

DIHANA: Sistema de diálogo para el acceso a la información mediante habla espontánea en diferentes entornos

José Miguel Benedí¹, Amparo Varona² y Eduardo Lleida³

[1] Universidad Politécnica de Valencia; Camino de Vera, s/n. 46022-Valencia <jbenedi@dsic.upv.es>

[2] Universidad del País Vasco; Barrio Sarriena, s/n. 48940-Leioa. <amparo@we.1c.ehu.es>

[3] Universidad de Zaragoza; María de Luna, 1. 50015-Zaragoza <lleida@posta.unizar.es>

Resumen

En este trabajo se presentan las metodologías de diseño y la implementación de un sistema de diálogo para el acceso a un sistema información. El principal objetivo es el estudio y desarrollo de aspectos metodológicos fundamentales en temas de habla espontánea, modelización del lenguaje, comprensión y diálogo. También se hace hincapié en el desarrollo de aquellas opciones de diseño que favorezcan la robustez del sistema ante los fenómenos del habla espontánea, los diferentes canales de telefonía y ambientes no telefónicos como entornos de automóvil. La tarea seleccionada como base para dicha investigación es la de consulta telefónica de horarios y precios de trenes de grandes líneas. Para esta tarea se ha implementado un prototipo previo y un entorno de adquisición de diálogos hablados mediante la técnica persona-sistema-mago. Con este entorno se han adquirido 900 diálogos de 75 interlocutores diferentes, lo que supone 6279 turnos de usuario en algo más de 10.8 horas de grabación.

1. Introducción

Un sistema de diálogo es un sistema automático capaz de emular a un ser humano en un diálogo con otra persona, con el objetivo de que el sistema cumpla con una cierta tarea (dar una cierta información, o proporcionar ciertos servicios). Un sistema de estas características se puede describir en términos de los siguientes módulos:

- *Reconocimiento automático del habla*, reconoce la señal vocal pronunciada por el usuario y proporciona la secuencia de palabras reconocida más probable (o las k más probables).
- *Comprensión*, a partir de las(s) secuencia(s) de palabra(s) reconocida(s), el sistema obtiene una representación semántica de su significado.
- *Diálogo*, considera la interpretación semántica de la petición del usuario, la historia del proceso de diálogo, la información de la aplicación disponible en ese punto y el estado del sistema, y determina la estrategia de diálogo a seguir: *acceder a la base de datos de la aplicación*, recogiendo y organizando la información proporcionada; y/o *generar respuestas al usuario*, para comunicarle el resultado de su consulta o para solicitar aclaraciones, cuando la información disponible no sea suficiente. Este módulo está conectado a un sintetizador texto-voz.

Para tareas complejas, el desarrollo de sistemas donde la iniciativa pueda ser compartida entre el usuario y el sistema, requiere un alto grado de desarrollo de las tecnologías involucradas:

representación robusta de la voz, reconocimiento del habla, tratamiento del habla espontánea, tratamiento y comprensión del lenguaje y modelización del diálogo. En esta línea, existen una gran cantidad de sistemas clásicos, entre otros: ARISE [6], LIMSI [14], CMU ATIS [31], MIT ATIS [26] y AT& ATIS [21]. Entre los sistemas españoles cabría citar: ATOS [1] y sobre todo BASURDE [2] como claro precursor del proyecto aquí presentado.

En este trabajo se presenta un proyecto para el desarrollo de un sistema de diálogo para el acceso a la información mediante habla espontánea en diferentes entornos DIHANA [7]. Este proyecto comenzó en diciembre de 2002 y finalizará en noviembre de 2005. El objetivo del DIHANA es el estudio y desarrollo de un sistema robusto de diálogo modular y distribuido para el acceso a sistemas de información. En concreto, se pretende profundizar en aspectos metodológicos fundamentales en los campos del *modelado acústico* en diferentes entornos, *tratamiento del habla espontánea*, *modelado del lenguaje*, *comprensión* y *diálogo*. Así mismo, se pone especial énfasis en aspectos tanto tecnológicos como metodológicos necesarios para extender nuestra propuesta a otros ámbitos de aplicación diferentes, como la *representación robusta de la voz*, el *acceso manos libres en entorno de automóvil* y la *adaptación de los modelos acústico-fonéticos*.

A continuación se va a realizar una descripción general del proyecto mostrando en detalle sus componentes principales: módulos de reconocimiento del habla, comprensión y diálogo. Igualmente, se presentará los resultados más importantes alcanzados hasta este momento.

