DESARROLLO DE UN SISTEMA DE DIÁLOGO ORAL EN DOMINIOS RESTRINGIDOS

Antonio Bonafonte[1], Pablo Aibar[2], Núria Castell[1], Eduardo Lleida[3],
José B. Mariño[1], Emilio Sanchis[4] y M. Inés Torres[5]

[1] Centro de Investigación TALP, Universitat Politècnica de Catalunya (antonio@talp.upc.es).
[2] Dept. de Informática, Universitat Jaume I.
[3] Dept. de Electrónica y Comunicaciones, Universidad de Zaragoza.
[4] Dept. LSI, Universidad Politécnica de Valencia.
[5] Dept. de Electricidad y Electrónica, Universidad del País Vasco (UPV/EHU).

RESUMEN

En esta comunicación se describe el proyecto titulado "Desarrollo de un Sistema de Diálogo Oral en Dominios Restringidos"; proyecto financiado por la CICYT que empezó en octubre de 1998 y terminará en septiembre del 2001. El objetivo de dicho proyecto es investigar metodologías y tecnología para desarrollar interfaces orales. La tarea que se ha escogido como sustento de dicha investigación es la de consulta sobre horarios y precios de trenes regionales y grandes líneas. El objetivo del proyecto se concretará mediante la implementación de un sistema prototipo que ofrezca dicha información.

En este momento, en el proyecto se han producido dos corpora. En el primero de ellos (corpus persona-persona) se han adquirido y transcrito 200 conversaciones de usuarios que llamaban al centro de información de RENFE. Este corpus se utilizó para analizar el dominio semántico de la tarea y sirvió de referencia al definir las estrategias de diálogo y generación de respuesta. También se utilizo como referencia al generar escenarios (situaciones) que pueden utilizarse al solicitar llamadas de colaboradores voluntarios al sistema. En el segundo corpus (OZ1) se han adquirido 225 diálogos siguiendo el paradigma del Mago de Oz, utilizados en la inferencias de modelos y reglas a distintos niveles: modelos de lenguaje para reconocimiento y para comprensión, modelo probabilístico de actos de diálogos, etc.

Dado que las prestaciones de los sistemas de reconocimiento empeoran sustancialmente al reconocer la llamada habla espontánea (frente a lectura), una de las actividades del proyecto se ha centrado en el análisis, etiquetado, modelado y tratamiento de los así denominados fenómenos de habla espontánea. Se ha realizado un etiquetado exhaustivo del corpus OZ1 y se ha definido un esquema de anotación. Actualmente se está estudiando el modelado –acústico, léxico, sintáctico- de algunos de los fenómenos etiquetados con objeto de mejorar las tasas de reconocimiento.

En el proyecto se está implementado un sistema distribuido, donde cada desarrollador de un módulo lo activa en forma de servidor. Esta arquitectura flexible facilita la colaboración entre los distintos grupos de trabajos y facilita el test de varias alternativas para cada módulo, disponer de un prototipo con la última versión de cada uno de los módulos, etc. En la comunicación se describen brevemente los componentes del sistema: servidor de audio, gestor de la aplicación, reconocimiento del habla, comprensión del habla, gestor del diálogo, generación de respuesta y conversión de texto a habla.

DESARROLLO DE UN SISTEMA DE DIÁLOGO ORAL EN DOMINIOS RESTRINGIDOS

Antonio Bonafonte[1], Pablo Aibar[2], Núria Castell[1], Eduardo Lleida[3],
José B. Mariño[1], Emilio Sanchis[4] y M. Inés Torres[5]

[1] Centro de Investigación TALP, Universitat Politècnica de Catalunya (antonio@talp.upc.es).
[2] Dept. de Informática, Universitat Jaume I.
[3] Dept. de Electrónica y Comunicaciones, Universidad de Zaragoza.
[4] Dept. LSI, Universidad Politécnica de Valencia.
[5] Dept. de Electricidad y Electrónica, Universidad del País Vasco (UPV/EHU).

1. INTRODUCCIÓN

En esta comunicación se describe un proyecto de diálogo oral que comenzó en octubre de 1998 y finalizará en septiembre del 2001. El proyecto está siendo realizado por los grupos de investigación del área del tratamiento del habla y del lenguaje que firman este articulo. En [BASURDE] puede encontrarse información sobre cada uno de los grupos que integran el proyecto, así como información para contactar con los coordinadores del proyecto en cada grupo.

