

Validación de un Sistema de Diálogo Mediante el Uso de Diferentes Umbrales de Poda en el Proceso de Reconocimiento Automático de Voz

R. López-Cózar, A. J. Rubio, J. E. Díaz-Verdejo, J. M. López-Soler

Dpto. Electrónica y Tecnología de Computadores, 18071 Universidad de Granada
Tel.: +34 958243271, FAX: +34 958243230 E-mail: {rlopezc,rubio,jedv,juanma}@ugr.es

Resumen: Este artículo presenta una nueva metodología cuyo objetivo es validar el funcionamiento de los sistemas de diálogo centrándose en dos aspectos: tiempo de respuesta y porcentaje de comprensión de frases. En primer lugar, el artículo realiza una descripción de la interfaz de entrada del sistema de diálogo usado en los experimentos, incluyendo una clasificación de las tareas de reconocimiento consideradas. A continuación, presenta los fundamentos de la técnica propuesta y muestra una aplicación de la misma para validar el funcionamiento del sistema de diálogo. Seguidamente muestra los resultados experimentales, los cuales indican, por una parte, que seis de las nueve tareas de reconocimiento diseñadas pueden ser consideradas válidas, pues se cumplen los requisitos impuestos respecto a tiempo de reconocimiento y porcentaje de comprensión de frases. Por otra parte, los resultados indican que para validar el funcionamiento del sistema es necesario cambiar las estrategias empleadas en las tres tareas restantes. Finalmente, el artículo muestra algunas líneas de trabajo futuro, encaminadas a utilizar nuevas estrategias para mejorar el funcionamiento del sistema de diálogo.

Palabras clave: Sistemas de diálogo, reconocimiento de voz, síntesis de voz, control del diálogo.

Abstract: This paper presents a new technique to validate the performance of dialogue systems focusing on two measures: response time a sentence understanding. Initially, the paper presents a description of the input interface of the dialogue system used in the experiments, including a classification of the recognition tasks considered. Later, it presents the basic features of the proposed technique and describes how the technique has been applied to validate the performance of the dialogue system. Later the paper presents the experimental results which, on the one hand, show that six out of the nine recognition tasks employed by the system can be considered validated, since the imposed restrictions on recognition time and sentence understanding are kept. On the other hand, the results show that for improving the system it is necessary to change the strategies used for the remainder three tasks. Finally, the paper shows some possibilities for future work related to the new strategies employable to enhance the performance of the dialogue system.

Keywords: Dialogue systems, speech recognition, speech synthesis, dialogue management.

1 Introducción

Los sistemas de diálogo (*Spoken Dialogue Systems*) son sistemas automáticos capaces de interactuar con los usuarios mediante el habla a fin de realizar determinadas tareas, como por ejemplo, consultas a bases de datos, reservas de viajes, obtención de información meteorológica, localización de números de teléfono en directorios, localización de información en Internet, etc. (Gustafson et al. 1999; Asoh et al. 1999; Chao et al. 1999). Dado que el habla constituye la forma más

directa y natural de comunicación entre las personas, estos sistemas puede aumentar la rapidez, efectividad y facilidad a la hora de realizar dichas tareas de forma automática. No obstante, para que dichos sistemas sean realmente útiles para el público en general, es necesario que funcionen en tiempo real y cuenten con técnicas que permitan procesar satisfactoriamente los errores cometidos por los módulos de reconocimiento empleados. Por tanto, sus diseñadores deben considerar simultáneamente cuestiones puramente

tecnológicas junto con otras relacionadas con el uso de los mismos.

Inicialmente, las interfaces basadas en el habla se desarrollaron para trabajar a partir de texto escrito. Sin embargo, los sistemas de diálogo deben enfrentarse a fenómenos no existentes en el texto escrito, como por ejemplo, diversidad de las características de la voz de cada usuario, tamaño del diccionario (conjunto de posibles palabras a reconocer), existencia de palabras fragmentadas (por ejemplo, por solapamiento con otras), falsos comienzos, etc. Además, estos sistemas deben enfrentarse a la posibilidad de que los usuarios utilicen palabras desconocidas para ellos.