2. Corpus

La tarea seleccionada para la adquisición del corpus oral de diálogos ha sido: *consulta telefónica de horarios y precios de trenes de grandes líneas*. Esta tarea también fue elegida en el proyecto BASURDE [2]. Las principales características son: vocabulario estable y regular, restringida semánticamente y estructura del diálogo relativamente rica y abierta.

La adquisición de un corpus específico de diálogo usuario-sistema plantea una gran dificultad; ya que, para que esta adquisición se realice de una manera *natural* se precisa un sistema de diálogo que funcione eficientemente, pero para desarrollar un sistema de diálogo eficiente es necesario una gran cantidad de datos (diálogos naturales) para el entrenamiento de sus modelos. La solución típica para obtener una adquisición inicial consiste en emplear la técnica de *Mago de Oz*. La labor del Mago consiste en ayudar al usuario a obtener respuestas a sus consultas simulando al sistema; por lo que, en sucesivos turnos de diálogo, interacciona con el usuario siguiendo una estrategia dada. Esta técnica se puede desarrollar, o bien, mediante una adquisición completamente simulada *persona-mago*, o bien, empleando un prototipo simple, mediante una adquisición limi-

tada persona-sistema-mago. En el último caso el mago interviene corrigiendo también los errores del sistema.

En DIHANA se ha seguido la segunda de estas estrategias. Para ello se ha desarrollado una plataforma de adquisición conectable directamente al sistema permitiendo al *mago* controlar todo el sistema de diálogo: escuchando al interlocutor, recibiendo la información de los módulos de reconocimiento y comprensión del habla e interactuando con el propio interlocutor [20].

Dada la dificultad de la tarea seleccionada, en BASURDE se introdujeron fuertes restricciones semántica, limitando el número de escenarios posibles, adquiriéndose un total de 215 diálogos. El número de turnos de usuarios fue de 1460, con un total aproximado de 14902 palabras. En el marco del proyecto DIHANA han participado 225 hablantes (153 hombres y 72 mujeres). Todos ellos realizaron 4 diálogos y pronunciaron 16 frases leídas (8 de la tarea y 8 fonéticamente balanceadas). En resumen se han adquirido:

- 900 diálogos de la tarea elegida, mediante la técnica del Mago de Oz;
- 1800 frases (leídas) de la tarea, para el estudio del fenómeno de las disfluencias en el habla espontánea;
- 1.800 frases fonéticamente balanceadas, para el aprendizaje de los modelos acústicos;
- 1800 frases fonéticamente balanceadas, adquiridas en el interior del automóvil, para el aprendizaje de los modelos acústicos en dicho entorno;

El número de turnos de usuario en los 900 diálogos adquiridos es de 6279, con un total aproximado de 48631 palabras. La duración de la grabación es de aproximadamente 10.8 horas. Este material se ha dividido en un corpus de entrenamiento y otro de test. El corpus de entrenamiento está grabado por 180 locutores (122 hombres y 58 mujeres), con un total de 4929 turnos de usuario y aproximadamente 38015 palabras en 8.5 horas de grabación. El corpus de test está grabado por 45 locutores (31 hombres y 14 mujeres), con un total de 1350 turnos de usuario y aproximadamente 10616 palabras en 2.3 horas de grabación.

3. Tratamiento del habla espontánea

Uno de los principales objetivos del proyecto es la integración, tanto en el modelado acústico como en el de lenguaje, de los fenómenos de habla espontánea que aparezcan: pausas de silencio, pausas habladas, sonidos ajenos al léxico, omisión de fonemas, palabras cortadas, vacilaciones, repeticiones, sustituciones, incorrecciones sintácticas, etc. Algunos de estos fenómenos -eventos no léxicos como pausas y vacilaciones- pueden ser modelados explícitamente y filtrados por el reconocedor. Otros más complejos resultan muy difíciles de detectar y requieren la ampliación y flexibilización del léxico, así como un modelo de lenguaje específico [27].

