

# Sistema de diálogo multimodal basado en modelos estadísticos\*

## *Multimodal dialog system based on statistical models*

E. Sanchis, L-F. Hurtado, J.A. Gómez, F. García, J. Pastor, J. Planells, E. Segarra

Universitat Politècnica de València

Camí Vera sn, 46022 València

{esanchis, lhurtado, jon, fgarcia, jpastor, xplanells, esegarra}@dsic.upv.es

**Resumen:** En este trabajo presentamos un sistema de diálogo multimodal. Además de la multimodalidad de entrada y salida, la principal característica del sistema es que los módulos más importantes están basados en modelos estadísticos.

**Palabras clave:** sistema de diálogo, multimodalidad, modelos estadísticos

**Abstract:** In this paper, we present a multimodal dialog system. In addition to input and output multimodality, the main feature of the system is that its key modules are based on statistical models.

**Keywords:** dialog system, multimodality, statistical models

## 1. *Introducción*

El desarrollo de sistemas de diálogo hablado es un proceso complejo que implica el diseño, implementación y evaluación de un conjunto de módulos que tratan diferentes fuentes de conocimiento.

Los modelos estadísticos han sido utilizado ampliamente en reconocimiento de voz y comprensión del habla. A pesar de que en el caso de la gestión de diálogo la mayoría de las aproximaciones están diseñadas a mano, en los últimos años, se han desarrollado aproximaciones que utilizan modelos estadísticos también para la gestión de diálogo.

En el presente trabajo presentamos un sistema de diálogo multimodal donde los principales módulos están basados en el uso de modelos estadísticos aprendidos a partir de datos. Este sistema se utiliza en la tarea EDECAN-SPORTS que consiste en una tarea para la gestión de reservas de instalaciones deportivas en la Universidad.

## 2. *Sistema de diálogo multimodal*

La Figura 1 muestra el esquema del sistema de diálogo multimodal que presentamos. El sistema se ha implementado utilizando la arquitectura desarrollada en el proyecto SD-TEAM. Esta arquitectura permite la comunicación entre los módulos que forman el sis-

tema de diálogo incluso si están situados en ordenadores diferentes. Además de los módulos habituales de un sistema de diálogo hablado para interactuar con un sistema de información, nuestro sistema incorpora dos módulos adicionales encargados de la gestión de la multimodalidad tanto de entrada como de salida.

### 2.1. *Módulo de comprensión*

La principal característica de nuestra propuesta para la comprensión del habla es la integración de las restricciones sintácticas y semánticas en un único autómata estocástico de estados finitos.

Para aprender los modelos sintácticos y semánticos se necesita un corpus de frases segmentadas y etiquetadas. Cada frase del corpus debe ser segmentada y cada uno de sus segmentos debe tener una etiqueta semántica asignada. La etiqueta asignada a cada segmento representa una interpretación semántica de dicho segmento.

A partir del corpus segmentado y etiquetado se aprenden dos tipos de modelos de estados finitos: un modelo semántico general a partir de las secuencias de etiquetas semánticas asociadas a las frases, y para cada etiqueta semántica se aprende un modelo sintáctico con todos los segmentos asociados a esa etiqueta semántica.

A fin de realizar el proceso de comprensión, se genera un autómata global substituyendo los estados del autómata semántico por sus correspondientes modelos sintácticos.

\* Trabajo parcialmente subvencionado por el gobierno español con el proyecto TIN2008-06856-C05-02 y la Universitat Politècnica de València con el proyecto 20100982