Los corpora de grabaciones de diálogos, palabras y frases son esenciales para el desarrollo de estos sistemas ya que es necesario determinar el conjunto de palabras y expresiones lingüísticas relacionadas con el dominio de aplicación de cada nuevo sistema a desarrollar. Además, es necesario obtener los modelos acústicos independientes del locutor que utilizará el módulo de reconocimiento del sistema, así como estimar los modelos del lenguaje (Ziegenhain et al. 1998).

Si bien los avances logrados por esta tecnología durante los últimos años son considerables, existen aún diversos problemas por resolver. Asimismo, son numerosas las líneas de investigación abiertas, las cuales tienen como finalidad lograr una interacción más cómoda y natural para lograr una mayor aceptación por parte de los usuarios.

2 El sistema de diálogo SAPLEN

El sistema SAPLEN es un sistema de diálogo desarrollado en nuestro laboratorio para atender telefónicamente a los clientes de restaurantes de comida rápida (López-Cózar et al. 1999). Su finalidad es ofrecer un servicio automático durante las 24 horas del día que permita proporcionar información y mantener un registro de productos encargados telefónicamente, los cuales puedan ser enviados posteriormente a domicilio.

Para desarrollar el sistema, hemos utilizado un corpus de 520 diálogos grabados en un restaurante de comida rápida. A partir del análisis de este corpus hemos podido conocer el vocabulario del dominio de aplicación, las estructuras sintácticas y semánticas que los clientes suelen usar, los objetivos que tanto los clientes como los encargados del restaurante

tienen durante la conversación, y el conjunto de frases que el sistema debe generar, tanto para proporcionar la información solicitada por los clientes, como para interactuar con los mismos.

A grandes rasgos, el sistema consta de una interfaz de entrada, un módulo de control del diálogo y una interfaz de salida (compuesta por un generador de respuestas y un sintetizador de voz). La Figura 1 muestra la estructura modular de la interfaz de entrada, constituida por un reconocedor de voz y un analizador semántico.

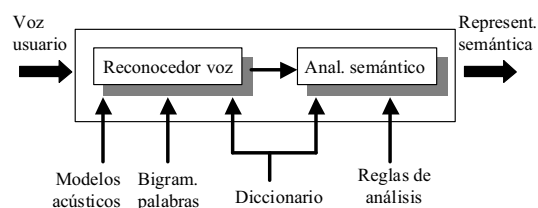


Figura 1. Interfaz de entrada del sistema SAPLEN

2.1 Reconocedor de voz

El sistema usa un reconocedor de voz continua independiente del locutor desarrollado en nuestro laboratorio, el cual está basado en unidades fonéticas independientes del contexto modeladas mediante SCHMM (modelos ocultos de Markov semicontinuos) (Rabiner et al. 1986, 1993). Este reconocedor se usa también en otros sistemas de diálogo desarrollados en nuestro laboratorio, como por ejemplo en el sistema STACC, cuya finalidad es proporcionar telefónicamente las calificaciones obtenidas por los estudiantes de diversas titulaciones de la Universidad de Granada (Rubio et al. 1997).

El tamaño del vocabulario del sistema SAPLEN es de 2.000 palabras aproximadamente, incluyendo nombres de productos del restaurante (principalmente comida y bebidas), números, nombres de calles, avenidas, plazas, etc.

El lenguaje se modela mediante 126 bigramáticas de palabras cuyas probabilidades se estiman a partir de conjuntos de frases representativas de las diversas tareas de reconocimiento consideradas en el dominio de aplicación (Nasr et al. 1999; Jelinek 1999). Para crear estas frases las palabras se agrupan en clases y se realiza una combinación automática de dichas clases siguiendo las estructuras sintácticas y semánticas conocidas a partir del análisis del corpus de diálogos.

Durante los experimentos se utilizan 109 bigramáticas para el reconocimiento de los datos de los usuarios (número de teléfono, código postal y dirección) y las restantes 17 bigramáticas se usan para el reconocimiento de los pedidos de productos, peticiones de información, confirmaciones, consultas y correcciones de los usuarios.