A partir de la transcripción ortográfica de la base de datos adquirida en DIHANA, se ha realizado un estudio y etiquetado de los fenómenos de habla espontánea a nivel acústico, léxico y sintáctico [22]. Los principales fenómenos encontrados que se han etiquetado son:

- 2050 ruidos aislados, de los cuales 1550 fueron debidos al locutor: por una aspiración (1160), ruido de labios (378) o tos (27);
- 1845 pausas de silencio;
- 1155 pausas habladas;
- 1640 alargamientos de sonidos, destacando: 497 de la vocal a, 603 de la vocal e y 182 de la vocal o;

- 450 disfluencias léxicas, de las cuales: 226 fueron palabras cortadas y 274 fueron palabras mal pronunciadas;
- 616 reformulaciones que incluyen: 406 repeticiones de una o varias palabras, 151 repeticiones con sustitución de alguna palabra y 54 repeticiones con inserción de alguna palabra.

La solución adoptada para tratar las disfluencias acústicas en el sistema de reconocimiento y con la que se ha obtenido una mejora muy significativa del rendimiento del reconocedor, ha sido la de modelar explícitamente los fenómenos de habla espontánea acústicos e incluirlos en el léxico y en el modelo de lenguaje como pseudo-palabras [13]. A la modelización explícita de los fenómenos de habla espontánea se ha añadido la modelización de los efectos de coarticulación en las unidades subléxicas mediante la definición de unidades contextuales que aportan reducciones muy significativas del error. Igualmente, se ha desarrollado también un nuevo algoritmo de clustering del conjunto de hablantes que trata de reducir la variabilidad debida al hablante dentro de los modelos acústicos, de modo que éstos sólo tengan en cuenta las diferencias entre los distintos sonidos de la lengua [23].

Para el tratamiento de las disfluencias léxicas se plantea la generación automática de modelos léxicos en forma de grafo, con objeto de incluir la enorme variedad de pronunciaciones que puede encontrarse en habla espontánea.

En los diálogos de habla espontánea es muy común que el usuario introduzca reformulaciones o repeticiones de una misma idea. Además, entre la primera pronunciación y la reformulación o repetición se suelen añadir pausas y/o *muletillas* que ayudan al usuario a pensar. Para el tratamiento de las disfluencias sintácticas se propone la construcción de modelos de lenguaje sintácticos que recojan estas peculiaridades y que se puedan integrar fácilmente con los modelos de lenguaje generales en el sistema completo.

4. Arquitectura del sistema

DIHANA está basado en una arquitectura distribuida (Figura 1) formada por 8 componentes: un servidor de audio (AUDIO), un servidor de reconocimiento automático del habla (RAH), un servidor de comprensión del habla (CH), un servidor de gestión del diálogo (GD), un servidor de Mago de Oz (MAGO), un servidor de generación de respuesta oral (GRO), un servidor de conversión texto-voz (CTV) y finalmente por un cliente gestor de comunicaciones (GC).

El gestor de comunicaciones está encargado de establecer todas las comunicaciones con el resto de módulos. Para iniciar una sesión de diálogo es necesario que el gestor de comunicaciones tenga conexiones activas con los módulos de: AUDIO, RAH, CH, GRO y CTV, y una conexión activa con: GD o el MAGO. El gestor de comunicaciones se puede controlar a través de un *applet* desde cualquier navegador. La comunicación entre módulos se realiza mediante mensajes. Todos los mensajes de información, a excepción de los mensajes de audio, pasan a través del gestor de comunicaciones para encaminarlos al servidor destino. La comunicación entre módulos se realiza utilizando mensajes en formato XML.

5. Módulo de reconocimiento del habla

Este módulo procesa y reconoce la señal vocal pronunciada por el usuario y proporciona al módulo de comprensión la

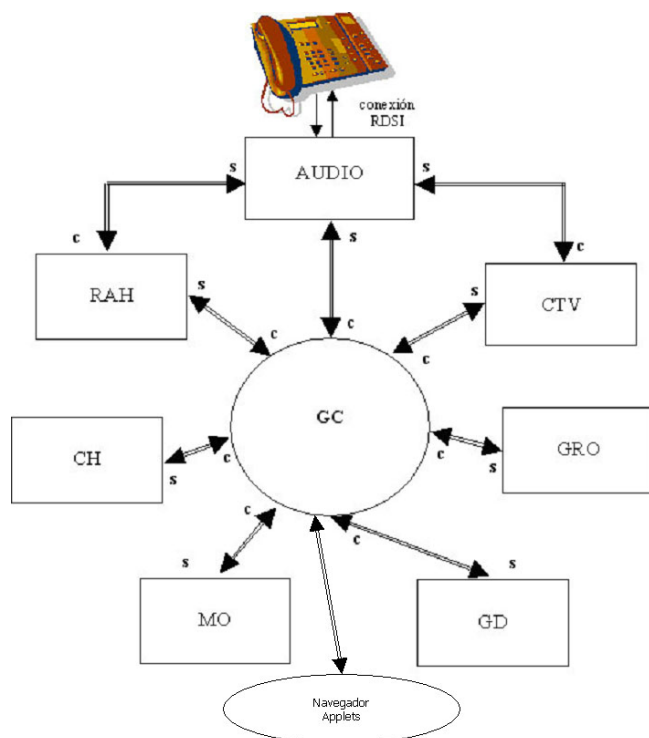


Figura 1: Arquitectura del sistema (c: cliente y s: servidor).

