

DEPARTAMENTO DE INGENIERÍA ELECTRÓNICA
ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE TELECOMUNICACIÓN



TESIS DOCTORAL

**MEJORA DE SERVICIOS AUTOMÁTICOS
POR TELÉFONO CON RECONOCIMIENTO
DE HABLA: NUEVA GENERACIÓN DE
SERVIDORES VOCALES INTERACTIVOS**

Rubén San Segundo Hernández
Ingeniero de Telecomunicación

Director de Tesis
Dr. Ingeniero José Manuel Pardo Muñoz

2002

Tesis Doctoral: MEJORA DE SERVICIOS AUTOMÁTICOS POR
TELÉFONOS CON RECONOCIMIENTO DE
HABLA: NUEVA GENERACIÓN DE SERVIDORES
VOCALES INTERACTIVOS

Autor: RUBÉN SAN SEGUNDO HERNÁNDEZ

Director: Dr. INGENIERO JOSÉ MANUEL PARDO MUÑOZ

El tribunal nombrado para juzgar la tesis doctoral arriba citada, compuesto
por los doctores:

Presidente:

D. Javier Ferreiros López

Vocales:

D. Antonio José Rubio Ayuso

D. Eduardo Lleida Solano

D. Daniel Tapias Merino

Secretario:

D. Ricardo de Córdoba Herralde

Firma:



Handwritten signatures of the tribunal members: D. Javier Ferreiros López, D. Antonio José Rubio Ayuso, D. Eduardo Lleida Solano, D. Daniel Tapias Merino, and D. Ricardo de Córdoba Herralde.

Acuerda otorgarle la calificación de *sobresaliente con honde*
por unanimidad

Madrid, a *19* de *NOVIEMBRE* de *2002*

Resumen

En este trabajo se ha realizado un análisis e investigación en tres aspectos importantes que forman parte de un Servidor Vocal Interactivo (SVI): reconocimiento automático del habla, obtención de medidas de confianza para la detección de errores en los módulos de reconocimiento y comprensión de lenguaje natural, y por último, se ha invertido un esfuerzo importante en el módulo de gestión del diálogo.

En cuanto al módulo de reconocimiento, se ha realizado un estudio de la tarea de deletreo en castellano y se ha implementado el primer reconocedor de nombres deletreados en castellano con tasas de acierto comparables a los realizados en otros idiomas. En un primer paso se han evaluado diferentes estrategias de reconocimiento eligiendo una solución basada en una arquitectura de hipótesis y verificación que ofrece un mejor compromiso entre tasa de reconocimiento y tiempo de proceso. Sobre esta arquitectura, se han incorporado nuevas ideas para hacer frente a las peculiaridades de la tarea de deletreo en nuestro idioma, como la generación de modelos de silencios contextuales. Por otro lado, se ha desarrollado un reconocedor de habla continua para frases que expresan fechas y horas. Ambos sistemas han sido diseñados y entrenados para su funcionamiento por línea telefónica e independiente del locutor.

En relación con el análisis de medidas de confianza, se ha trabajado fundamentalmente sobre el sistema DARPA Communicator desarrollado en el Centro de Investigación de Lenguaje Hablado (CSLR: The Center for Spoken Language Research) de la Universidad de Colorado (Boulder) en Estados Unidos. Sobre este sistema se han realizado estudios independientes para los niveles de palabra, concepto semántico y frase completa. Por otro lado, también se han realizado análisis para los reconocedores implementados en la presente tesis, centrándonos en los niveles de frase para el sistema de nombres deletreados, y en el nivel de palabra para el reconocedor desarrollado en el dominio de fechas y horas. En esta parte del estudio se propone la utilización de las medidas de confianza como heurístico para la combinación de varias hipótesis de reconocimiento obtenidas de diferentes decodificadores.