La Tabla 1 muestra las diversas tareas de reconocimiento consideradas en el dominio de aplicación del sistema de diálogo, así como el número de palabras diferentes que pueden combinarse entre sí para formar posibles frases de los usuarios.

Tarea	Nº palabras diferentes
(1) Confirmación guiada	6
(2) Confirmación libre	24
(3) Petición de información	8
(4) Corrección	16
(5) Código postal	116
(6) Pedido de productos	158
(7) Número de teléfono	116
(8) Consulta	220
(9) Dirección	1723

Tabla 1. Tareas de reconocimiento consideradas en el dominio de aplicación

Denominamos *confirmaciones guiadas* a aquéllas generadas por el sistema de diálogo para confirmar datos críticos para la aplicación (número de teléfono, código postal, dirección, productos encargados, precio a pagar y tiempo estimado de entrega a domicilio de los productos encargados). Para estos datos el sistema solicita confirmaciones del tipo “sí/no” a los usuarios a fin de obtener el mayor porcentaje posible de reconocimiento. Por ejemplo, para confirmar un número de teléfono el sistema puede generar la siguiente confirmación: “¿Has dicho 9, 5, 8, 17, 13, 28?, Por favor, responde sí o no”.

Denominamos *confirmaciones libres* a aquéllas otras generadas por el sistema para confirmar datos no críticos en la aplicación. En este caso, el sistema no indica a los usuarios las palabras que deben utilizar para realizar la confirmación, pudiendo utilizar múltiples palabras y expresiones (tales como, “vale”, “claro”, “por supuesto”, “de acuerdo”, “en absoluto”, etc.).

2.2 Analizador semántico

El reconocedor de voz procesa la señal de voz del usuario y proporciona como resultado una secuencia de palabras que constituye la entrada del analizador semántico.

Para construir la representación semántica correspondiente a la entrada, el analizador semántico detecta las palabras clave (*keywords*) existentes en la misma y utiliza reglas semánticas que ignoran las palabras no significativas desde un punto de vista semántico (Boros et al. 1999; Schadle et al. 1999).

Las ventajas principales de este tipo de análisis son dos. Por una parte, se puede evitar ambigüedades sintácticas si las interpretaciones semánticas que se obtienen no tienen significado. Por otra parte, se puede ignorar los detalles sintácticos que no afectan a la interpretación semántica. Concretamente, el analizador utiliza 45 de estas reglas, las cuales permiten obtener las representaciones semánticas de las frases mediante unas estructuras conceptuales denominadas *frames* (Allen, 1995).

A fin de contar con cierta robustez frente a errores de reconocimiento, el analizador emplea una estrategia de recuperación implícita de errores que permite, en ocasiones, obtener la interpretación semántica correcta a pesar de que algunas palabras no hayan sido reconocidas correctamente por parte del reconocedor de voz.

3 Técnica propuesta para validar el funcionamiento del sistema de diálogo

El porcentaje de reconocimiento y el tiempo de reconocimiento son dos medidas del funcionamiento de los sistemas de reconocimiento que se encuentran en mutua oposición. Generalmente, para lograr un porcentaje de reconocimiento suficientemente alta es necesario explorar un número considerable de frases candidatas, lo cual conlleva un tiempo de reconocimiento generalmente alto.

El tiempo de respuesta es un factor primordial para los sistemas de diálogo. Para que estos sistemas sean aceptados por los usuarios es necesario que generen sus respuestas casi en tiempo real. De lo contrario,

tales sistemas son considerados lentos y, por consiguiente, son rechazados por los usuarios.

En la práctica, es usual llegar a un compromiso respecto a los valores admisibles de las dos medidas mencionadas anteriormente, a fin de contar con porcentajes de reconocimiento suficientemente altas y tiempos de reconocimiento suficientemente bajos. Para alcanzar tal compromiso, se suele usar un valor conocido como *umbral de poda* (Up), el cual permite descartar frases poco probables durante el proceso de reconocimiento. Dado que dicho umbral permite reducir el número total de posibles frases a explorar, su uso permite reducir el tiempo requerido por el proceso de reconocimiento de las frases.

