

# Sistema de información ferroviaria por teléfono: propuesta de una metodología de diseño de gestores de diálogo

R. San-Segundo, J.M. Montero, J. Ferreiros, J. Macías-Guarasa y J.M. Pardo

Grupo de Tecnología del Habla. Departamento de Ingeniería Electrónica. UPM  
E.T.S.I. Telecomunicación. Ciudad Universitaria s/n, 28040 Madrid, Spain

<mailto:{lapiz|juancholjfl|macias|pardo}@die.upm.es>

**Resumen:** En esta contribución se describen las fases llevadas a cabo en el desarrollo un servicio de información ferroviaria en castellano, presentándose una nueva metodología para el diseño de gestores de diálogo en servicios automáticos de información. En esta metodología se propone la combinación de diferentes fuentes de información: intuición, observación y simulación, para definir y evaluar alternativas de diseño. En este artículo, presentamos el diseño de los mecanismos de confirmación y las técnicas de modelado de usuario. En las pruebas de campo realizadas, los usuarios han valorado al sistema con una puntuación global de 3.9 (de 1 a 5), obteniendo un tiempo medio de llamada de 204 segundos.

## 1 Introducción

Los importantes avances en la tecnología del habla han permitido desarrollar sistemas automáticos para servicios finales por teléfono. Durante la última década, han aparecido gran cantidad de sistemas de información y reserva de billetes de viaje, como ARISE [1][2]. Otro proyecto es Darpa Communicator (<http://fofoca.mitre.org>), que ofrece información y reserva de billetes de avión, hotel y alquiler de coches. En este artículo proponemos una metodología para el diseño de gestores de diálogo en este tipo de sistemas. Esta propuesta es similar al modelo "Life-Cycle"[3][4], pero en nuestro caso incorporamos una fase de diseño por observación y presentamos medidas para evaluar las alternativas de diseño en cada fase.

## 2 Análisis de la base de datos y Diseño por intuición

El objetivo del análisis de la base de datos es describir mediante un diagrama Entidad-Relación E-R, la información disponible para ofrecer el servicio. En este diagrama, se definen

las principales entidades, sus atributos y claves (atributos que definen unívocamente un elemento dentro del conjunto entidad), y las relaciones entre entidades. Las entidades serán los objetivos que el sistema podrá ofrecer (partes del servicio: información de horarios de trenes, reserva, precios, etc.). Las claves son los datos obligatorios que el sistema deberá preguntar al usuario para acceder a un elemento concreto del conjunto entidad y, por tanto definen las interacciones mínimas que tendrá nuestro diálogo.

Posteriormente se debe realizar un diseño por intuición o "brain-storming" sobre el diagrama E-R, en el que se *proponen alternativas* acerca de los objetivos a cubrir por el sistema, su secuencia y los datos necesarios (p.e. destino, origen, fecha, etc.), para satisfacer cada uno de los objetivos. El resultado de este análisis debe ser una tabla con las alternativas propuestas.

## 3 Diseño por observación

En esta fase se analizarán diálogos reales de usuarios con operadores en un servicio similar, y se anotarán los eventos observados. Esta fase pretende medir y evaluar el impacto de cada una de las alternativas propuestas anteriormente.

### 3.1 Análisis de objetivos

a) *Qué objetivos son solicitados con más frecuencia por el usuario y en qué orden.*

	%	Orden				
		1º	2º	3º	4º	5º
<b>Horarios</b>	<b>64</b>	57	6	1	-	-
<b>Horarios para viaje de vuelta</b>	<b>20</b>	-	14	5	1	-
<b>Precios</b>	<b>46</b>	6	30	10	-	-
<b>Reservas</b>	<b>26</b>	14	4	2	3	3
<b>Frecuencia de trenes</b>	<b>2</b>	2	-	-	-	-
<b>Itinerario</b>	<b>14</b>	8	4	1	1	-
<b>Descuentos</b>	<b>5</b>	1	2	1	1	-
<b>Otros</b>	<b>12</b>	8	2	-	2	-

Tabla 1: Análisis de objetivos

El resultado de este análisis es el número de veces que se solicita cada objetivo y su posición. En la tabla 1 se presenta el análisis para 100 llamadas. En una misma llamada pueden aparecer uno o varios objetivos. En nuestro sistema final hemos considerado únicamente aquellos objetivos que aparecen en más del 10% de las llamadas, excepto *Itinerario* (no se dispone de información) y *Otros* (contiene varios objetivos pero ninguno llega al 10%).