secuencia de palabras reconocida más probable (o las k más probables).

Interfaz de audio. En DIHANA, se han definido dos modelos de interfaz de audio. Un interfaz de audio fijo que permite interactuar con el sistema a través de la línea telefónica, y un interfaz de audio móvil pensado principalmente para interactuar con el sistema desde vehículos o dispositivos móviles (p.ej. dispositivos tipo PDA) con una conexión remota con el gestor de comunicaciones.

En la conexión del servidor de audio fijo con el sistema de reconocimiento, el servidor de audio puede enviar directamente las muestras de audio, codificadas en Ley A (8 bits/muestra), o puede enviar los parámetros de la señal de acuerdo con el estándar ETSI ES 201 108 [8].

En la conexión del servidor de audio móvil, existen dos posibilidades. En un tipo de servidor, todo el *front-end* del sistema de reconocimiento y el servidor de conversión-texto voz se encuentra en el vehículo o en el dispositivo móvil. De esta forma la comunicación remota con el resto de módulos se puede realizar con un ancho de banda reducido. Sin embargo, para entornos con conexión de alta velocidad, p.ej. *wi-fi*, el servidor de audio móvil actúa como el servidor de audio fijo, enviando y recibiendo muestras de la señal de voz, para utilizar el mínimo de recursos del dispositivo móvil.

En DIHANA se utiliza un sistema de manos libres para el acceso al sistema de diálogo desde vehículos que incluye cancelación de eco, para la comunicación *full-duplex* y cancelación de ruido para mejorar la relación señal a ruido de la señal de voz captada por los micrófonos [17].

Modelado acústico. El aprendizaje de los modelos acústicos en las mismas condiciones que se van a dar en el reconocimiento, requiere no sólo el aprendizaje de las características del habla sino también las del entorno acústico. Evidentemente, es

imposible disponer de modelos acústicos para todos los posibles entornos acústicos donde se va a utilizar el sistema. Una forma de solucionar el problema es compensar/adaptar los modelos entrenados en condiciones ideales al nuevo entorno o procesar la señal afectada por el entorno acústico para eliminar los efectos del mismo. En esta última opción entrarían los métodos propuestos tanto para el realzado de la señal de voz como para la cancelación de ecos.

Estos tratamientos permiten mejorar la relación señal a ruido y compensar los efectos del canal de comunicación. Sin embargo, esta mejora se consigue a costa de un aumento en la distorsión de la señal de voz original que en ocasiones puede hacer que el sistema de reconocimiento pierda prestaciones. En DIHANA se proponen métodos de adaptación y compensación rápidas para aplicaciones de reconocimiento automático del habla tanto en el entorno de telefonía como en el entorno del automóvil utilizando la proyección de los vectores de parámetros de un entorno acústico concreto sobre un conjunto de *auto-entornos* previamente definidos. Estos métodos de compensación permiten obtener mejoras en la tasa de error de hasta un 80 % en el entorno del automóvil [3].

Modelado del lenguaje. El modelo de lenguaje establece las restricciones y relaciones existentes en la combinación de las palabras de la tarea. En DIHANA se modelizan los distintos turnos del usuario utilizando modelos de lenguaje k -explorables en sentido estricto [28]. Este tipo de modelado proporciona autómatas o redes de estados finitos estocásticas otorgando probabilidades a secuencias de hasta k palabras. Adicionalmente, se incorporan modelos de categorías mediante una integración parcial de algunos modelos sencillos [30].

El modelo de lenguaje se construye mediante una única red que integre distintos autómatas de estados finitos (modelos de palabras, modelos sintácticos específicos para recoger las disfluencias sintácticas y modelos basados en categorías), para lo que es necesario formalizar la metodología de integración de las distintas redes.

Finalmente, también se han explorado la integración de estos modelos mediante modelos híbridos basados en gramáticas incontextuales estocásticas [12, 15].