La técnica propuesta en este artículo para validar el funcionamiento del sistema de diálogo consiste en encontrar un umbral de poda Up_i que, para cada tarea de reconocimiento considerada i , satisfaga los requisitos de porcentaje de reconocimiento y tiempo de reconocimiento, de forma tal que el funcionamiento del sistema pueda ser considerado aceptable por los usuarios.

Si es posible encontrar dicho umbral, se construye un par (B_i, Up_i) , donde B_i representa la bigramática usada por el sistema para reconocer las frases correspondientes a la tarea i . La finalidad de este par es lograr que el reconocedor de voz del sistema utilice una bigramática específica a la hora de analizar cada frase del usuario, junto con un umbral de poda específico para dicha bigramática.

Si no es posible encontrar un umbral de poda que satisfaga ambos requisitos (porcentaje de reconocimiento y tiempo de reconocimiento) para una determinada tarea, ello significa que para validar el sistema es necesario cambiar la estrategia utilizada para realizar el reconocimiento de frases relacionadas con dicha tarea, y utilizar otra estrategia que sí permita cumplir ambos requisitos.

4 Resultados experimentales

Para aplicar la técnica propuesta en este artículo al sistema de diálogo SAPLEN, utilizamos un corpus de frases previamente grabado por seis locutores en condiciones (no ruidosas) de laboratorio. El corpus consta de 100 frases por cada una de las tareas de reconocimiento mostradas en la Tabla 1, esto es, $100 \times 9 = 900$ frases.

Para crear el corpus, las frases fueron leídas espontáneamente por los locutores. Teóricamente, todas ellas podían ser reconocidas correctamente pues eran permitidas por las bigramáticas. Cada grabación se almacenó mediante un fichero de muestras de voz, usando 16 bits/muestra a una frecuencia de muestreo de 8 KHz.

A fin de determinar si era posible encontrar un umbral de poda (Up) que satisficiera los requisitos de porcentaje y tiempo de reconocimiento para las nueve tareas mostradas en la Tabla 1, se probaron 6 valores diferentes de Up (10, 20, 30, 40 50 y 60). Usamos concretamente estos valores pues experimentos previos con otros sistemas de diálogo probaron que, usando el mismo reconocedor de voz, se podían obtener resultados aceptables para ambas medidas.

De entre las diversas medidas existentes en la bibliografía para medir el porcentaje de reconocimiento, en los experimentos hemos utilizado la conocida como *exactitud de palabras* (*Word Accuracy, WA*), según la cual, una determinada frase es reconocida correctamente si la salida del reconocedor coincide exactamente con la secuencia de palabras existente en la frase.

También hemos utilizado como medida la conocida como *recuperación implícita* (*Implicit Recovery*), la cual representa la capacidad del sistema de diálogo para comprender correctamente frases reconocidas con errores.

Hemos utilizado una medida más en los experimentos denominada *comprensión de frases* (*Sentence Understanding*) (Danieli y Gerbino, 1995), según la cual, una frase es comprendida correctamente si la representación semántica obtenida por el analizador semántico coincide con la representación semántica correcta de la frase.

Para obtener los valores de dichas medidas, procesamos mediante la interfaz de entrada del sistema (es decir, reconocedor de voz y analizador semántico) las 900 frases, usando cada uno de los 6 umbrales de poda. Por consiguiente, realizamos $900 \times 6 = 5.400$ análisis de frases.

4.1 Tiempo de reconocimiento

La Tabla 2 muestra los resultados obtenidos con respecto al tiempo promedio de reconocimiento de las frases. Puede observarse que un incremento en el valor del umbral de poda

conlleva un incremento en el tiempo de reconocimiento. En término medio, el tiempo de reconocimiento de las tareas (1)-(6) requirió menos de 5 segundos, el reconocimiento de los números de teléfono (tarea 7) requirió 5.22 segundos, el reconocimiento de las consultas (tarea 8) requirió 7.20 segundos y el reconocimiento de las direcciones (tarea 9) requirió 16.45 segundos.