b) *También se debe analizar, de igual forma, la información que el sistema debe ofrecer para satisfacer cada uno de los objetivos.*

### 3.2 Análisis de datos

En este análisis se debe:

- a) *Definir los datos necesarios para cubrir un objetivo y su secuencia en el diálogo.*
- b) *Clasificar cada dato como obligatorio u opcional:*
  - *Obligatorio:* dato imprescindible para satisfacer un objetivo. Este dato siempre debe ser preguntado al usuario (p.e. "¿Qué día desea viajar?"). Si el usuario no lo define, se le puede asignar un valor por defecto (p.e. "hoy") o se puede satisfacer el objetivo con información parametrizada por ese dato (p.e. para todos los días de la semana actual).
  - *Opcional:* el sistema nunca pregunta este dato, pero si el usuario lo especifica, el sistema lo utilizará.
- c) *Analizar las diversas formas de especificar un dato y su frecuencia de aparición.*

### 3.3 Análisis de la negociación

La "negociación" se produce cuando el usuario debe elegir una opción de entre varias ofertadas. En este caso debemos definir:

a) *Qué información le permite al usuario elegir la opción y su importancia.*

Criterio	%	Criterio	%
Hora salida	41.0	Nº de enlaces	2.6
Hora llegada	15.4	Enlace	2.6
Est. de Salida	2.6	Clase	10.3
Precios	7.7	Duración	5.0
Tipo de tren	12.8		

Tabla 3: Análisis del criterio de negociación.

El número de transbordos es un criterio importante, pero no es muy frecuente en las llamadas analizadas puesto que el operador

filtra la información y sólo ofrece viajes directos. Por otro lado, el tipo de tren es muy útil puesto que da una idea conjunta de precios y servicios ofrecidos en el tren.

b) *La estrategia de negociación.*

En la negociación podemos presentar la mejor opción de viaje y permitir que el usuario pregunte por la opción anterior/posterior, o presentar varias opciones a la vez y pedir al usuario que elija una de ellas. En este caso es muy importante analizar el número de opciones a considerar.

El diseño por observación permite analizar alternativas de diseño sin tener implementado ningún sistema, pero este análisis se realiza sobre interacciones usuario-operador, que son diferentes de las interacciones usuario-sistema. Por ello, podemos analizar aspectos generales del diálogo pero no características específicas de la interacción (p.e. cambios en la velocidad de locución, utilización de frases más o menos formales, etc.).

## 4 Diseño por simulación

En esta fase se analizan las características de las interacciones usuario-sistema, simulando el sistema con la herramienta de Mago de Oz. En esta evaluación, 15 usuarios llamaron al sistema para completar 6 escenarios y rellenar un cuestionario. Los elementos a evaluar así como las medidas de evaluación propuestas (Sistema y Cuestionario) se presentan a continuación:

### 4.1 Evaluación de objetivos

- a) *Para ello debemos validar la secuencia y cobertura de los objetivos considerados.*
  - Sistema: nº de veces que un objetivo es requerido, así como tiempo e interacciones utilizadas para satisfacerlo.
  - Cuestionario: nuevos objetivos sugeridos.

### 4.2 Evaluación de datos

Los elementos a evaluar en este caso son:

- a) *Diseño de preguntas y vocabulario de reconocimiento.*
  - Sistema: número de veces que el usuario se queda en silencio ante una pregunta, tasa de reconocimiento (si se dispone de una primera versión del reconocedor).
  - Cuestionario: facilidad de especificación.
- b) *Secuencia de datos.*

Se proponen diferentes secuencias posibles de datos y para cada una de ellas se evalúa:

  - Sistema: Tasa de reconocimiento de la secuencia de datos, calculada como el

producto de las tasas de cada dato por separado (cuando sea posible).

- **Cuestionario:** preferencia del usuario.
- c) *Cómo gestionar los datos obligatorios cuando no son especificados:* parametrizar la información o fijar un valor por defecto.
- **Cuestionario:** preferencia del usuario.

### 4.3 Negociación

En nuestro caso, presentamos varias alternativas de viaje y pedimos al usuario que elija una de ellas. Aleatoriamente, el sistema presenta 1, 2 ó 3 opciones, dependiendo del número total. Se definen diferentes patrones de información a presentar por opción y, aleatoriamente, se elige uno diferente para cada grupo de opciones.

- **Sistema:** nº de preguntas y tiempo.
- **Cuestionario:** qué información ayuda mejor al usuario a elegir y cual es el nº de opciones que se pueden recordar a la vez.