Resultados de reconocimiento. Con la base de datos de voz, generada en el proyecto TIC95-0884, sobre red telefónica fija, consistente en 4349 frases, se ha realizado un entrenamiento de los Modelos Ocultos de Markov, dependientes del contexto, mono estado con funciones de densidad de probabilidad continuas representadas por mezclas de hasta 16 gaussianas. Posteriormente se ha utilizado la base de datos BASURDE, consistente en 1349 frases correspondientes a diálogos reales de la tarea, para estimar los modelos acústicos mediante el algoritmo MAP.

Se ha creado un modelo de lenguaje estocástico utilizando los turnos de diálogo de usuario de la base de datos de BASURDE. El vocabulario contiene un total de 792 entradas y la perplejidad del conjunto de test es de 15,7.

La Tabla 1 muestra los resultados que se obtienen, para distintos anchos del haz de búsqueda, sobre las frases de prueba definidas en la base de datos de BASURDE. Hay que destacar que se trata de habla con un alto grado de espontaneidad.

Utilizando las 3600 locuciones disponibles en la partición de frases de la nueva base de datos DIHANA, se ha realizado un reentrenamiento de los modelos acústicos obtenidos con la anterior base de datos. En la Tabla 2 se muestran los resultados obtenidos para un porcentaje de activación del 30 %, comparándolos con los de la Tabla 1.

Como se puede apreciar en la Tabla 2, los modelos acústicos reentrenados con los nuevos datos de DIHANA presenta una

%err	%I	%O	%S	%Acti	xTR
32.31	8.09	8.24	15.99	7.90	1
27.55	8.24	6.36	12.95	12.15	1.3
26.39	8.34	6.33	11.72	17.51	1.6
25.29	8.40	6.08	10.81	23.50	1.85
24.95	8.40	6.01	10.53	30.00	2.2
24.83	8.40	5.96	10.47	36.68	2.43
24.83	8.40	5.96	10.47	77.03	3.29

Tabla 1: Tasas de error en palabras (%err), inserciones (%I), omisiones (%O) y sustituciones (%S) para distintos porcentajes de actividad (%Acti) y factores de proceso en tiempo real (xTR) sobre Pentium IV Xeon a 1,8 GHz.

	%err	%I	%O	%S	xTR
BASURDE	24.95	8.40	6.01	10.53	2.2
DIHANA	17.24	6.77	3.60	6.86	2.2

Tabla 2: Tasas de error en palabras (%err), inserciones (%I), omisiones (%O), sustituciones (%S) y factores de proceso en tiempo real (xTR) sobre Pentium IV Xeon a 1,8 GHz, para un porcentaje de actividad del 30 %. Para modelos entrenados con BASURDE (Tabla 1) y reentrenados con DIHANA.

reducción de un 7,71 % la tasa de error, lo que supone un mejora superior al 30 %.

6. Módulo de comprensión

En un sistema de diálogo hablado, el objetivo del módulo de comprensión es proporcionar una representación semántica de la salida del reconocedor. La representación semántica escogida se basa en el concepto de *frame* (registro semántico) [6]: todo mensaje enviado por el módulo de comprensión, y por cada intervención del usuario, al módulo de diálogo es un frame. Para esta tarea, se han definido un conjunto de frames que permiten representar todas las intervenciones del usuario.

El proceso de traducción que lleva a cabo el módulo de comprensión se divide en dos grandes fases. En una primera, se obtiene, de forma automática, un *traductor estocástico* que proporciona una interpretación de la frase de entrada en términos de una secuencia de unidades semánticas definidas sobre un cierto alfabeto semántico. Por ejemplo:

Frase de entrada	Frase de salida
por favor	consulta
a que hora salen trenes	<hora_salida>
hacia	marcador_destino
Alicante	ciudad_destino

En una segunda fase, se traduce esta secuencia de unidades en su frame correspondiente.

En DIHANA se propone el uso de modelos estocásticos para abordar la construcción de estos traductores estocásticos que caractericen la comprensión del lenguaje hablado. Además, se exploran nuevas técnicas de aprendizaje automático de modelos estocásticos de estados finitos (traductores finitos, inferencia de lenguajes k -explorables con umbral) [25].