Tarea	Up=10	Up=20	Up=30	Up=40	Up=50	Up=60
(1)	3.04	3.05	3.05	3.05	3.05	3.06
(2)	3.17	3.18	3.20	3.21	3.23	3.24
(3)	3.68	3.70	3.79	4.02	4.33	4.54
(4)	3.65	3.64	3.71	3.76	3.94	4.24
(5)	3.50	3.55	3.65	3.81	4.0	4.31
(6)	3.65	3.65	3.75	3.92	4.05	4.32
(7)	4.02	4.12	4.91	4.93	5.84	7.05
(8)	5.35	5.82	6.25	7.14	8.35	10.30
(9)	4.35	6.35	10.14	15.55	24.15	38.20

Tabla 2. Tiempo promedio de reconocimiento (segundos)

4.2 Exactitud de palabras

La Tabla 3 muestra los resultados obtenidos con respecto a la exactitud de palabras. En esta tabla puede observarse, asimismo, que el valor de la exactitud de palabras aumenta conforme lo hace el valor del umbral de poda.

Tarea	Up=10	Up=20	Up=30	Up=40	Up=50	Up=60
(1)	60.0	85.0	93.0	100	100	100
(2)	20.0	48.0	55.0	60.0	61.0	61.0
(3)	5.0	10.0	12.0	41.0	58.0	65.0
(4)	48.0	72.0	82.0	85.0	90.0	91.0
(5)	76.63	92.42	98.62	98.62	98.96	98.96
(6)	67.16	79.10	83.58	85.07	86.56	93.41
(7)	27.53	52.17	80.67	90.33	90.82	91.30
(8)	51.35	80.99	81.08	90.99	91.20	93.72
(9)	45.97	57.47	80.45	81.60	83.90	85.05

Tabla 3. Exactitud de palabras (%)

La mayor valor de la exactitud corresponde a las *confirmaciones guiadas* (tarea 1) pues el vocabulario usado en esta tarea es muy reducido y la gramática utilizada es muy simple.

Los valores obtenidos para las confirmaciones libres (tarea 2) y para las peticiones de información (tarea 3) son muy bajos. Con respecto a las *confirmaciones libres*, ello se debe a que el reconocedor cambia, por error, muchas palabras por otras acústicamente similares. Además, se produce

un considerable número de inserciones de disfluencias (como por ejemplo, “*eh*”), las cuales son permitidas por las bigramáticas. No obstante, muchos de estos errores son recuperados implícitamente durante el análisis semántico.

Dado que los usuarios pueden solicitar información en cualquier momento del diálogo, no se definió ninguna bigramática específica para el reconocimiento de las *peticiones de información* (tarea 3). El resultado mostrado en la tabla para esta tarea es obtenido utilizando la bigramática correspondiente al pedido de productos. Como consecuencia, se produce un considerable número de palabras insertadas, ya que las salidas del reconocedor tienden a seguir las estructuras sintácticas correspondientes a los pedidos de productos.

4.3 Recuperación implícita

La Tabla 4 muestra los resultados obtenidos con respecto al porcentaje de recuperación implícita.

Tarea	Up=10	Up=20	Up=30	Up=40	Up=50	Up=60
(1)	17.50	26.66	42.85	0	0	0
(2)	43.54	65.0	83.78	78.12	81.81	81.81
(3)	-	-	-	-	-	-
(4)	51.85	33.33	0	0	0	0
(5)	-	-	-	-	-	-
(6)	36.36	42.85	50.0	40.0	50.0	36.36
(7)	-	-	-	-	-	-
(8)	6.25	12.50	16.66	33.33	33.33	33.33
(9)	13.33	20.0	55.55	66.66	75.0	66.0

Tabla 4. Recuperación implícita (%)

Como puede observarse, no existe recuperación implícita ni para los códigos postales (tarea 5) ni para los números de teléfono (tarea 7). Ello se debe a que un error de reconocimiento en cualquier dígito (o par de dígitos) provoca la construcción de una interpretación semántica aceptable por el sistema, pero que es incorrecta.

Tampoco existe recuperación implícita para las peticiones de información (tarea 3) pues las salidas del reconocedor correspondientes a esta tarea, en caso de ser correctas, únicamente podían ser de la forma “*necesito información*”, “*quiero información*”, etc., las cuales siempre son comprendidas correctamente; y cuando las salidas son erróneas siempre se obtienen representaciones semánticas incorrectas.