Análisis de la negociación (Cuestionarios)				
1 en 1		2 en 2		3 en 3
21.4%		21.4%		57.2%
Criterio de Negociación				
Tipo	Hora de Salida	Hora de Llegada	Precios	Duración
Tren	37.5%	25.0%	15.6%	6.2%

Tabla 5: Análisis de la negociación.

Los usuarios prefirieron disponer de 3 opciones a la vez (menor número de preguntas) pero el tiempo de negociación es mayor. En nuestro caso decidimos hacer una negociación de 3 en 3, pero reduciendo la información por opción a: "hora de salida/llegada" y "tipo tren".

Para una evaluación global se debe medir:

- **Sistema:** nº medio de interacciones, tiempo medio de llamada y tasa de reconocimiento.
- **Cuestionario:** rapidez del sistema, facilidad de aprendizaje, etc...

La herramienta de Mago de Oz permite analizar varios diseños del diálogo sin tener el sistema completo implementado.

## 5 Diseño por mejora iterativa

En la mejora iterativa, vamos a implementar la primera versión de un sistema completamente automático, mediante un proceso iterativo de prueba y corrección. Con la herramienta de Mago de Oz, se definió una primera versión del diálogo, pero en un sistema automático, debemos diseñar también, los mecanismos de confirmación de datos. Este diseño requiere de medidas de confianza en el reconocedor.

### 5.1 Estrategias de confirmación

Según el número de datos a confirmar tenemos:

- De un solo dato. ("¿Has dicho Madrid?")
- De varios datos a la vez. ("¿Desea ir de Madrid a Sevilla?")

Según la facilidad de corrección [5]:

- **Confirmación explícita:** el sistema utiliza una pregunta directa. ("Entiendo que desea salir de Madrid. ¿Es correcto?")
- **Confirmación implícita:** no se permite la corrección, simplemente se informa ("Desea salir de Madrid. ¿A dónde desea llegar? ")
- **Confirmación semi-implícita:** similar a la confirmación implícita, pero se permite corregir ("Entiendo que desea salir de Madrid, diga corregir si no es correcto, o diga la ciudad destino.").
- **Sin confirmación:** el sistema no informa del valor reconocido (p.e. preguntas de sí o no).
- **Rechazo de lo reconocido y repetición de la pregunta:** cuando la confianza es muy baja, el sistema rechaza lo reconocido y vuelve a preguntar ("Perdone, no le entendido muy bien. Diga la ciudad de origen.").

### 5.2 Medidas de confianza

El módulo de reconocimiento utilizado es un sistema por teléfono para gran vocabulario, capaz de reconocer palabras aisladas y expresiones sencillas como "el lunes", "la semana que viene" o "el catorce". Este reconocedor está basado en una arquitectura de hipótesis y verificación. Con las medidas de confianza propuestas en [6] sobre una base de datos con 2.204 ejemplos, el 39.1% de los errores son detectados (para un falso rechazo del 5%), reduciendo el error de clasificación de 15.8% a 14.0%.

### 5.3 Mecanismos de confirmación

Para diseñar los mecanismos de confirmación, es necesario representar las distribuciones de los errores y aciertos de reconocimiento según el valor de confianza y definir diferentes umbrales. En nuestro caso hemos definido 4 niveles (3 umbrales) de confianza (Figura 1):

1. **Nivel muy alto:** el número de palabras correctas es mucho mayor que el de errores.
2. **Nivel alto:** el número de palabras correctas es mayor que el de errores.
3. **Nivel bajo:** ambas distribuciones son similares.
4. **Nivel muy bajo:** hay más errores que aciertos.

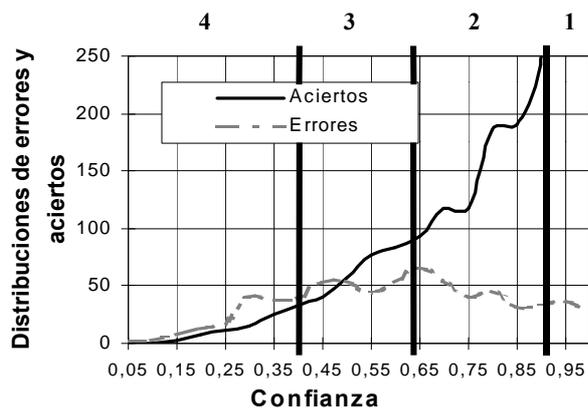


Figura 1. Distribuciones de errores y aciertos.

