Sistemas de interacción **hombre-máquina por voz**

En el Laboratorio de Procesamiento y Transmisión de Voz (LPTV) a cargo del profesor. Néstor Becerra Yoma, del Departamento de Ingeniería Eléctrica de la Facultad, se está desarrollando un trabajo con el sistema *Communicator* en colaboración con la Universidad de Colorado, USA, para desarrollar interfaces hombre-máquina conversacionales.

Este sistema *Communicator* es el más avanzado en su tipo y forma parte de un programa de la agencia DARPA (Defense Advanced Research Projects Agency, USA), que también financió el proyecto de redes de comunicaciones que dio origen a la Internet hace 40 años.

El proyecto *Communicator*, del cual participan el MIT, Carnegie Mellon University, Colorado University y el Artificial Intelligence Center at SRI International, está basado en la arquitectura Gallaxy II, contempla el acceso a base de datos en Internet mediante la red telefónica, y permite probar tecnologías tanto del punto de vista académico como práctico. La aplicación implementada por estos cuatro centros en Estados Unidos contempla la reserva de pasajes de avión a cualquier lugar del mundo, reserva de hotel y de alquiler de automóvil. La tasa de éxito, que es completar las tres tareas a través del diálogo conversacional, es de alrededor de 75%. Esto indica que en una aplicación real solo el 25% de los usuarios requerirían un(a) operador(a). En otras palabras, con el mismo número de operadores un *Call Center* podría atender cuatro veces más clientes maximizando así su eficiencia.

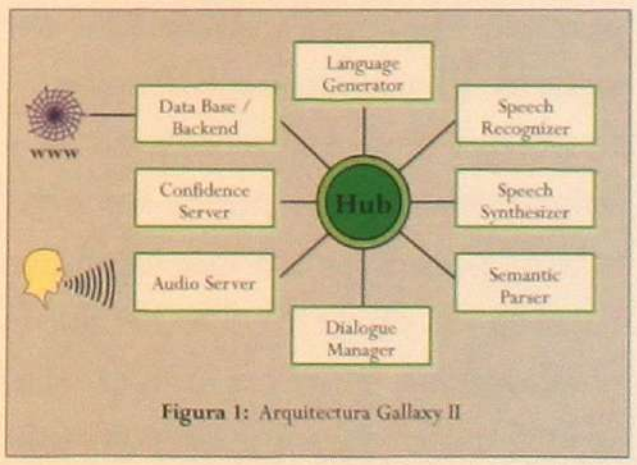

Figura 1: Arquitectura Gallaxy II

La arquitectura Gallaxy II está diseñada para proveer un ambiente accesible y comprensible para investigación y desarrollo de interfaces conversacionales hombre-máquina que utilicen tecnologías de diferentes proveedores. Así mismo, permite también el desarrollo de tecnologías específicas en ambientes muy similares a los de aplicaciones prácticas. Esta arquitectura (Figura 1) se desarrolló en el MIT. La idea es que todos los módulos, independientemente de la tecnología que utilicen, se comuniquen mediante el proceso Hub central empleando los protocolos de la arquitectura. El reemplazo de cualquier módulo por otro equivalente es perfectamente posible si mantiene el protocolo de comunicación (*plug-and-play*). Además del Hub, el sistema puede llegar a componerse de ocho elementos:

Servidor de Audio – Recibe la señal voz de una persona que esté llamando a través de la línea telefó-

nica, y envía la voz sintetizada en respuesta a las solicitudes del usuario.

Reconocedor de Voz – Toma la señal de voz que viene de la línea telefónica y reconoce lo que la persona ha dicho. Este procedimiento se le puede considerar conversión voz-a-texto.

Parser Semántico – Toma el resultado del reconocedor de voz y extrae la información relevante para la aplicación. Por ejemplo, si la aplicación corresponde a un sistema de reserva de pasajes de avión, el parser buscará por palabras claves que corresponden a origen, destino, fecha, etc.

Conductor del Dialogo – Conduce el diálogo de modo que pueda obtener toda la información que necesite para realizar la consulta a la base de datos local o remota vía Internet. En el caso de reserva de pasajes de avión, el conductor del diálogo pedirá que se le pregunte al usuario cuántas veces sea necesario para completar la información requerida por la aplicación. El procedimiento que utiliza es similar al de completar los campos de un formulario (*filling-form*).

