupo en colaboración con Telefónica I+D) y desarrollos posteriores[4][5]. El objetivo del proyecto TIDAISL fue el desarrollo de un módulo de preselección para sistemas de reconocimiento de habla aislada de gran vocabulario. Un módulo de preselección consiste en un conjunto de algoritmos de reconocimiento poco costosos a través de los cuales se permite reducir la lista de candidatos a reconocer. Esta reducción en los candidatos permite aplicar después técnicas de reconocimiento más potentes sin aumentar excesivamente el tiempo de proceso. El módulo obtenido permitía, partiendo de una serie de ficheros parametrizados por Telefónica I+D, reducir un vocabulario de 10000 palabras a uno de 1000 (lista de preselección) con el 98% de certeza de que la palabra a reconocer permanecía en dicha lista. El módulo de preselección consta de dos partes:

- *Extracción de la cadena fonética (decodificador acústico)*: para la implementación de esta etapa se ha utilizado el algoritmo de un paso, cuyo objetivo es determinar la secuencia de alófonos que con mayor probabilidad se ajusta a la secuencia de observaciones acústicas recibidas. Los modelos probabilísticos utilizados para la decodificación acústica son los HMM semicontinuos (SCHMM).
- *Acceso léxico*: se trata de un algoritmo de programación dinámica que recibe a su entrada una cadena o bien una malla de unidades fonéticas (alófonos), y genera a su salida una serie de palabras candidato que el sistema clasifica como las que con más probabilidad representan a la cadena (malla) de entrada.

En el marco de este proyecto se trabajó también con modelos SCHMM múltiples para la parte de decodificación acústica y con el diccionario estructurado en forma de árbol para la etapa de acceso léxico.

### III.- DESARROLLO DE LA APLICACIÓN DE PÁGINAS BLANCAS

Utilizando el entorno de desarrollo aplicaciones telefónicas comentado anteriormente (TADE), nos propusimos el desarrollo del demostrador objeto del proyecto IDAS. La aplicación desarrollada es la que se describe en los siguientes pasos:

- 1) En primer lugar, se descuelga y se da un mensaje de bienvenida al usuario y se procede al reconocimiento de la ciudad de la que desea conocer el teléfono, pidiendo confirmación al usuario del primer candidato (y si fuese necesario, del segundo) resultado del reconocimiento. En el caso de que el usuario no confirme lo reconocido, se anota esta circunstancia para la intervención posterior de un operador. Una vez conocida la ciudad, se le pregunta al usuario si desea un teléfono particular o de empresa.
- 2) Si la opción elegida es la de empresa, se le pide al usuario el nombre de la empresa, realizando un reconocimiento análogo al visto para el caso de la ciudad (pidiendo confirmación al usuario de los dos primeros candidatos).
- 3) En el caso de teléfono particular, se procede al reconocimiento del primer apellido, segundo apellido y nombre, en turnos de diálogo independientes.
- 4) Si ha habido algún reconocimiento no confirmado por parte del usuario se le ofrece al operador un cuadro de diálogo donde se le permite escuchar lo dicho por el usuario y rellenar un cuadro de texto con el dato correcto. En ningún momento hay comunicación directa con el cliente de modo que el paso a operador es transparente para el usuario. Una vez obtenidos los datos correctos, se accede a la base de datos y se proporciona el teléfono solicitado.

Con la posibilidad de intervención de un operador se garantiza que la totalidad de las consultas son atendidas correctamente, condición indispensable en un sistema real. Lo que variará será la tasa de intervención del operador dependiendo, fundamentalmente, de la calidad del reconocimiento.

### IV.- NUEVAS PRIMITIVAS DEL LENGUAJE

El reto más importante para la realización de esta aplicación ha sido la implementación de las nuevas primitivas del lenguaje necesarias para dotarle de la nueva funcionalidad.