Sobre la transcripción de los 215 diálogos de la tarea adquiridos en BASURDE, se han construido unos modelos estocásticos de traducción en términos de las unidades semánticas definidas. Los primeros resultados sobre este corpus se muestran

en la Tabla 3. Donde (%cssu), es el porcentaje de secuencias

Modelo	% cssu	% csu	% cf
DIHANA	59.85	85.73	72.5

Tabla 3: Resultados de comprensión

de unidades semánticas correctas; (%csu), es el porcentaje de unidades semánticas correctas; y (%cf), es el porcentaje de frases correctas. La gran diferencia entre %cssu y %cf se debe a que aunque la sentencia de unidades semánticas de salida sea errónea, con respecto a la de referencia, su frame correspondiente si que es exacto.

Además del desarrollo de estos modelos, también se está trabajando en dos líneas de trabajo relacionadas:

- Desarrollo de modelos de comprensión que manejen medidas de confianza. Este trabajo se realiza en dos direcciones: utilizando para la comprensión las medidas de confianza que proporciona el reconocedor y generando las medidas de confianza del propio proceso de comprensión [10].
- Definición de modelos específicos de comprensión dependientes del estado del diálogo [11, 24].

7. Módulo de diálogo

El objetivo de este módulo es el diseño y la estimación de los *modelos estocásticos de diálogo* que caractericen el comportamiento del sistema en base a la información suministrada por el usuario, la historia del proceso de diálogo, la información semántica capturada (frame semántico) y el estado del sistema.

Igualmente, también será necesario construir un *gestor (estocástico) de diálogo* que modelice la estrategia de diálogo que determina la acción a tomar. Este gestor de diálogo debe permitir la interrelación tanto en los módulos internos del sistema: reconocimiento del habla, comprensión y diálogo, como con otros módulos que se relacionan con el exterior: generación de requerimientos a la base de datos de la aplicación y generación de respuestas al usuario.

La característica principal de esta aproximación es que los parámetros de los modelos estocásticos se aprenden (estiman) automáticamente a partir de ejemplos reales de diálogos, convenientemente procesados y etiquetados. En DIHANA, el etiquetado de los turnos del usuario se define en tres niveles [18]: 1) **Actos de diálogo**, independiente de la tarea, representan la intención de una parte de la intervención del usuario. 2) **Frames**, específico para la tarea, representa el tipo de mensaje proporcionado por el usuario. 3) **Cases**, representa los valores específicos dados en las frases.

En DIHANA se propone un marco estocástico formal para modelar el módulo de diálogo e integrarlo en el sistema completo. Estos modelos permiten predecir el acto de diálogo del sistema atendiendo a la petición del usuario, al estado del proceso de diálogo y a la historia del mismo [19]. Sobre los 215 diálogos de la tarea adquiridos en BASURDE, se ha diseñado y estimado un modelo estocástico de diálogo cuyos resultados preliminares se muestran en la Tabla 4 [19]. La talla de entrenamiento se refiere al número de diálogos. La precisión corresponde al porcentaje de actos de diálogo del usuario correctamente detectados (es decir, el acto o actos de diálogo asignado coincide con la anotación hecha de dicho turno). Y los modelos con los que se comparan son: FKW [9] y SCB [27]. La baja precisión de nuestro modelo

Modelo	#etiq.	Talla	Precisión
DIHANA	35	1060	57.4 %
FKW	26	3584	81.2 %
SCB	42	198000	65.0 %

Tabla 4: Resultados de precisión para distintos modelos con distinto número de etiquetas y distintas tallas del corpus de entrenamiento

con respecto a la del resto de modelos se explica por la falta de datos de entrenamiento (nuestro corpus de entrenamiento es mucho más reducido) y el gran número de etiquetas empleado. Esto hace que los resultados no sean comparables entre sí. Sin embargo, a pesar de las adversas condiciones en las que se ha desarrollado el modelo, éste parece presentar un comportamiento lo suficientemente bueno como para ser un punto de partida a usar en desarrollos posteriores.

Además del desarrollo de este marco formal, también se está trabajando en otras líneas de trabajo relacionadas:

- Detección y clasificación de actos de diálogo [4, 5]
- Desarrollo de modelos de diálogo dirigidos por la semántica que incorporen medidas de confianza de la etapa de comprensión para la adaptación de la estrategia del gestor de diálogo [29].
- Exploración de nuevas arquitecturas para sistema de diálogo basadas en VoiceXML [16].