Confidence Server – Debido a que el usuario puede sentirse confundido a la hora de responder preguntas del sistema, por no entender bien lo que la máquina le dice, es posible que el diálogo no avance como lo esperado. Para evitar estas situaciones se contempla un módulo cuya finalidad es tratar de evaluar si el usuario está respondiendo de acuerdo a lo previsto. Si se detecta que el usuario no responde la información deseada por el sistema se puede re-iniciar el diálogo y obligar al usuario a ser más específico en su respuesta para reducir la posibilidad de error del reconocedor de voz.

Base de datos/Aplicación – Recibe solicitudes de acceso a bases de datos locales o vía Internet. Estas solicitudes las envía el módulo Conductor del Dialogo.

Generador de Lenguaje Natural – Construye sentencias o frases inteligibles en el idioma local a partir de abreviaciones y códigos enviados por el Conductor de Diálogo o la Base de Datos/Aplicación. Estas sentencias serán posteriormente convertidas a una señal acústica oíble por el sintetizador de voz o convertidor texto-a-voz (TTS – *Text-to-speech*).

Sintetizador de Voz (TTS – *Text-to-speech*) – Recibe frases del Generador de Lenguaje Natural y genera la señal de voz correspondiente que es enviada por la línea telefónica al usuario.

El sistema *Communicator* que se desarrolla en el LPTV del Departamento de Ingeniería Eléctrica de la Universidad Chile está destinado a realizar investigaciones avanzadas sobre tecnologías de procesamiento de voz (reconocimiento, síntesis y verificación y locutor) y sobre sistemas de diálogos conversacionales del punto de vista de la aceptabilidad del usuario. El sistema también permite probar tecnologías existentes y aplicaciones lo que puede ser interesante para empresas que estén trabajando en el sector. De hecho, la colaboración empresa-universidad se está transformando en una realidad cada vez más concreta y se le está considerando como una pieza clave en el desarrollo tecnológico del país, afirmó el profesor Néstor Becerra.

"El sistema implementado en la Facultad es la reserva de pasajes aéreos en vuelos domésticos. La aplicación utilizó una configuración *Wizard of Oz* (Mago de Oz) para simular un sistema de reconocimiento de voz con 100% de exactitud. Esto se hizo para evaluar el sistema como un todo y un sintetizador de voz independientemente de los posibles errores de un reconocedor de voz real. El sistema se probó con 30 alumnos los que mostraron una reacción bastante positiva con relación a estos tipos de sistemas", explicó el profesor Becerra.

Otras investigaciones realizadas en el LPTV están relacionadas con la transmisión de voz en Internet y con el estudio experimental de redes de comunicaciones modernas. Estos temas están en el ámbito de la convergencia entre las telecomunicaciones y las tecnologías de la información, y tienen un gran interés tanto teórico como aplicado.

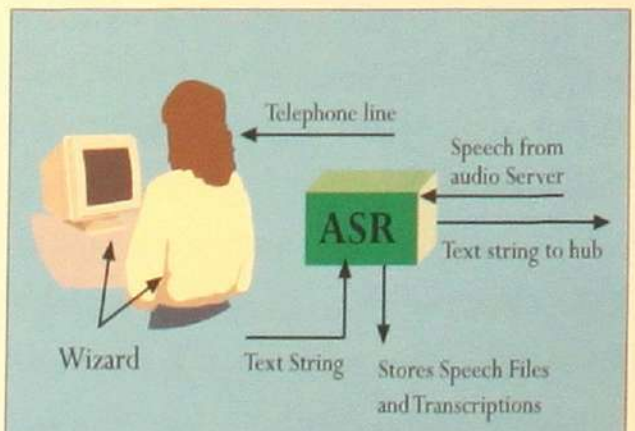

Figura 2: Configuración *Wizard of Oz* para reemplazar un reconocedor de voz (ASR-*Automatic Speech Reconizer*). Para el usuario es transparente si se está utilizando un reconocedor real o un wizard (mago).