#### IV.1. Reconocimiento de gran vocabulario.

El sistema de partida ofrecía el reconocimiento de los dígitos y de algunas palabras de control como SÍ, NO, AYUDA y CANCELAR. Para ofrecer el reconocimiento de gran vocabulario se han añadido cuatro primitivas que pasamos a comentar a continuación:

1. RECONOCIMIENTO\_DICC\_ESTATICO: esta función realiza el reconocimiento de la palabra pronunciada por el usuario de entre las pertenecientes a un diccionario determinado. Como resultado del reconocimiento se ofrece una lista ordenada de candidatos. El reconocimiento se realiza en dos etapas: hipótesis (preselección) y verificación (reconocimiento más detallado).
  - *Hipótesis* [4][5]: para la implementación de esta etapa del reconocimiento se ha utilizado el módulo de preselección desarrollado en el proyecto TIDAISL en el que se han introducido importantes mejoras: la primera de ellas ha consistido en la introducción de un módulo de parametrización propio (no disponible en TIDAISL) con resultados mejores a los obtenidos en dicho proyecto. En nuestro caso utilizamos RASTA-LPC con filtros PLP de acuerdo a la escala MEL. Otra mejora importante ha sido la ampliación del alfabeto de alófonos considerado. Este nuevo alfabeto tiene en cuenta una mayor variedad de alófonos para las vocales, considera la diferenciación de los dos alófonos (oclusivo y fricativo) para los fonemas [b], [d] y [g] e introduce un alófono adicional para el caso del fonema [n].
  - *Verificación*: en esta etapa se escoge un número reducido de candidatos (10%-15% del tamaño del diccionario) que obtuvieron un menor coste en la etapa anterior, y se realiza un proceso de reconocimiento, basado en el algoritmo de Viterbi, utilizando modelos más potentes, modelos de palabra HMM-continuos (resultado de la concatenación de modelos de alófonos). En esta etapa se ha realizado un esfuerzo importante en el entrenamiento de los modelos HMM-continuos siguiendo las pautas establecidas en [6] y [7]. El entrenamiento utilizado finalmente, permite la obtención de

modelos con un número diferente de mezclas en cada estado. Estas mezclas son de tipo gaussiano y su número depende, para cada iteración del algoritmo, del número de vectores pertenecientes a ese estado, obtenido en la fase de alineamiento previa.

En la implementación del algoritmo de Viterbi para la obtención de la palabra más probable, se han desarrollado técnicas de reducción del tiempo de proceso. La primera de ellas ha consistido en la introducción de una etapa de preproceso, anterior al algoritmo de reconocimiento. En esta etapa se realizan una serie de cálculos generales al proceso de reconocimiento cuyos resultados se almacenan en memoria de forma que dichos cálculos no se repitan posteriormente. Mediante esta técnica se ha conseguido reducir en un orden de magnitud el tiempo de proceso. Otras técnicas utilizadas y que han permitido otra reducción importante han sido la utilización de restricciones sobre el camino óptimo [8] y la técnica de reducción del haz de búsqueda (*Beam Search* [9]).

La finalidad de utilizar estas dos etapas en nuestro sistema es la de acelerar el reconocimiento. Esta característica es más importante cuanto mayor sea el tamaño del diccionario a tratar.

Otra funcionalidad importante que se ha introducido en esta primitiva de reconocimiento ha sido la detección de homófonos. Los homófonos son palabras que tienen la misma pronunciación pero se escriben de forma diferente (Ej: Esther y Ester). Nuestro sistema de reconocimiento, a través de esta primitiva, es capaz de detectar los homófonos reconocidos en primer y segundo lugar, de forma que mediante el desarrollo de una sencilla rutina en el lenguaje de alto nivel, establecemos un pequeño diálogo con el usuario que nos permite desambiguar y obtener la palabra deseada.

La base de datos utilizada para entrenar los modelos HMM (tanto HMM-SC como HMM-C) ha sido SPEECH-DAT [10] (<http://www.upc.es/castella/recerca/recerca.htm>).