8. Conclusiones

En este trabajo se ha presentado el proyecto DIHANA para el desarrollo de un sistema de diálogo hablado de acceso a un sistema de información. Igualmente se han descrito las propuestas metodológicas para abordar los temas de tratamiento de habla espontánea, reconocimiento del habla, comprensión y diálogo; junto con los resultados más importantes alcanzados hasta este momento en cada uno de ellos. Finalmente, se ha descrito el proceso de adquisición de los diálogos, así como el propio corpus adquirido.

En cuanto al trabajo futuro, y hasta la finalización del proyecto, se propone las siguientes grandes tareas: re-entrenar los modelos estocásticos de los diferentes módulos (reconocimiento del habla, comprensión y diálogo) con los nuevos datos del corpus adquirido; integrar las nuevas versiones re-entrenadas de los diferentes módulos en la arquitectura distribuida del sistema; y finalmente, evaluar el sistema tanto globalmente como individualmente para cada uno de sus componentes principales.

9. Referencias

- [1] Álvarez, J., D. Tapias, C. Crespo, I. Cortazar and F. Martínez: "Development and evaluation of the ATOS spontaneous speech conversational system". Proceedings of ICASSP, pp. 1139-1142 (1997).
- [2] Bonafonte, A., P. Aibar, N. Castell, E. Lleida, J.B. Mariño, E. Sanchis and M.I. Torres. "Desarrollo de un sistema de diálogo oral en dominios restringidos", I Jornadas en Tecnología del Habla, Sevilla, 2000.
- [3] Buera, L., E. Lleida, A. Miguel and A. Ortega: "Multi-Environment Models Based Linear Normalization for Robust Speech Recognition". SPECOM-2004, San Petersburgo (Rusia), Septiembre 2004 (Dinamarca), Sept. 2001.
- [4] Castro, M.J., D. Vilar, E. Sanchis and P. Aibar: "Uniclass and Multiclass Connectionist Classification of Dialogue Acts". En Progress in Pattern Recognition, Speech and Image Analysis, volumen 2905. Lecture Notes in Computer Science, páginas 266-273. Springer, 2003. (8th Iberoamerican Congress on Pattern Recognition (CIARP 2003)). Havana, Cuba, November 2003.
- [5] Castro, M.J., D. Vilar, P. Aibar and E. Sanchis: "Dialogue Act Classification in a Spoken Dialogue System". En Current Topics in Artificial Intelligence, volumen 3040 de Lecture Notes in Artificial Intelligence, páginas 260-270. Springer-Verlag, 2004.
- [6] den Os, E., L. Boves, L. Lamel and P. Baggia: "Overview of the ARISE project" Proceedings of EuroSpeech'99, 1999.
- [7] DIHANA. URL: <http://www.dihana.upv.es/>
- [8] ETSI ES 201 108 v1.1.2 (2000-04) Speech Processing, Transmission and Quality Aspects (STQ); Distributed Speech Recognition; Front-end feature extraction algorithm; Compression Algorithms.
- [9] Fukada, T., D. Koll, A. Waibel, K. Tanigaki: "Probabilistic Dialogue Act Extraction for Concept Based Multilingual Translation Systems". Proceedings of ICSLP'98, 1998.
- [10] García, F., Ll. Hurtado, E. Sanchis and E. Segarra: "The incorporation of Confidence Measures to Language Understanding". Lecture Notes in Artificial Intelligence series 2807. Springer Verlag. (International Conference on Text Speech and Dialogue (TSD 2003)). Vol LNAI 2807, pp. 165-172, 2003.
- [11] García, F., E. Sanchis, Ll. Hurtado, and E. Segarra: "Modelos específicos de comprensión en un sistema de diálogo". Procesamiento del Lenguaje Natural. ISSN: 1135-5948, N 31, pp. 99-106, 2003.
- [12] García, J., J.A. Sánchez and J.M. Benedí: "Performance and Improvements of a Language Model based on Stochastic Context-Free Grammars". F.J.Perales, A.J.C.Campilho, N.Pérez (Eds.), Springer Verlag, LNCS-2652, pp.271-278. Iberian Conference on Pattern Recognition and Image Analysis, Mallorca, junio 2003.
- [13] Guijarrubia, E., I. Torres and L.J. Rodríguez: "Evaluation of a Spoken Phonetic Database in Basque Language". Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004) Vol. 6, pp. 2127-2130. Lisboa, Portugal, May 26-28, 2004.
- [14] Lamel, L.F., S. K. Bennacef, et al. "The LIMSI Rail-Tel System: Field trial of telephone service for rail travel information". Speech Communication 23: 67-82, 1997.
- [15] Linares, D., J.M. Benedí and J.A. Sánchez: "A hybrid language model based on stochastic context-free grammars". In C. de la Higuera, P. Adriaans, M. van Zaanen, and J. Oncina, editors, ECML/PKDD 2003 Workshop on Learning Context-Free Grammars, pages 41-52, Septiembre 2003.
- [16] López-Cózar, R., R. Granell: "Sistema de Diálogo Basado en VoiceXML para Proporcionar Información de Viajes en Tren". Procesamiento del Lenguaje Natural. ISSN: 1135-5948, pages 171-178, 2004.
- [17] Lleida, E., E. Masgrau and A. Ortega: "Acoustic Echo Control and Noise Reduction for Cabin Car Commu-