Las confirmaciones libres (tarea 2) obtienen el mayor porcentaje de recuperación implícita cuando $Up=30$, pues aproximadamente un 84% de las frases con algún error de reconocimiento son comprendidas correctamente.

4.4 Comprensión de frases

La Tabla 5 muestra los resultados obtenidos con respecto al porcentaje de comprensión de frases. Puede observarse que dicho porcentaje aumenta conforme lo hace el valor del umbral de poda.

El menor porcentaje de comprensión se obtiene para los números de teléfono (tarea 7) ya que la bigramática permite que cualquier dígito o par de dígitos sea seguido por cien palabras (dígitos o pares de dígitos). Además, un error en una palabra (dígito o par de dígitos) provoca que el número de teléfono sea incorrectamente comprendido.

Tarea	$Up=10$	$Up=20$	$Up=30$	$Up=40$	$Up=50$	$Up=60$
(1)	67.0	89.0	96.0	100	100	100
(2)	65.0	86.0	93.0	94.0	94.0	94.0
(3)	48.0	55.0	75.0	81.0	86.0	90.0
(4)	87.0	90.0	91.0	94.0	95.0	96.0
(5)	48.0	80.0	88.0	88.0	96.0	96.0
(6)	42.0	65.0	83.0	87.0	90.0	91.0
(7)	4.0	34.0	51.0	64.0	67.0	67.0
(8)	25.0	65.0	75.0	90.0	90.0	91.0
(9)	60.0	65.0	80.0	85.0	88.0	90.0

Tabla 5. Comprensión de frases (%)

4.5 Asignación de un umbral de poda a cada tarea de reconocimiento

A fin de asignar un umbral de poda a cada tarea de reconocimiento tenemos en cuenta, por una parte, los tiempos de reconocimiento obtenidos, y por otra, los porcentajes de exactitud de palabras (WA) junto con los de comprensión de frases (SU).

Como restricciones para obtener un funcionamiento aceptable del sistema, imponemos que el tiempo de reconocimiento no sea superior a 5 segundos, y la exactitud de palabras sea tal que proporcione, al menos, un 90% de comprensión de frases.

Teniendo en cuenta ambos requisitos hemos seleccionado, para cada tarea de reconocimiento, un valor del umbral de poda que permita llegar a un buen compromiso, teniendo en cuenta los resultados experimentales mostrados la Tablas 2, 3 y 5

(correspondientes a tiempo de reconocimiento, exactitud de palabras y comprensión de frases, respectivamente). La Tabla 6 muestra los valores del umbral de poda seleccionados para cada tarea.

Tarea	Up	Tiempo	WA	SU
(1)	50	3.05	100	100
(2)	60	3.24	61.0	94.0
(3)	60	4.54	65.0	90.0
(4)	60	4.24	91.0	96.0
(5)	60	4.31	98.96	96.0
(6)	60	4.32	93.41	91.0
(7)	50	5.84	90.82	67.0
(8)	40	7.14	90.99	90.0
(9)	40	15.55	81.60	85.0

Tabla 6. Umbrales de poda (Up) seleccionados para cada tarea de reconocimiento

5 Conclusiones y trabajo futuro

Como podemos observar en la Tabla 6, la selección de umbrales permite obtener resultados aceptables para las tareas (1)-(6), pues el tiempo de reconocimiento no supera los 5 segundos y el porcentaje de comprensión de frases alcanza el 90% en el peor de los casos. En cambio, tal selección no ha sido posible para el resto de tareas.

En la tarea 7 (números de teléfono) el tiempo de reconocimiento supera ligeramente los 5 segundos pero el porcentaje de comprensión no llega al 70%. Por consiguiente, para validar el funcionamiento del sistema debemos cambiar la estrategia empleada para esta tarea de reconocimiento. Una posibilidad podría consistir en permitir únicamente el reconocimiento de dígitos (0-9), y no el de pares de dígitos. De esta forma, podrían obtenerse mejores resultados respecto a ambas medidas, pero a costa de reducir la libertad de interacción de los usuarios, los cuales suelen utilizar combinaciones de dígitos aislados y pares de dígitos para indicar los números de teléfono (por ejemplo, 9 5 8 17 13 26).