2. RECONOCIMIENTO\_DICC\_DINAMICO: esta función realiza la misma labor que la comentada anteriormente con la peculiaridad de que el reconocimiento no se realiza sobre todo el diccionario de palabras, sino solamente sobre una lista de candidatos propuesta (subdiccionario dinámico) lo cual mejora la tasa de reconocimiento. El objetivo es disponer de diccionarios dinámicos que puedan ir variando según las necesidades del diálogo de la aplicación.
3. CARGAR\_DICCIONARIO Y DESCARGAR\_DICCIONARIO: estas dos funciones nos permiten cargar y descargar en memoria, uno o varios diccionarios, de modo dinámico, durante la ejecución de la aplicación. Existe un número máximo de diccionarios que pueden estar cargados simultáneamente y que es configurable por el desarrollador de aplicaciones.

#### IV.2. Intervención de un operador.

La posibilidad de intervención de un operador ha requerido añadir dos nuevas primitivas al lenguaje:

1. OPERADOR\_DATO: esta función permite presentar en pantalla un cuadro de diálogo con las opciones necesarias para la intervención del operador. En este cuadro se permite al operador escuchar lo dicho por el usuario y se le facilita un recuadro de texto donde puede escribir la palabra correcta.
2. ESPERAR\_OPERADOR\_DATO: la misión de esta función es suspender el curso normal de la aplicación hasta que el operador cierre el cuadro de diálogo. De esta forma, garantizamos que la intervención del operador se lleva a cabo y disponemos siempre de un dato correcto.

La intervención del operador es totalmente transparente al usuario. El usuario únicamente percibe un pequeño retraso en la obtención de su número de teléfono que puede ser fácilmente atribuible al proceso de búsqueda en la base de datos. De esta forma un mismo operador puede atender a varias líneas telefónicas.

#### V.- ADAPTACIÓN DEL SISTEMA A UN HARDWARE COMERCIAL

Otro desarrollo importante ha sido la adaptación del entorno a un hardware comercial. Inicialmente este sistema utilizaba dos tarjetas desarrolladas en nuestro propio grupo de trabajo: VISHA[11] e IFTEL. La tarjeta VISHA es una tarjeta de sonido con un DSP32C capaz de realizar parte de las tareas de procesamiento de voz, lo que permite al ordenador realizar otras tareas en paralelo. Por otro lado la tarjeta IFTEL es la encargada de la interfaz y adaptación de señales entre la línea telefónica y la tarjeta de sonido. El cambio ha

consistido en sustituir la tarjeta VISHA por un dispositivo WAVE (Sound Blaster 16) [12]. Este cambio ha supuesto que las tareas de análisis de la señal de voz que anteriormente realizaba la VISHA, ahora las deba realizar el procesador del ordenador. Para permitir que el ordenador pueda procesar las muestras al mismo tiempo que las va capturando a través de la tarjeta de sonido, hacemos uso de la multitarea que nos ofrece el sistema operativo Windows'95. En esta fase de adaptación ha sido necesario instalar un segundo dispositivo de audio WAVE para permitir al operador escuchar los ficheros grabados al mismo tiempo que se ejecuta la aplicación.

## VI.- EXPERIMENTOS

Para evaluar la tasa de reconocimiento obtenida en condiciones reales, se ha realizado una primera evaluación del sistema con posibles usuarios finales. Cada uno de los usuarios debía acceder a 10 teléfonos de empresas y 10 teléfonos de particulares. Las pruebas las realizaron 33 personas (21 hombres y 12 mujeres). En la siguiente tabla se ofrece la tasa de reconocimiento para cada diccionario (ciudades, empresas, apellidos y nombres) en primer (1 candidato) y segundo (2 candidato) lugar, junto con la tasa de no reconocidos (No reconocido). Todos los diccionarios utilizados tienen un tamaño de 1000 palabras.