- nication". Proceedings of the EUROSPEECH-2001, Aalborg (Dinamarca), Sept. 2001.
- [18] Martínez, C., E. Sanchis, F. García-Granada and P. Aibar: "A labelling proposal to annotate dialogues". Proceedings of the Third International Conference on Language Resources and Evaluation, 5, 1566–1582, Las Palmas de Gran Canaria, 2002.
 - [19] Martínez, C., and F. Casacuberta: "Evaluating a Probabilistic Dialogue Model for a Railway Information Task". Proceedings of the Fifth International Conference on Text, Speech, Lecture Notes in Artificial Intelligence LNCS/LNAI 2448, 381–388, 2002.
 - [20] Miguel, A., M.I. Galiano, R. Granell, L.I.F. Hurtado, J.A. Sánchez, E. Sanchis: "La plataforma de adquisición de diálogos en el proyecto DIHANA". Procesamiento del Lenguaje Natural. N.31, pp.343-344, septiembre 2003.
 - [21] Pieraccini, R. and E. Levin: "AMICA: the AT&T Mixed Initiative Conversational Architecture". Proceedings of EUROSPEECH'87, 1875-1878, 1997.
 - [22] Rodríguez, L.J. and I. Torres: "Annotation and analysis of acoustic and lexical events in a generic corpus of spontaneous speech Spanish". Procc of ISCA and IEEE workshop on Spontaneous Speech Processing and Recognition. pp 187-190. Tokio, Abril 2003.
 - [23] Rodríguez, L.J. and I. Torres. "A speaker clustering algorithm for fast speaker adaptation in continuous speech recognition". 7th International Conference on TEXT, SPEECH and DIALOGUE. Lecture Notes in Artificial Intelligence. Brno (Czech Republic). Setiembre 2004.
 - [24] Sanchis, E., M.J. Castro and D. Vilar "Stochastic Understanding Models Guided by Connectionist Dialogue Acts Detection". En Proceedings of 2003 IEEE Workshop on Automatic Speech Recognition and Understanding Workshop, páginas 501–506, St. Thomas, U.S., diciembre 2003.
 - [25] Segarra, E., E. Sanchis, F. García and L. Hurtado. "Extracting semantic information through automatic learning techniques". International Journal of Pattern Recognition and Artificial Intelligence, 16:3, 301–307, 2002.
 - [26] Seneff, S., L. Hirschman, V. Zue: "Interactive problem solving and dialogue in the ATIS domain". Proceedings of Fourth DARPA Speech and Natural Language Workshop, 354-359, 1991.
 - [27] Stolcke, A, N. Coccaro, R. Bates, P. Taylor, C. van Ess-Dykema, K. Ries, E. Shriberg, D. Jurafsky, R. Martin and M. Meteer: "Dialogue Act Modelling for Automatic Tagging and Recognition of Conversational Speech". Computational Linguistics 26(3),339-373. 2000.
 - [28] Torres, I. and A. Varona: "k-TSS language models in speech recognition systems". Computer Speech and Language, 15:2, pp 127-149, 2001.
 - [29] Torres, F., E. Sanchis and E. Segarra: "Development of a stochastic dialog manager driven by semantics". European Conference on Speech Communication and Technology. (EUROSPEECH'03). ISSN 1018-4074. pp. 605-608. (Suiza) 2003.
 - [30] Varona, A. and I. Torres: "Integrating High and Low Smoothed LMs in a CSR system". Procc of CIARP2003. Lecture Notes in Computer Science. Pp 236-243. Noviembre 2003.
 - [31] Ward, W. "Evaluation of the CMU ATIS System".

Proc. of DARPA Speech and Natural Language Workshop, 101- 105, 1991.