

# *Hacia una educación inclusiva y personalizada mediante el uso de los sistemas de diálogo multimodal*

**David Griol, Ismael Baena, Víctor Corrales, José Manuel Molina, Araceli Sanchis de Miguel**

Universidad Carlos III de Madrid

Avda. de la Universidad, 30 – 28911 Leganés (Madrid)

{david.griol, josemanuel.molina, araceli.sanchis}@uc3m.es; {100065131, 100048294}@alumnos.uc3m.es

**Resumen:** Los continuos avances en el desarrollo de tecnologías de la información han dado lugar actualmente a la posibilidad de acceder a los contenidos educativos desde cualquier lugar, en cualquier momento y de forma casi instantánea. Sin embargo, la accesibilidad no es siempre considerada como criterio principal en el diseño de aplicaciones educativas, especialmente para facilitar su utilización por parte de personas con discapacidad. Diferentes tecnologías han surgido recientemente para fomentar la accesibilidad a las nuevas tecnologías y dispositivos móviles, favoreciendo una comunicación más natural con los sistemas educativos. En este artículo se describe el uso innovador de los sistemas de diálogo multimodales en el campo de la educación, con un especial énfasis en la descripción de las ventajas que ofrecen para la creación de aplicaciones educativas inclusivas y adaptadas a la evolución de los estudiantes.

**Palabras clave:** Sistemas de Diálogo Hablado, Sistemas Multimodales, Interacción Oral, Educación, E-learning.

**Abstract:** Continuous advances in the development of information technologies have currently led to the possibility of accessing learning contents from anywhere, at anytime and almost instantaneously. However, accessibility is not always the main objective in the design of educative applications, specifically to facilitate their adoption by disabled people. Different technologies have recently emerged to foster the accessibility of computers and new mobile devices favouring a more natural communication between the student and the developed educative systems. This paper describes innovative uses of multimodal dialog systems in education, with special emphasis in the advantages that they provide for creating inclusive applications and adapted to the students specific evolution.

**Key words:** Spoken Dialog Systems, Multimodal Systems, Spoken Interaction, Education, E-learning.

## **1. Introducción**

El auge de las tecnologías de la información ha propiciado la posibilidad cada vez mayor de acceder a la información y servicios en la red desde cualquier lugar, en cualquier momento y de forma casi instantánea. Los avances tecnológicos han favorecido además la creación de dispositivos con tamaños cada vez más reducidos, capaces de ejecutar nuevas aplicaciones y acceder a los datos mediante conexiones inalámbricas, como por ejemplo, los tablets-PC y los teléfonos inteligentes, utilizados ampliamente hoy en día para el acceso a la web.

Sin embargo, en el diseño de estas aplicaciones y uso de estos dispositivos no siempre suele tenerse en cuenta el objetivo fundamental de maximizar la accesibilidad, especialmente para facilitar el acceso a personas con discapacidades. Para cumplir este objetivo han surgido recientemente diversas tecnologías, entre ellas, los sistemas de diálogo multimodales [Pieraccini 12] [Wahlster 06] [López-Cózar y Araki 05] [McTear 04]. Estos sistemas pueden definirse como programas informáticos diseñados para emular la capacidad de comunicación de un ser humano utilizando diversas modalidades de comunicación, entre ellas, el habla.

Disponiendo de un interfaz hablado, se libera al usuario de utilizar otros canales tradicionales como el teclado, el ratón o la pantalla. Dado este interfaz vocal, el número de entornos y tareas en los que pueden aplicarse estos sistemas es enorme: sistemas que proporcionen información para la planificación de viajes [Callejas y López-Cózar 08] [Griol et al. 08] [Bohus y Rudnicky 05] [Andeani et al. 06], sistemas de información meteorológica [Zue et al. 00], servicios de banca electrónica [Melin et al. 01], aplicaciones accesibles desde los vehículos [Hofmann et al. 12] [Weng et al. 06] [Jokinen et al. 04] [Mattasoni et al. 02], sistemas orientados a personas con discapacidades y aplicaciones médicas [Bickmore et al. 13] [Callejas et al. 07], control remoto de dispositivos y robots especialmente en entornos inteligentes [Lutfi et al. 13] [Augusto 09] [Montoro et al. 06] [Ábalos et al. 06], portales de voz [Griol y García-Jiménez 12], resolución de tareas colaborativas [Nestorovič 12], agentes y compañeros virtuales [Corradini et al. 05] [Catizone et al. 03] [Hubal et al. 00], etc.

Con el desarrollo de las Tecnologías del Habla, las posibilidades de integrar estos sistemas para el desarrollo de aplicaciones educativas están recibiendo una atención creciente, incluyendo aplicaciones dentro del ámbito de la tutorización [Kopp et al. 12] [D’Mello et al. 12] [Latham et al. 12] [Pon-Barry et al. 06], aplicaciones del tipo pregunta-respuesta (*question answering*) [Wang et al. 07], la práctica de la pronunciación en el aprendizaje de idiomas [Fryer et al. 06], agentes pedagógicos [Segedy et al. 13] [Pareto et al. 12] [Cavazza et al. 10], y sistemas de diálogo para promover la reflexión y las habilidades metacognitivas [Kerly et al. 08a].

Este artículo se centra la descripción del potencial educativo de los sistemas de diálogo, que puede observarse por el gran número y variedad de aplicaciones prácticas que se han desarrollado durante los últimos años. Asimismo, se describen los esfuerzos actuales para desarrollar interfaces educativos multimodales, inteligentes, adaptativos, proactivos, portátiles y afectivos. Todas estas características no son mutuamente excluyentes, por ejemplo, la inteligencia del sistema puede atribuirse a su adaptación para permitir una mejor portabilidad a diferentes entornos de interacción.

