

## 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 “brainstorming” 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.