Para las ciudades origen y destino del viaje definimos NC(orig.) y NC(dest.) como los niveles de confianza (NC) obtenidos. Dependiendo de los NCs, definiremos la estrategia de confirmación:

- NC(orig.)=1 y NC(dest.)=1: *confirmación implícita de ambos datos.*
- NC(orig.)=2 o NC(dest.)=2: *confirmación explícita de ambos datos.*
- NC(orig.)=3 o NC(dest.)=3: *confirmación explícita de cada dato.*
- NC(orig.)=4 o NC(dest.)=4: *rechazo del dato.*

Si el reconocedor falla, el sistema pide al usuario que deletee el nombre de la ciudad [7]. No hemos considerado la confirmación semi-implícita por no ser bien aceptada por el usuario. Más adelante comentamos el comando CORREGIR que ofrece la misma funcionalidad sin alargar la pregunta.

#### 5.4 Recuperación frente a errores

Cuando en un sistema automático se utiliza la confirmación implícita, es necesario definir estrategias de recuperación de errores que permitan corregir suposiciones del sistema que puedan ser erróneas:

**VOLVER A EMPEZAR:** esta comando permite al usuario comenzar desde el principio. En nuestro sistema, en lugar de borrar todos los datos obtenidos, comenzamos confirmando explícitamente grupos de datos. Cuando uno de estos grupos no es confirmado, el sistema comienza desde ese punto:

S: "La opción elegida es un intercity..."  
 U: "Volver a empezar"  
 S: "Comencemos la consulta. ¿Desea ir de Madrid a Barcelona?"  
 U: "Sí"  
 S: "¿Desea viajar el 19 de Junio?"  
 U: "No"  
 S: "¿Cuándo desea viajar?"

**CORREGIR:** cuando el sistema comete un error y considera un dato reconocido como correcto (confirmación implícita), el usuario puede decir este comando y el sistema vuelve a preguntar el último dato introducido.

#### 5.5 Modelado de usuario

La técnica de modelado de usuario implementada está basada en [8]. En esta técnica se definen niveles de destreza del usuario en su interacción con el sistema (en nuestro caso hemos considerado 4 niveles). Dependiendo del nivel actual, las preguntas son más o menos explicativas (contienen más información sobre cómo debe contestar el usuario, posibles valores admitidos, etc...):

- **1<sup>er</sup> nivel.** La pregunta explica cómo interactuar con el sistema, el dato solicitado, los posibles valores y cómo especificarlos (p.e. período del día: "Hable después de escuchar la señal. Diga el período del día en el que desea viajar; por la mañana, por la tarde o por la noche.").
- **2<sup>do</sup> nivel.** La pregunta incluye el dato requerido, los valores aceptados y la manera de especificarlos ("Diga el período del día en el que desea viajar; por la mañana, por la tarde o por la noche.")
- **3<sup>do</sup> nivel.** Sólo se incluye el dato requerido. ("Diga el período del día en el que desea viajar.")
- **4<sup>to</sup> nivel.** El usuario conoce toda la información y podemos relajar la pregunta ("¿Cuándo desea salir?").

El nivel en cada instante depende del estado inicial y del nº de errores o confirmaciones positivas (aciertos) a lo largo de la interacción. En nuestro caso, el sistema comienza en el nivel 2 (después de ofrecer una ayuda general de cómo interactuar). Cuando ocurren errores (o aciertos) de reconocimiento, el sistema decrementa (o aumenta) el nivel. El nº de aciertos o errores que fuerzan un cambio de nivel depende del propio nivel actual. De esta manera el sistema se adapta dinámicamente a la destreza del usuario. Ejemplo:

[El sistema está en el nivel 3]  
 S: "Diga el período del día en que desea viajar."  
 U: "Después de comer"  
 [El sistema reconoce "por la noche"]  
 S: "¿Ha dicho por la noche?"  
 U: "No"  
 [El sistema pasa del nivel 3 al 2]  
 S: "Diga el período del día en que desea viajar; por la mañana, por la tarde o por la noche."  
 U: "Por la tarde"

## 6 Evaluación de campo

En esta evaluación, 30 personas llamaron al sistema para completar 4 escenarios (120 consultas). Las medidas de evaluación han sido obtenidas mediante anotaciones del sistema (tabla 6) y mediante las respuestas de los usuarios en un cuestionario (tabla 7).

Medida	Valor
Duración media de la consulta (seg.)	204
Número medio de interacciones	21.2
Número de "Volver a Empezar"	0.08
Número de "Corregir"	0.43
Duración de la negociación (seg.)	58

Tabla 6. Medidas obtenidas por el sistema.