	<b>Ciudades</b>	<b>Empresas</b>	<b>Apellidos</b>	<b>Nombres</b>	<b>TOTAL</b>
<b>1 candidato</b>	75,9%	85,6%	66,6%	67,9%	74,0%
<b>2 candidato</b>	7,3%	4,6%	7,9%	9,2%	7,3%
<b>No reconocido</b>	16,8%	9,8%	25,5%	22,9%	18,7%

Tabla 1. Tasa de reconocimiento (primer y segundo candidato) y de error.

Como se puede observar la tasa de reconocimiento para el diccionario de empresas es superior al resto. Este hecho se debe a la mayor longitud media de los nombres de empresas (14,3 letras) frente a las palabras del resto de diccionarios, circunstancia que permite a nuestro reconocedor una mejor discriminación. La peor tasa se obtuvo para los nombres propios y apellidos, coincidiendo también con una menor longitud media de palabra (7,2 y 6,2 letras respectivamente). Las ciudades con una longitud media de 7,9 letras se sitúan en una posición intermedia en tasa de reconocimiento. Otro factor importante de las diferentes tasas para cada diccionario es la confusión entre sus palabras, por ejemplo, Fernández y Hernández son palabras muy parecidas puesto que se diferencian en un sólo alófono. Mediante un algoritmo de programación dinámica hemos calculado la distancia alofónica media de cada palabra con la más parecida dentro del mismo diccionario, obteniendo 2,0 y 2,3 para apellidos y nombres respectivamente y, 3,3 y 6,8 para los casos de ciudades y empresas. Como se puede apreciar las peores tasas corresponden a los diccionarios con menor distancia media y por tanto mayor parecido.

El porcentaje de consultas completadas con éxito (el usuario accedió al número de teléfono correcto) para el caso de empresas fue del 75% mientras que en el caso de teléfono particular fue del 35%. Esta diferencia tiene su explicación en que para acceder a un teléfono de empresas es necesario realizar dos reconocimientos (ciudad y empresa) mientras que para conseguir el teléfono de un particular son necesarios cuatro (ciudad, primer apellido, segundo apellido y nombre). Esta fase de pruebas tenía el objetivo de evaluar el sistema de reconocimiento; por esta razón se desactivó la generación de diccionarios dinámicos en el transcurso de la aplicación. Es decir, que aunque se obtenga correctamente la ciudad donde está localizada una empresa, a la hora de reconocer el nombre de dicha empresa, al sistema de reconocimiento se le ofrece el diccionario completo de empresas y no una lista reducida con las que existen en esa ciudad. Por esta razón, los resultados de tasas de llamadas completadas con éxito son los obtenidos en las condiciones de mayor perplejidad y deben ser interpretados como un límite inferior de la tasa obtenible.

## VII.- CONCLUSIÓN

El demostrador desarrollado funciona en tiempo real en un ordenador Pentium II-350Mhz. Según la evaluación realizada, la tasa de intervención de operador en las peores condiciones es menor al 45%, lo que nos da una tasa mínima de 55% de llamadas procesadas automáticamente (tasa que irá mejorando a lo largo del proyecto). Podemos concluir que las tecnologías del habla son un campo de investigación de gran interés

comercial y que los *Servidores Vocales Interactivos* son una solución interesante para el camino de automatización y abaratamiento de los servicios por línea telefónica.

## VIII.- LÍNEAS DE INVESTIGACIÓN FUTURAS

Una primera línea de investigación, que está implícita en el propio objetivo del proyecto IDAS, es el aumento del tamaño de los diccionarios utilizados. Este aumento requiere dos esfuerzos importantes: por un lado, en los algoritmos de entrenamiento para obtener unos modelos más potentes siguiendo la línea de estudios anteriores [8], y por otro lado, la optimización de los algoritmos de reconocimiento para permitir el funcionamiento del sistema en tiempo real. Otro reto importante dentro del marco del proyecto es el desarrollo de una nueva primitiva que nos permita realizar reconocimiento de nombres deletreados. Ello nos ofrecerá la posibilidad de solicitar al usuario el deletreo del nombre cuando su reconocimiento falle.