En esta comunicación se presentará una visión general del proyecto en su conjunto presentándose los corpora adquiridos, un análisis de los fenómenos de habla espontánea que aparecen, la arquitectura del sistema y una breve descripción de cada componente.

2. CORPORA

2.1 Elección de la tarea

El consorcio formado para este proyecto ya había colaborado en proyectos de reconocimiento del habla, pero este es el primer trabajo en el área del diálogo oral. Para aproximarse al problema, el proyecto se ha centrado en una tarea concreta: un sistema de información telefónica de horarios y precios de trenes regionales y de grandes líneas. Esta tarea ha sido la elegida por varios grupos europeos, entre ellos en el marco de los proyectos TABA [TABA] y ARISE [ARISE]. Presenta algunas características que la hacen preferible a otras tareas, ya que permite centrar los esfuerzos en el tema al que fundamentalmente se dedica el proyecto: el diálogo.

- Es una tarea bien limitada semánticamente, lo que facilita la comprensión respecto a otras tareas más abiertas.
- El vocabulario es relativamente estable y regular: no aparecen palabras extranjeras como sería el caso de información sobre viajes de aviones o información sobre espectáculos.
- Presenta una estructura de diálogo relativamente rica (medida en turnos de interacción entre usuario y sistema).
- Al tratarse de información telefónica el interfaz es unimodal.

Muchas de las metodologías que se deseaban investigar eran basadas en técnicas de inferencia estadísticas. Ello requiere la existencia de un corpus de la tarea a modelar. En el proyecto se han adquirido dos corpora: un corpus persona-persona y un corpus adquirido bajo el paradigma del Mago de Oz.

2.2 Corpus persona-persona.

El corpus persona consta de las transcripciones ortográficas de 200 diálogos de usuarios que solicitaban información del servicio de RENFE. No son, por tanto, simulaciones a consultas sino consultas reales. La adquisición y transcripción se realizó en los primeros meses del proyecto y no se impuso ningún otro criterio en su selección salvo que la llamada se cursara en castellano (fueron adquiridas en Barcelona) y que no utilizara el servicio de reserva (debido a la confidencialidad de los datos). Se procuró además adquirir una llamadas variadas tanto en lo que respecta al tema de la llamada (viajes simples, viajes con enlace, se conocen horarios y se piden precios, viajes en grupos, descuentos, bonos, etc.) como en cuando a la forma en la que se desarrollaba la llamada y –relacionado con esto- el operador que atendía la llamada.

El resultado de la adquisición es un corpus muy rico y complejo. Se aprecian todos los fenómenos de habla espontánea que han sido observados en otros

trabajos: rellenos, reformulaciones, ofrecimientos implícitos del turno, sutiles solicitudes del turno mientras uno de los interlocutores está hablando, interrupciones, etc.

Como ya era de esperar, incluso en situaciones tan sencillas como la solicitud de información entre dos personas que no se conocen, la expresión verbal es muy compleja, más de lo que nuestros sistemas de reconocimiento y comprensión son capaces de tratar actualmente.

El corpus persona-persona se utilizó para las siguientes tareas. En primer lugar para definir la semántica de la tarea de información y delimitar lo que se quería tratar. La decisión, como ya se ha comentado, fue la de dar información sobre horarios y precios de trenes regionales y de grandes líneas. No se han incluido trenes de cercanías por dos motivos: la complicación léxica y el hecho de que los diálogos sobre trenes de cercanías tienen habitualmente una estructura muy simple. Se contemplan viajes de ida y vuelta, clases preferente y turista, litera, coche-cama. No se contemplan en esta primera aproximación otros servicios tales como equipamiento de las estaciones, servicios de restaurantes, abonos, reservas, etc.

El corpus persona-persona también fue utilizado como guía al definir la estrategia de diálogo y la generación de la respuesta oral. Fue además útil en la definición de situaciones –escenarios- a utilizar en llamadas al sistema de informantes a los que se les pide que accedan al servicio automático de información.

2.3 Corpus Mago de Oz (OZ1)

Como ya se ha comentado, el corpus persona-persona presenta un léxico y unas estructuras lingüísticas muy complejas. Es muy previsible que ese tipo de comunicación, fluida y coloquial, no se presente en cuanto se trate con una máquina. Además, al utilizar un sistema automático, se dan situaciones nuevas que no son contempladas en los diálogos persona-persona y que provocarán por tanto reacciones distintas en el llamante; por ejemplo, errores graves del sistema de comprensión. Por tanto, el persona-persona no es adecuado para inferir modelos estadísticos del usuario.