El 35.4% de las llamadas pidieron información tanto para el viaje de ida como el de vuelta y el 31.5% completó la reserva (ida o ida-vuelta). El tiempo medio de consulta fue de 204 s, mayor que el tiempo medio para el sistema atendido por operador (152 s), pero similar a otros servicios automáticos [2].

Medida (escala de 1 a 5)	Valor
Experiencia del usuario en este tipo de sistemas.	1.8
El sistema comprende lo que digo.	3.6
Comprendo lo que el sistema dice.	4.5
Consigo la información rápidamente.	4.0
Es fácil de aprender.	3.9
Ante un error la corrección fue fácil.	3.1
Me pregunta en un orden lógico.	4.6
Generalmente, es un buen sistema.	3.9

Tabla 7. Medidas obtenidas del cuestionario

La inteligibilidad del sistema obtuvo una calificación de 4.5 gracias a la síntesis de voz de alta calidad desarrollada en [9]. El flujo del diálogo obtuvo la mejor calificación (4.6) debido al detallado análisis llevado a cabo.

## 7 Conclusiones

En este trabajo proponemos una nueva metodología para el diseño de gestores de diálogo en servicios automáticos por teléfono con reconocimiento de voz. Esta metodología ha sido aplicada con éxito para el desarrollo de un sistema de información ferroviaria. Los primeros pasos propuestos son el análisis de la base de datos (diagrama E-R) y el diseño por intuición en el que se realiza un "brainstorming" para proponer diferentes alternativas de diseño. En el diseño por observación, evaluamos cada propuesta mediante el análisis de transcripciones de diálogos usuario-operador.

El problema de la observación es que la interacción usuario-operador es diferente de la interacción usuario-sistema. La etapa de análisis por simulación resuelve este problema mediante la técnica de Mago de Oz. En el diseño por mejora iterativa, describimos la consideración de las medidas de confianza en la gestión de las confirmaciones, proponemos dos comandos para la recuperación frente a errores del sistema y además, describimos la técnica de modelado de usuario utilizada para adaptar dinámicamente el sistema a la destreza del usuario.

Con esta metodología hemos desarrollado un sistema automático con buena aceptabilidad por parte del usuario: valoración global de 3.9 y tiempo medio de llamada 204 seg.

## 8 Agradecimientos

Este trabajo ha sido financiado parcialmente con fondos Feder 2FD1997-1062-C02 (EU y CICYT). Los autores agradecen la colaboración de sus compañeros en GTH y RENFE.

## Referencias

- [1] Lamel, L., Rosset, S., Gauvain, J.L., Bennacef, S., Garnier-Rizet, H., Prouts, B., 2000. "The LIMSI ARISE system". *Speech Comm.* V. 31, pp 339-355.
- [2] Baggia, P., Castagneri, G., Danieli, M., 2000. "Field trails of the italian ARISE train timetable system". *Speech Comm.* V. 31, pp. 356-368.
- [3] Bernsen, N.O. Dykjaer, H. and Daykjaer, L. 1998 "Designing interactive speech systems. From first ideas to user testing" Springer Verlag.
- [4] [www.disc2.dk](http://www.disc2.dk)
- [5] Lavelle, A.C., Calmes, H., Pérennov, G., 1999. "Confirmation strategies to improve correction rates in a telephonic inquiry dialogue system". *Eurospeech'99.* V 3, pp. 1399-1402.
- [6] Macías-Guarasa, J., Ferreiros, J., San-Segundo, R., Montero, J.M., and Pardo, J.M., 2000. "Acoustical and Lexical Based Confidence Measures for a Very Large Vocabulary Telephone Speech Hypothesis-Verification System". *ICSLP'00.* Beijing, China.
- [7] San-Segundo, R., Colás, J., Ferreiros, J., Macías-Guarasa, J., Pardo, J.M., 2000. "Spanish Recognizer of continuously spelled names over the telephone". *ICSLP'00,* Beijing, China.
- [8] Veldhuijzen van Zanten, G., "User modelling in adaptive dialogue management". 1999. *Eurospeech'99,* Vol. 3, pp. 1183-1186.
- [9] Montero, J.M., Córdoba, R., Vallejo, J.A., Gutiérrez-Arriola, J., Enríquez, E., Pardo, J.M., 2000. "Restricted-domain female-voice synthesis in Spanish: from database design to a prosodic modeling". *ICSLP'00.* Beijing, China.