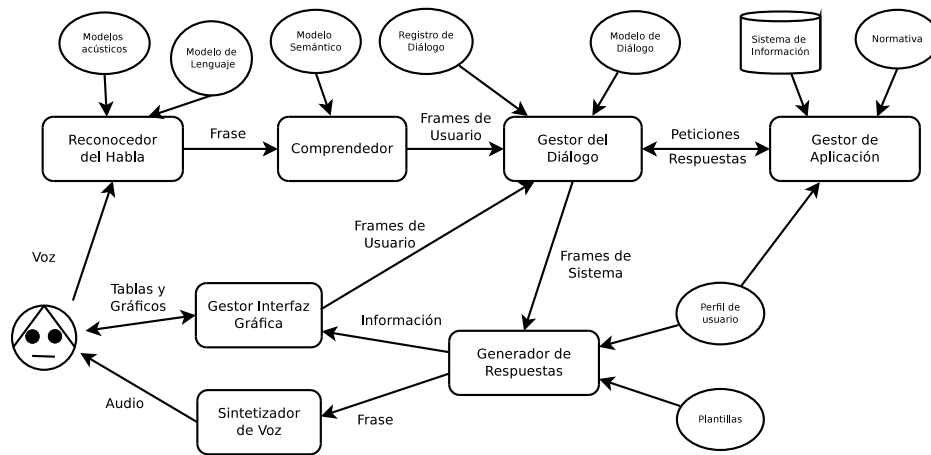


Figura 1: Arquitectura del sistema de diálogo multimodal

## 2.2. Gestor de Diálogo

En un sistema de diálogo, el Gestor de Diálogo (GD) es el módulo encargado de escoger la mejor respuesta del sistema en cada turno de diálogo. El modelo de diálogo propuesto en este trabajo se basa en el uso de un transductor de estados finitos. Dado un estado del sistema y una intervención del usuario, el GD genera una respuesta y se transita a un nuevo estado. La gestión de diálogo se basa, por tanto, en la modelización de secuencias de pares: intervención del usuario, respuesta del sistema. Un diálogo puede ser visto como un camino en el modelo del transductor de su estado inicial a un estado final en el cual el usuario ha alcanzado su objetivo.

El GD dispone de un modelo que cubre el comportamiento típico (más frecuente) del usuario. Sin embargo, durante un diálogo con un usuario real puede surgir una situación inesperada. Para abordar este problema, el GD es capaz de aumentar dinámicamente la cobertura del modelo de diálogo para cubrir esas situaciones no vistas en el entrenamiento y para las que en principio un GD no tendría respuesta. Cada vez que una de estas situaciones no vistas se produce, el GD genera un conjunto de subdiálogos sintéticos que comenzando en el estado no visto finalizan con éxito en un estado final. Estos subdiálogos se añaden al modelo permitiendo al GD continuar con el diálogo.

## 2.3. Interfaz multimodal

Una vez que el GD ha decidido cual será la respuesta del sistema, esta respuesta es enviada al Generador de Respuestas. El Generador de Respuestas se encarga de generar dos ti-

pos de respuestas distintos: respuesta oral y respuesta gráfica.

Para la generación de la respuesta oral se utiliza un conjunto de plantillas prediseñadas. Mediante estas plantillas se genera una frase que se envía al sintetizador de voz. El caso de la generación de la información gráfica es un poco más compleja. El primer lugar, la información suministrada por el GD es procesada para generar una serie de elementos gráficos. En segundo lugar, a cada uno de estos elementos gráficos se le añade información semántica. Por último, la información gráfica etiquetada semánticamente se envían al Gestor de la Interfaz Gráfica que será el encargado de mostrarla en la pantalla de táctil.

La multimodalidad de entrada se consigue interpretando semánticamente las pulsaciones del usuario sobre la pantalla táctil. El Gestor de la Interfaz Gráfica es el encargado de recuperar la información semántica asociada al elemento sobre el que pulsa el usuario y enviarla al Gestor de Diálogo.

## 3. Conclusiones

En este trabajo hemos presentado un sistema de diálogo multimodal basado en modelos estadísticos que ha sido aplicado a una tarea de reserva de instalaciones deportivas. El sistema muestra que es posible construir sistemas de diálogo basados en modelos estadísticos. Una ventaja del uso de este tipo de modelos radica en la posibilidad de adaptación dinámica, como hemos visto para el caso del Gestor de Diálogo. También se ha explorado la posibilidad de combinar la multimodalidad de entrada y salida.