El paradigma del Mago de Oz busca recoger muestras de intervenciones del usuario con el sistema antes de disponer del sistema. Para ello, una persona (el Mago de Oz) actúa como sistema, caracterizando de la forma mejor posible el sistema que se desea construir. Si el mago actúa bien, la persona que llama creerá efectivamente que es un sistema y actuará en consecuencia. Para facilitar el trabajo al Mago de Oz, es recomendable incluir la parte del sistema que se disponga. Por ejemplo, si se dispone de un sistema de reconocimiento, este debe incluirse para que el Mago de Oz no necesite simular los errores de dicho sistema de reconocimiento.

Para adquirir el corpus OZ1, en el proyecto se ha desarrollado la plataforma HAL. Dicha plataforma consiste de un PC conectado a una línea telefónica digital (RDSI). En el momento de adquirirse el corpus no se disponía de un sistema de reconocimiento con prestaciones suficientes (fundamentalmente debido al modelo del lenguaje), por lo que la señal telefónica recibida por HAL se reproducía por el auricular del Mago de Oz. Tampoco se disponía de sistema de comprensión, pero si que se definió, antes de empezar las adquisiciones, la estrategia de diálogo precisa que se había de seguir. La estrategia estaba definida en forma de reglas del tipo:

> *Si se formula una pregunta sobre horarios o precios y no se conoce el destino, preguntar la ciudad de destino.*
>
> *Si el resultado de la consulta son más de tres trenes, decir el número de trenes y las características del primer y último tren.*

Un aspecto importante de la verosimilitud de la actuación del mago está en la voz producida. Por ello HAL generaba dicha voz utilizando un sistema de conversión de texto a voz. El inconveniente principal de este método es el retardo que se produce debido al tiempo que necesita el mago para escribir los mensajes. Para reducir este tiempo HAL incorpora una herramienta independiente, un generador de respuesta [SESMA], que permite seleccionar rápidamente entre un conjunto de plantillas debidamente clasificadas y dar valor a variables que aparecen en dicha plantilla. Por ejemplo, se puede seleccionar la plantilla

> *¿Qué día le interesa viajar a <CIUDAD-DESTINO>?*

y a continuación seleccionar entre los posibles, el valor que toma la variable <CIUDAD-DESTINO>. Además de reducirse el tiempo de generación de respuesta (que sigue siendo elevado), el hecho de tener definidas las posibles respuestas ayuda al Mago de Oz a seguir la estrategia de diálogo definida. Para formular las respuestas, el Mago disponía del software InfoTren, cedido por RENFE.

En la generación de un corpus de diálogo, bien sea con un sistema automático, bien mediante un sistema Mago de Oz, un aspecto interesante es cómo se logran usuarios que llamen al sistema. Idealmente, el sistema debería conectarse a un punto de servicio real; sin embargo esto no es siempre fácil de conseguir. En el proyecto se ha solicitado la participación de 75 informantes que habían de conseguir cierta información de acuerdo a unos escenarios diseñados previamente. Cada informante debía completar tres escenarios. En dos de ellos el objetivo estaba perfectamente definido –por ejemplo, horarios del primer tren de Castellón a Valencia- aunque era presentado mediante unos guiones. El tercero era más abierto, indicando lo que se espera pero sin concretar fechas ni ciudades, etc. Los informantes al terminar rellenaron un cuestionario en el que valoraban la experiencia y los distintos subsistema (por ejemplo la fluidez del diálogo, relacionado con la estrategia, o la calidad de la voz generada).

El corpus Oz1 contiene, para cada una de las 225 sesiones, las señales adquiridas (un canal para la entrada y uno para la salida), la segmentación entre turnos y el texto asociado a cada turno con algunos símbolos para indicar algunos fenómenos acústicos concretos: palabras mal pronunciadas, truncadas, ruido del locutor o externo, etc. Posteriormente se han realizado otras transcripciones de las mismas señales como son: transcripción semántica, transcripción en actos de diálogos, transcripción con anotación detallada de las disfluencias.