En relación con la gestión del diálogo se propone una metodología de diseño en la que se combina información de diferentes fuentes: análisis de base de datos, observación de conversaciones reales, simulación del servicio y funcionamiento con usuarios reales. Esta metodología está formada por 5 fases. En la primera fase se realiza un análisis de la base de datos con la información disponible para ofrecer el servicio. En la segunda etapa "diseño por intuición", se propone la técnica de "brain-storming" para plantear diferentes opciones de diseño. En el diseño por observación (fase tercera), se analizan conversaciones entre los usuarios y operadores humanos para evaluar diferentes alternativas de diseño. En la cuarta fase (diseño por simulación) utilizamos la herramienta de Mago de Oz para simular una interacción usuario-sistema. Por último, en la etapa de mejora iterativa se describe la utilización de medidas de confianza para el diseño de los mecanismos de confirmación y se describe una técnica para el modelado del usuario basada en niveles de destreza.

La presentación de esta metodología se ha realizado mediante su aplicación al caso de un servicio de información y reserva de billetes de tren.

Abstract

In this work, we have analyzed and studied three important aspects of a Spoken Dialogue System: speech recognition, confidence measures analysis (for detecting recognition and understanding errors), and dialogue management.

About the recognition module, we have studied the spelling task in Spanish and we have developed the first spelled name recognizer in this language. With this system, we have got similar recognition rates than other systems developed for English or French. At the beginning, we evaluated different recognition architectures, selecting that one based in a hypothesis-verification approach. This architecture provides a good compromise between recognition rate and speed. Over this approach, we have included new ideas to deal with the specific characteristics observed in Spanish, like the contextual silence models. On the other hand, we have developed a continuous speech recognizer for dates and times. Both systems are speaker independent and they work over the telephone.

In the confidence measures analysis, we have focused mainly over the DARPA Communicator system, developed in CSLR: The Center for Spoken Language Research) at University of Colorado (Boulder) in USA. In this system, we have analyzed confidence measures at three different levels: word, semantic concept and phrase. We have also worked over the recognition systems developed in this thesis, focusing on the word level for the continuous speech decoder and the phrase level for the spelled name recognizer. In this study, we propose the confidence measures as heuristic for different recognition hypotheses combination.

About the dialogue management, we propose a design methodology that combines different information sources: database analysis, user-operator conversations, simulation and iterative testing with real users. This methodology consists of 5 phases. In the first one, we do an analysis of the database containing the information used to provide the service. In the second step, "design by intuition", we consider the "brain-storming" technique of proposing several design alternatives. In the design by observation (third phase), we analyze real dialogues between users and human operators. The fourth step is the design by simulation. We use the Wizard of Oz technique to simulate the system. Finally, in the design by iterative improvement, we describe how to use the confidence measures for the confirmation mechanism design and we propose an approach for user modeling, based in levels of skill.

This methodology has been presented showing the steps done for developing a railway information and reservation system for Spanish.

Índice de contenidos

Índice de contenidos	VII
Índice de figuras	XII
Índice de tablas	XV
Resumen	XXI
Abstract	XXIII

INTRODUCCIÓN

1.1 Servidores Vocales Interactivos	1-1
1.2 Objetivos de la tesis	1-3
1.2.1 Módulo de reconociendo del habla	1-3
1.2.2 Medidas de confianza	1-4
1.2.3 Gestión del diálogo	1-5
1.3 Contenido de la tesis	1-5

ENCUADRE CIENTÍFICO-TECNOLÓGICO

2.1 Introducción	2-1
2.2 Reconocimiento de nombres deletreados	2-1
2.3 Reconocimiento de habla continua	2-4
2.4 Medidas de confianza	2-6
2.5 Gestión de diálogo	2-10

SISTEMA DE RECONOCIMIENTO DE NOMBRES DELETREADOS

3.1 Introducción	3-1
3.2 Tarea de deletreo en Castellano	3-1
3.3 Arquitecturas de reconocimiento	3-4
3.3.1 Medidas de evaluación	3-5
3.3.1.1 Bandas de probabilidad	3-6
3.3.2 Arquitectura en dos niveles	3-6
3.3.3 Arquitectura integrada	3-8
3.3.4 Arquitectura de Hipótesis y Verificación	3-9
3.4 Características generales del sistema de reconocimiento utilizado	3-10
3.4.1 Modelado acústico utilizado	3-10
3.4.2 Estructura de reconocimiento	3-12
3.5 Base de datos	3-14
3.6 Etapa de Hipótesis	3-15
3.6.1 Sistema de referencia	3-15
3.6.2 Nueva topología con silencios contextuales	3-16
3.6.3 Incorporación de modelos de ruido	3-17
3.6.4 Modelos de lenguaje de letras	3-19
3.6.5 Obtención de las N mejores cadenas de letras	3-22
3.6.6 Consideración de un grafo de palabras	3-24