En la línea de adaptar el sistema a un hardware comercial nos planteamos utilizar una sola tarjeta para la captura de muestras y la gestión de la línea telefónica. La tarjeta elegida será alguna de las ofrecidas por la compañía Dialogic (<http://www.dialogic.com>), de forma que sea compatible con el estándar WAVE para la captura de muestras y con el estándar TAPI para la gestión de la línea telefónica.

Una línea de investigación que se planteará al final de este proyecto y que será clave en proyectos venideros será dotar al sistema de módulos de comprensión y diálogo que den una mayor flexibilidad a la comunicación hombre-máquina. En este punto cabe comentar los estudios realizados en [9].

## IX.- REFERENCIAS

- [1] A. Casas Guijarro: "Sistema telefónico multilínea con reconocimiento de voz y acceso a base de datos remota". Proyecto Fin de Carrera. Grupo de Tecnología del Habla. DIE. UPM. Madrid, 1997.
- [2] J.M. Pardo et al "Spanish text to speech: from prosody to acoustic" International Conference on Acoustic 95 vol III, 1995.
- [3] J.Macías-Guarasa, A. Gallardo, J. Ferreiros. "Módulo de búsqueda rápida para el reconocedor de habla aislada y grandes vocabularios". Informe interno. Grupo de Tecnología del Habla. DIE. UPM. Madrid, 1996.
- [4] J.Ferreiros, J. Macías-Guarasa, A. Gallardo, J. Colás, R. Cordoba, J.M. Pardo. and L. Villarrubia "Recent Work on Preselection Module for Flexible Large Vocabulary Speech Recognition System in Telephone Enviroment", ICSLP'98, Sidney (Australia), Nov. 98.
- [5] J. Macías-Guarasa, A. Gallardo, J.Ferreiros, J.M. Pardo. and L. Villarrubia. "Initial Evaluation of a Preselection Module for a Flexible Large Vocabulary Speech Recognition System in Telephone Enviroment". ICSLP'96, pag 1343-1346. Philadelphia (USA), Oct. 96.
- [6] A. López. "Modelos continuos de Markov para reconocimiento de habla continua". Proyecto Fin de Carrera. Grupo de Tecnología del Habla. DIE. UPM. Madrid, 1997.
- [7] R. Córdoba. "Sistemas de reconocimiento de habla continua y aislada: comparación y optimización de los sistemas de modelado y parametrización". Tesis Doctoral. Grupo de Tecnología del Habla. DIE. UPM. Madrid, 1995.
- [8] J. Ferreiros. "Aportación a los métodos de entrenamiento de Markov para reconocimiento de habla continua.". Tesis Doctoral. Grupo de Tecnología del Habla. DIE. UPM. Madrid, 1996.
- [9] J. Colás. "Estrategias de Incorporación de Conocimiento Sintáctico y Semántico en Sistemas de Comprensión de Habla Continua en Castellano". Tesis Doctoral. Grupo de Tecnología del Habla. DIE. UPM. Madrid, 1999.
- [10] A. Moreno, H. Hvge, J. Koehler, J.B. Mariqo First. "SpeechDat across Latin America. Project SALA". LREC'98, Granada (Spain). 28-30 May 1998.
- [11] S. Aguilera, M.A. Berrojo, F.M. Giménez de los Galanes, J. Colás, J. Macías-Guarasa, J.M. Montero. "Impaired Persons Facilities Based on a Multi-Modality Speech Processing Sustum". Proceedings of the International Conference on Speech and Language Technology for Disabled Persons, pp 129-132. ISSN 1018-4554. Stocholm, May-June 1993.
- [12] R. San-Segundo: "Optimización de un sistema de reconocimiento de habla aislada por teléfono sobre un ordenador compatible (PC)". Proyecto Fin de Carrera. Grupo de Tecnología del Habla. DIE. UPM. Madrid, 1997.