3. HABLA ESPONTÁNEA

Es bien conocido que las prestaciones de los sistemas de reconocimiento empeoran sustancialmente al reconocer habla espontánea frente a habla leída. Por ello, una de las actividades del proyecto se ha centrado en el análisis, etiquetado, modelado y tratamiento de los así denominados fenómenos de habla espontánea. Durante el análisis preliminar de un grupo reducido de diálogos se detectaron fenómenos muy variados: silencios amplios en mitad de un turno, ruidos diversos que en general son puntuales, alargamientos de vocales y también de algunas consonantes (especialmente /m/, /n/, /l/ y /s/), pausas habladas (vocalizaciones o nasalizaciones que marcan una pausa), palabras pronunciadas parcialmente y particularmente lo que se conoce como disfluencias sintácticas: repeticiones, sustituciones e inserciones que rompen el discurso y lo hacen gramaticalmente incorrecto. Al mismo tiempo se realizó una revisión exhaustiva de diversos formatos y herramientas de anotación lingüística [Rodríguez99], buscando aquéllos que integraran los distintos niveles de anotación necesarios en tareas de diálogo: ortográfico, fonético, prosódico, morfosintáctico, semántico y de actos de diálogo. De esta búsqueda se extrajeron dos conclusiones: una, que resulta muy difícil integrar todos los niveles en un mismo esquema de anotación; y dos, que los esfuerzos mejor encaminados en este sentido parecen los que se han iniciado con el proyecto europeo MATE [MATE], basado en el uso de XML [OASIS]. Finalmente, tras catalogar los fenómenos de habla espontánea más significativos, se definió un formato de anotación tipo XML [Rodríguez00]. El etiquetado ha sido realizado en fechas recientes, y ha estado a cargo de tres personas. Las anotaciones han sido supervisadas posteriormente por una sola persona, de modo que se asegura un alto grado de coherencia. Actualmente se está estudiando el modelado - acústico, léxico, sintáctico- de algunos de los fenómenos etiquetados, con objeto de mejorar las tasas de reconocimiento.

4. ARQUITECTURA DEL SISTEMA

En el desarrollo del proyecto se esperan desarrollar métodos, algoritmos y recursos para las tecnologías que se utilizan en sistemas de diálogo oral: reconocimiento del habla espontánea, técnicas de comprensión del habla, gestión de diálogo, generación de respuesta oral, etc. No obstante hemos considerado que un método idóneo tanto para el desarrollo como la evaluación de dichas tecnologías es el desarrollo de un prototipo de sistema automático de diálogo oral. El objetivo fijado es disponer de una primera versión integrada en diciembre del 2000.

Se ha adoptado una arquitectura distribuida, cliente servidor, donde los distintos componentes del sistema se ejecutan (potencialmente) en máquinas distintas. Mediante la definición de los interfaces entre módulos y del un protocolo de comunicación se consiguen varias ventajas. En primer lugar se simplifica la integración del sistema. Cada componente es ensamblado directamente por el grupo que lo desarrolla, sin interferir con el resto de componentes ni presentar incompatibilidades de software, sistema operativo o plataforma. Además, es sencillo desarrollar en paralelo varias versiones de cada componente con distintas metodologías, que pueden compararse en igualdad de condiciones y con la mejor versión del resto del sistema. Finalmente, se relajan las condiciones en cuanto a coste computacional. Al ser posible instalar componentes distintos en máquinas

distintas es más fácil conseguir versiones de desarrollo con tiempo de respuesta cortos.

Un posible problema que presenta esta configuración es los retrasos que puedan surgir debido a congestión en la red de comunicaciones que une los distintos componentes. Para aliviar el problema, en el diseño de la arquitectura se prevé instalar los componentes que tratan directamente con muestras de señal de voz (extracción de características del reconocedor y sintetizador del habla) en la misma máquina que el servidor de audio o en máquinas en las que el acceso sea rápido. Para el resto de los componentes el intercambio de datos es textual y no se espera que presente problemas importantes.