3.6.6.1 Obtención de las N mejores cadenas de letras	3-26
3.6.6.2 Incorporación del modelo de lenguaje 3-gram	3-27
3.6.7 Análisis de los conjuntos de letras con mayor confusión	3-29
3.7 Etapa de Verificación	3-31
3.8 Evaluación de campo	3-35
3.9 Conclusiones	3-37

SISTEMA DE RECONOCIMIENTO EN DOMINIOS RESTRINGIDOS: FECHAS Y HORAS

4.1 Introducción	4-1
4.1.1 Base de datos	4-2
4.1.2 Generación del vocabulario	4-2
4.1.3 Medidas de evaluación	4-3
4.2 Características generales del sistema de reconocimiento	4-4
4.2.1 Modelado acústico a utilizar	4-4
4.2.2 Efectos acústicos modelados	4-5
4.2.3 Entrenamiento de modelos acústicos	4-6
4.2.4 Algoritmo de decodificación	4-8
4.3 Análisis del modelado acústico	4-11
4.3.1 Longitud de los modelos	4-11
4.3.2 Entrenamiento selectivo	4-12
4.3.2.1 Aumento del número de trifenemas	4-14
4.3.2.2 Aumento del número de centroides	4-15
4.4 Grafo de palabras	4-17
4.4.1 Obtención del grafo de palabras	4-17
4.4.2 Procesado del grafo de palabras	4-20
4.5 Modelos de lenguaje	4-21
4.5.1 Gramáticas N-gram	4-21
4.5.2 Incorporación del modelo de lenguaje	4-21
4.5.2.1 Modelo 2-gram en el One-pass	4-21
4.5.2.2 Modelo 3-gram en el grafo de palabras	4-22
4.5.3 Modelos de lenguaje para habla leída	4-24
4.5.4 Modelos de lenguaje para habla espontánea	4-26
4.6 Consideración de las N mejores hipótesis	4-27
4.7 Conclusiones	4-29

ANÁLISIS DE MEDIDAS DE CONFIANZA EN SISTEMAS DE RECONOCIMIENTO DE HABLA CONTINUA

5.1 Introducción	5-1
5.2 Medidas de confianza en el sistema CU Communicator	5-1
5.2.1 Base de datos	5-1
5.2.1.1 Etiquetación automática de los ejemplos	5-2
5.2.2 Parámetros de medidas de confianza	5-4
5.2.2.1 Nivel de palabra	5-4
5.2.2.2 Nivel de concepto semántico	5-6

5.2.2.3 Nivel de frase	5-8
5.2.3 Combinación de parámetros	5-9
5.2.4 Evaluación de las medidas de confianza	5-10
5.2.5 Detección de errores	5-13
5.2.5.1 Nivel de palabra	5-13
5.2.5.2 Nivel de concepto	5-14
5.2.5.3 Nivel de frase	5-16
5.2.6 Aplicación de las medidas de confianza para la recuperación de errores	5-19
5.2.6.1 Métodos para la combinación de hipótesis	5-19
5.2.6.2 Algoritmo para la generación de un grafo de palabras a partir de varias hipótesis de reconocimiento	5-20
5.2.6.3 Experimentos al nivel de palabra	5-24
5.2.6.4 Experimentos al nivel de concepto semántico	5-27
5.3 Medidas de confianza sobre el reconocedor de nombres deletreados	5-28
5.3.1 Base de datos	5-29
5.3.2 Parámetros de medidas de confianza	5-29
5.3.2.1 Parámetros de la etapa de hipótesis	5-30
5.3.2.2 Parámetros de la etapa de verificación	5-31
5.3.3 Combinación de parámetros	5-31
5.3.4 Detección de errores de reconocimiento	5-32
5.3.5 Detección de nombres fuera del vocabulario de reconocimiento (OOV: Out of Vocabulary).....	5-33
5.3.6 Detección simultánea de errores y nombres fuera del vocabulario de reconocimiento	5-35
5.4 Medidas de confianza para el sistema de reconocimiento de fechas y horas	5-36
5.4.1 Base de datos	5-37
5.4.2 Parámetros utilizados	5-37
5.4.3 Combinación de parámetros	5-38
5.4.4 Detección de errores en habla leída	5-39
5.4.5 Detección de errores en habla espontánea	5-40
5.4.6 Aplicación de las medidas de confianza para la recuperación de errores de reconocimiento	5-42
5.5 Conclusiones	5-46