Con respecto a la tarea 8 (consultas), ninguno de los umbrales de poda utilizados permite obtener resultados aceptables; el valor mostrado en la Tabla 6 proporciona un porcentaje de comprensión aceptable, pero a costa de un tiempo de reconocimiento ligeramente excesivo. Por consiguiente, para validar el funcionamiento del sistema se debe

utilizar una estrategia diferente para esta tarea de reconocimiento. Por ejemplo, se podría dividir dicha tarea en dos subtarefas relacionadas con los dos tipos de productos del restaurante (es decir, una relacionada con consultas acerca de “comidas” y otra con consultas acerca de “bebidas”). De esta forma, el sistema podría utilizar una bigramática u otra de acuerdo con el contexto de la conversación. Dado que cada una de estas dos bigramáticas tendría una menor perplejidad y un vocabulario más reducido, podemos esperar que los tiempos de reconocimiento se reduzcan y aumenten los porcentajes de comprensión de frases.

La tarea 9 (direcciones) parece ser la más compleja. Como se puede observar en la Tabla 6, el tiempo de reconocimiento supera con creces la limitación impuesta y el porcentaje de comprensión de frases es de sólo un 85%. Este resultado nos indica que para validar el funcionamiento del sistema debemos emplear una técnica diferente. Una posibilidad sería evitar el proceso de reconocimiento para este tipo de información, por ejemplo, grabando en un fichero la voz del usuario en el momento en que éste proporciona los datos de su dirección.

Bibliografía

- Allen J. 1995. Natural language understanding. The Benjamin/Cummings Publishing Company Inc.
- Asoh H., Matsui T., Fry J., Asano F. Hayamizu S. 1999. A Spoken Dialog System for a Mobile Office Robot. Eurospeech '99, pág. 1139-1142
- Boros M., Heisterkamp P. 1999. Linguistic Phrase Spotting in a Single Application Spoken Dialogue System. Eurospeech '99, pág. 1983-1986
- Chao H., Xu P., Zhang X., Zhao S., Huang T., Xu B. 1999. LODESTAR: A Mandarin Spoken Dialogue System for Travel Information Retrieval. Eurospeech '99, pág. 1159-1162
- Danieli M., Gerbino E. 1995. Metrics for evaluating dialogue strategies in a spoken language system. 1995 AAAI Spring Symposium on Empirical Methods in Discourse Interpretation and Generation, pp. 34-39
- Gustafson J., Lindberg N., Lundeberg M. 1999. The August Spoken Dialogue System. Eurospeech '99, pág. 1151-1154
- Jelinek F., 1999. Statistical Methods for Speech Recognition”, MA, MIT Press
- López-Cózar R., Rubio A. J., García P., Segura J. C. 1999. A New Word-Confidence Threshold Technique to Enhance the Performance of Spoken Dialogue Systems. Eurospeech '99, pág. 1395-1398
- Nasr A., Esteve Y., Béchet F., Spriet T., de Mori R. 1999. A Language Model Combining N-grams and Stochastic Finite State Automata. Eurospeech '99, pág. 2175-2178
- Rabiner L. R., Juang B. H. 1986. An Introduction to Hidden Markov Models. IEEE ASSP Magazine, Enero
- Rabiner L. R., Juang B. H. 1993. Fundamentals of Speech Recognition. Prentice-Hall
- Rubio A.J., García P., De la Torre A., Segura J.C., Díaz-Verdejo J.E., Benítez M.C., Sánchez V., Peinado A.M., López-Soler J.M., Pérez-Córdoba J.L. 1997. STACC: An Automatic Service for Information Access Using Continuous Speech Recognition Through Telephone Line. Eurospeech '97, pág. 1779-1782
- Schadle I., Antoine J. Y., Memmi D. 1999. Connectionist Language Models for Speech Understanding: The Problem of Word Order Variation. Eurospeech '99, pág. 2035-2038
- Ziegenhain U., Harengel S., Kaiser J., Wilhem R. 1998. Creating Large Pronunciation Lexica for Speech Applications. First International Conference on Language Resources and Evaluation, pág. 1039-1043