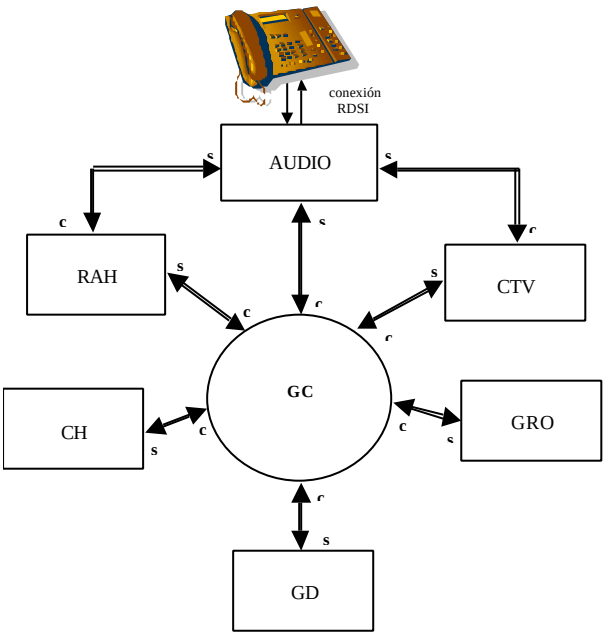

Figura 1 Diagrama de bloques de la arquitectura del sistema. RAH: sistema reconocimiento automático del habla, CH: sistema de comprensión del habla, GD: sistema de gestión del diálogo, GRO: sistema de generación de la respuesta oral, CTV: sistema de conversión texto-voz. Las letras "c" y "s" hacen referencia a la función del socket. **c**: cliente, **s**:servidor.

5. COMPONENTES DEL SISTEMA

5.1 Audio server

El servidor de audio es el que gestiona la conexión del sistema con la línea telefónica. Además, para reducir la carga sobre la red al comunicarse con el módulo de reconocimiento, calcula unos parámetros de la señal de voz cada 10 milisegundos. En la versión actual, dichos parámetros constan de la media del valor absoluto de las muestras (para la detección de voz/silencio) y de los coeficientes cesptrum, calculados a partir de bancos de filtros frecuenciales escalados MEL. Actualmente se están investigado parametrizaciones alternativas basadas en las derivadas frecuenciales de las energías logarítmicas de 14 bandas espectrales, por su mejor robustez al ruido en comparación con los clásicos coeficientes cepstrales [Nadeu00]

5.2 Sistema de Reconocimiento del Habla

El sistema de reconocimiento recibe los parámetros de la señal de voz y los transforma en las palabras que han sido pronunciadas con mayor probabilidad. También informa al gestor de la aplicación de las detecciones de voz/silencio y silencio/voz. Varios grupos que participan en el proyecto disponen de sistemas de reconocimiento competitivos que se integrarán a lo largo del proyecto al sistema de diálogo. En este apartado, no obstante, se presenta el sistema de reconocimiento que se utiliza en la primera versión del prototipo.

En lo que se refiere al Modelado Acústico, se utilizan modelos ocultos de Markov, semicontinuos, estimados a partir de la base de datos SpeechDat [SPEECHDAT]. La unidad subléxica utilizada son los semifonemas ligados utilizando árboles de decisión y clustering aglomerativo [Mariño00].

En cuanto al modelo del lenguaje, se han utilizado x-gramas [Bonafonte96] (n-gramas de longitud variable) estimados a partir de las frases pronunciadas por los usuarios en el corpus OZ1. Se utiliza un único modelo del lenguaje general durante todo el diálogo. Para paliar el pobre entrenamiento se han utilizado clases de palabras: se ha inferido el x-grama no directamente sobre las palabras, sino jerárquicamente, sobre clases de equivalencia. Las frases del usuario se han transformado utilizando clases de equivalencia según criterios sintácticos y semánticos, estimandose un x-grama sobre esas clases de equivalencia. El modelado de cada clase de equivalencia depende de su complejidad. Para clases sencillas como palabras o locuciones, se ha utilizado un autómata de estados finitos (aceptor canónico). Para las clases más complejas (por ejemplo *fechas*) se ha utilizado un x-grama, bien directamente en palabras, bien utilizando a su vez otras clases de equivalencia (por ejemplo, *meses*).