DISEÑO DE GESTORES DE DIÁLOGO

6.1 Introducción	6-1
6.1.1 Definiciones	6-2
6.2 Modelos de diálogo	6-5
6.2.1 Modelo basado en un autómata de estados finito	6-5
6.2.2 Modelo basado en plantillas	6-7
6.2.3 Modelo basado en árboles de objetivos	6-11
6.2.4 Modelo basado en árboles de objetivos dinámicos	6-13
6.3 El problema de la evaluación de los flujos del diálogo	6-15
6.4 Metodología de diseño	6-19
6.4.1 Análisis de la base de datos	6-19

6.4.1.1 Descripción formal	6-19
6.4.1.2 Aplicación al caso de ejemplo	6-21
6.4.2 Diseño por intuición	6-23
6.4.2.1 Descripción formal	6-23
6.4.2.2 Aplicación al caso de ejemplo	6-26
6.4.3 Diseño por observación	6-29
6.4.3.1 Descripción formal	6-29
6.4.3.2 Aplicación al caso de ejemplo	6-31
6.4.4 Diseño por simulación	6-35
6.4.4.1 Diseño de la herramienta de Mago de Oz (WOZ)	6-36
6.4.4.2 Descripción formal del análisis	6-38
6.4.4.3 Aplicación al caso de ejemplo	6-42
6.4.5 Diseño por mejora iterativa	6-45
6.4.5.1 Diseño de las confirmaciones de los datos	6-46
6.4.5.2 Modelado de usuario	6-58
6.4.5.3 Funcionalidad de carácter general	6-63
6.4.5.4 Ajustes finales	6-64
6.4.6 Evaluación: pruebas de campo	6-66
6.5 Conclusiones	6-69
 CONCLUSIONES Y LÍNEAS FUTURAS	
7.1 Conclusiones	7-1
7.1.1 Reconocedor de nombres deletreados	7-1
7.1.2 Reconocedor de fechas y horas	7-3
7.1.3 Análisis de medidas de confianza	7-4
7.1.4 Metodología de diseño de gestores de diálogo	7-5
7.2 Líneas futuras de trabajo	7-6
 APÉNDICES	
Apéndice A. Descripción de los servidores vocales interactivos utilizados	A-1
A.1 El sistema Communicator de la Universidad de Colorado: CU Communicator	A-1
A.1.1 Arquitectura GALAXY-II	A-2
A.1.2 Servidores del sistema CU Communicator	A-4
A.1.3 Captura de datos	A-6
A.1.4 Evaluación por NIST en Junio 2000	A-7
A.2 TADE (Telephone Application Development Enviroment)	A-8
A.2.1 El entorno TADE	A-8
A.2.2 Funcionalidad del Lenguaje en el entorno TADE	A-9
A.2.3 Servicios comerciales desarrollados con el entorno TADE	A-13
A.3 Servicio de páginas blancas	A-14
A.4 Servicio de información y reservas de billetes de tren	A-17
Apéndice B. Experimentos independientes del locutor en el reconocedor de nombres deletreados	B-1

Apéndice C. Inventarios de alófonos utilizado en el reconocedor de fechas y horas	C-1
Apéndice D. Escenarios y cuestionario diseñado para la evaluación en la etapa de simulación	D-1
D.1 Escenarios	D-1
D.2 Cuestionario	D-2
Apéndice E. Escenarios y cuestionario diseñado para la evaluación final del sistema de información de tren	E-1
Apéndice F. Recomendaciones auxiliares para el diseño de gestores de diálogo en SVIs	F-1
Apéndice G. Evaluación de campo del sistema de información y reserva ferroviaria según el tipo de usuario	G-1
G.1 Evaluación en el entorno universitario	G-1
G.2 Evaluación de los empleados de RENFE	G-2
 BIBLIOGRAFÍA	
Bibliografía	Biblio-1