Cada palabra reconocida va acompañada de un parámetro que mide la confianza de que la palabra reconocida sea la que ha dicho el usuario o, por el contrario, se trate de un error del sistema de reconocimiento. Para asignar este valor se compara componente acústica del coste que el reconocedor asigna a esa palabra con el coste que le asigna una red de fonemas. Se están valorando las prestaciones que ofrece la inclusión de la información contextual propuesta en [Hernández99]

5.3 Sistema de Comprensión

El módulo de comprensión proporciona una representación semántica de la salida del reconocedor. La representación semántica escogida se basa en el concepto de *frame*. Un *frame* es una plantilla que resume la intervención del usuario mediante una agrupación etiquetada de atributos. Por ejemplo, el turno "¿Me puede decir los horarios de los trenes de Barcelona a Sevilla?" se representa del siguiente modo:

```
(HORA-SALIDA)
   CIUDAD-ORIGEN: Barcelona
   CIUDAD-DESTINO: Sevilla
```

Se ha definido un conjunto de *frames* para esta tarea, de modo que se puedan representar todas las intervenciones del usuario, no sólo las específicas de consulta, sino también las intervenciones de apertura, cierre, afirmación, confirmación, etc.

En el proyecto se están desarrollando dos módulos de comprensión siguiendo metodologías distintas. Uno de ellos [Sanchis00] se basa en la aplicación de técnicas de aprendizaje automático de modelos estocásticos. El proceso de comprensión se realiza en dos etapas. En la primera se obtiene de forma automática un traductor estocástico que proporciona una interpretación de la frase de entrada en términos de un lenguaje intermedio cercano a la estructura del *frame*. En la segunda etapa se obtiene el correspondiente *frame*. El segundo módulo de comprensión [Arranz00] está basado en un procesamiento lingüístico de cada intervención del usuario dividido en tres etapas. En primer lugar, el turno ya transcrito es analizado y desambiguado morfológicamente. El resultado de este proceso es analizado sintácticamente obteniéndose un árbol de análisis parcial. En la tercera etapa, se aplican las reglas de extracción semántica al resultado del análisis sintáctico, obteniéndose los *frames* que resumen la intervención y que se envían al gestor de diálogo.

De la evaluación cualitativa de ambos módulos de comprensión se espera establecer las ventajas e inconvenientes de cada metodología y ver si es posible implementar una cooperación entre ellos.

5.4 Gestor del Diálogo

El gestor de diálogo utiliza los *frames* generados por el módulo de comprensión y, acorde a la estrategia de diálogo que implementa, decide la acción a tomar. Dicha acción habitualmente será, bien generar una respuesta para el llamante mediante una representación semántica (*frame* de generación), bien realizar una consulta a la base de datos, bien comunicar al gestor de la aplicación que cancele la comunicación con el usuario.

Al igual que ocurría con el módulo de comprensión se están desarrollando dos gestores de diálogo que se basan en metodologías distintas. En el primero de ellos se pretende aplicar técnicas de inferencia de traductores al problema del aprendizaje de la estrategia del diálogo (Martínez00). Para el aprendizaje del modelo de diálogo se debe disponer de un conjunto suficientemente grande de diálogos. Por ello se ha desarrollado un generador automático de diálogos a partir de los obtenidos mediante el Mago de Oz. El modelo de diálogo obtenido, proporciona el acto de diálogo más probable, a partir de los actos previos y de su contenido semántico. Este acto de diálogo irá acompañado de la correspondiente información semántica para que el generador de respuestas pueda construir la frase a sintetizar.

El segundo gestor [Arranz00] está siendo implementado mediante un sistema de razonamiento terminológico. El conjunto de axiomas representa de manera formal la estrategia del gestor. Los *frames* del módulo de comprensión constituyen los hechos que el motor de razonamiento combina con los hechos del histórico del diálogo. De esta combinación se obtienen los hechos que marcan la reacción del sistema. Finalmente, el gestor enviará al generador de respuesta oral el/los *frame(s)* que contienen la información necesaria para elaborar la frase del siguiente turno del sistema.

5.5 Generación de respuesta.

La generación de la respuesta oral tiene dos fases claramente diferenciadas: el generador del texto y el conversor de texto a voz. El generador de texto, transforma el *frame* semántico en frases ortográficas. Debido a que el generador de texto tiene información semántica puede producir, además del texto,

información que pueda ser utilizada por el conversor de texto a voz. Por ejemplo, puede indicar que aquellas partes de la frase que presentan información nueva y han de ser enfatizadas o cuándo se requieren pausas para que el usuario pueda tomar nota.

El conversor de texto a voz [Bonafonte00] es un sistema de síntesis por concatenación basado en corpus en el que se disponen de muchos ejemplos de las unidades básicas y se seleccionan para concatenar aquellas cuya prosodia es más próxima a la indicada por el modelo prosódico y que presentan mejor concatenación. Para el proyecto se ha adquirido un corpus específico que contiene 200 frases incluyendo las ciudades y estaciones más importantes y que complementa al corpus general. De esta forma se dispone de un sistema versátil, en el que es posible cambiar las frases de respuesta sin necesidad de grabar *prompts*, y con una calidad próxima a frases pregrabadas. El sistema soporta muchas de las marcas de control definidas SABLE [SABLE] lo que facilita la utilización de información semántica para controlar rasgos prosódicos de la respuesta.

6. RECONOCIMIENTO Y AGRADECIMIENTOS

Agradecemos a Estaciones comerciales de RENFE, en particular a Armando Brigos gerente de la Gerencia Territorial del Nordeste y al personal de la Jefatura de Centros de Viaje, en particular a Jesús Martínez, la cesión de las grabaciones de llamadas anónimas a su servicio de información y su colaboración en la transferencia de los diálogos y en otros aspectos relacionados con el proyecto.

El consorcio agradece a la empresa Natural Vox, que actúa como Ente Promotor y Observador del proyecto, por su participación en las reuniones del proyecto.

Finalmente, el consorcio agradece a la CICYT por su contribución en la financiación del proyecto TIC98-0423-C06, así como a la CIRIT por su financiación a través del proyecto 1999SGR150.

7. REFERENCIAS

[ARISE] E. Den Os, L. Boves, L. Lamel, P. Baggia. Overview of the ARISE project. Proc. EUROSPEECH'99, pp. 1527-1530, Budapest, Hungría, Septiembre 1999.

[Arranz00] V. Arranz, N.Castell, M.Civit. "La comprension del dialogo en un sistema de habla espontanea". Primeras Jornadas de Tecnologia del Habla (1JTH), Sevilla, 2000.

[BASURDE] URL: gps-tsc.upc.es/veu/basurde/

[Bonafonte00] A. Bonafonte, José B. Mariño. "Language modeling using X-grams". Proc. ICSLP'96, pp. 394-397, Philadelphia, USA, Octubre 1996.

[Bonafonte00] A.Bonafonte y Albert Febrer. "Actividades en el área de conversión de texto a voz en el centro TALP". I Jornadas de Tecnologías del Habla, Sevilla, Noviembre 2000.

[Hernández00] G. Hernández Ábrego, J.B. Mariño, Contextual Confidence Measures for Continuous Speech Recognition. Proc. ICASSP'2000, Vol III, pp. 1803-1806, Istambul. June 2000.

[Mariño00] José B. Mariño, A. Nogueiras. "Modelado acústico-fonético mediante semifonemas para el reconocimiento del habla fluida". I Jornadas de Tecnologías del Habla, Sevilla, Noviembre 2000.

[Martínez00] C.Martínez, F.Casacuberta "A pattern recognition approach to dialog labelling using finite-state transducers " V Ibero American Sympusium on Pattern Recognition, pp.669-677, Sept, 2000.

[MATE] URL: mate.nis.sdu.dk

[Nadeu00] C. Nadeu, D. Macho, J. Hernando. "Time & frequency filtering of filter-bank energies for robust HMM speech recognition". Speech Communication, 2000.

[OASIS] URL: www.oasis-open.org/cover/xml.html00

[Rodríguez99] L.J. Rodríguez. "Anotación de corpora para diálogo". Documento interno: BS12AV02. Proyecto TIC98-0423-C06. Diciembre 1999.

[Rodríguez00] L.J. Rodríguez, I. Torres, A. Varona. "Manual para el etiquetado de disfluencias". Documento interno: BS12BV30. Proyecto TIC98-0423-C06. Mayo 2000.

[SABLE] URL: www.bell-labs.com/project/tts/sable.html

[Sanchis00] E.Sanchis, E.Segarra, M.Galiano, F.García, L.Hurtado "Modelización de la comprensión mediante técnicas de aprendizaje automático" Primeras Jornadas de Tecnologia del Habla (1JTH), Sevilla, 2000.

[SESMA] URL:gps-tsc.upc.es/veu/sesma/software.html

[SPEECHDAT] URL: www.speechdat.org

[TABA] H. Aust, M. Oerder, F. Seide, V. Steinbiss. "The Phillips automatic train timetable information system". Speech Communication 17, pp. 249-262, 1995.