Para tratar los aspectos más importantes relativos al diseño de interfaces multimodales educativos, el

artículo se estructura del siguiente modo. La Sección 2 proporciona una visión general de los principales módulos y funcionalidades necesarios para el desarrollo de sistemas de diálogo hablado. La Sección 3 describe los principios fundamentales que intervienen en el desarrollo de interfaces multimodales educativos. Esta sección también proporciona importantes ejemplos que muestran los beneficios de la integración de este tipo de sistemas en aplicaciones educativas. La Sección 4 describe nuestro trabajo en este campo a través del análisis de dos sistemas desarrollados para facilitar, respectivamente, el acceso a Internet y aprender lenguas extranjeras. Finalmente, la sección 5 presenta las conclusiones y líneas de trabajo futuro.

## 2. Sistemas de diálogo: Arquitectura y procesos fundamentales

El diccionario de la Real Academia Española de la Lengua define el diálogo como la plática entre dos o más personas que alternativamente muestran sus ideas o afectos. La acción de conversar es uno de los modos más natural para resolver un gran número de acciones cotidianas entre los seres humanos. Por este motivo, un interés histórico dentro del campo de las Tecnologías del Habla ha sido utilizar estas tecnologías en aplicaciones reales, especialmente en aplicaciones que permitan a una persona utilizar su voz para obtener información mediante la interacción directa con una máquina o para controlar un determinado sistema.

Construir una aplicación informática que pueda mantener una conversación con una persona de manera natural sigue siendo hoy en día un reto, dada la gran cantidad de fuentes de conocimiento que son necesarias y las limitaciones de las tecnologías utilizadas para obtener información del usuario. No obstante, los constantes avances de la investigación en Tecnologías del Habla han permitido que sean factibles actualmente sistemas de comunicación persona-máquina mediante la voz, capaces de interactuar con cierto grado de flexibilidad (iniciativa mixta en el desarrollo del diálogo). Estos sistemas están siempre orientados a tareas de información muy específica (dominios semánticos restringidos).

Debido al gran número de operaciones que deben realizar estos sistemas, es habitual un desarrollo modular de los sistemas de diálogo hablado, lo que

permite desglosar las dificultades entre los diferentes componentes del sistema. Un sistema de estas características se puede describir en términos de los siguientes módulos: Módulo de Reconocimiento Automático del habla, Módulo de Comprensión del Habla, Gestor de Diálogo, Módulo de Consulta a las Bases de Datos de la Aplicación, Módulo de Generación de Respuestas y Sintetizador de Texto a Voz.

El Módulo de Reconocimiento Automático del Habla [Rabiner et al. 96], reconoce la señal vocal pronunciada por el usuario y proporciona la secuencia de palabras reconocida más probable (o las  $k$  más probables). Se trata de una tarea muy compleja dado la gran diversidad de factores que pueden afectar a la entrada, básicamente relativos al locutor, el contexto de la interacción y el canal de comunicación.

El Módulo de Comprensión del Habla [Minker 98], a partir de las(s) secuencia(s) de palabra(s) reconocida(s), el sistema obtiene una representación semántica de su significado. Existen dos aproximaciones fundamentales para el desarrollo de este módulo: modelos basados en reglas [Mairesse et al. 09] y modelos estadísticos [Meza et al. 08].

El Gestor de Diálogo, considera la interpretación semántica de la petición del usuario, la historia del diálogo, la información de la aplicación disponible en ese punto y el estado del sistema, y determina la siguiente acción que debe tomar el sistema siguiendo la estrategia del diálogo. El modelo de diálogo más simple se basa en máquinas de estados finitos, en la que los estados de la máquina representan los estados del diálogo y las transiciones entre los estados se determinan por las acciones del usuario.

La aproximación basada en *frames* tiene como principal objetivo superar la falta de flexibilidad de los modelos de estados finitos y se utilizan en los sistemas comerciales más actuales [Chu et al. 05]. En la actualidad, la aplicación de técnicas de aprendizaje automático en la gestión del diálogo es un área de investigación muy activa [Griol et al. 08] [Williams y Young 07].

El Módulo de Consulta a la Base de Datos de la Aplicación, recibe peticiones de consulta a la base de datos por parte del gestor de diálogo, las procesa y devuelve el resultado al gestor. Más recientemente, este módulo debe poseer la capacidad de acceder a

diferentes tipos de repositorios de información en la red (como páginas web, servicios web, etc.).

El Módulo de Generación de Respuestas, recibe la respuesta del sistema en forma de cierta representación formal y tiene como función la generación de una frase, gramaticalmente correcta y en un lenguaje lo más cercano posible al lenguaje natural, que transmita el mensaje generado por el gestor de diálogo [Reiter 95]. La respuesta del sistema proporcionada por el generador de respuestas puede incorporar otras modalidades de información (texto, imágenes, vídeo, tablas con datos, gestos a reproducir por un avatar, etc.).

Finalmente, el Sintetizador de Texto a Voz, recibe la respuesta del sistema como texto en lenguaje natural y genera la correspondiente señal de audio, que será la respuesta que llegará al usuario.

La Figura 1 muestra la arquitectura modular descrita para el desarrollo de sistemas de diálogo hablado. En cuanto a las metodologías utilizadas para el desarrollo de los diversos módulos que componen el sistema de diálogo.

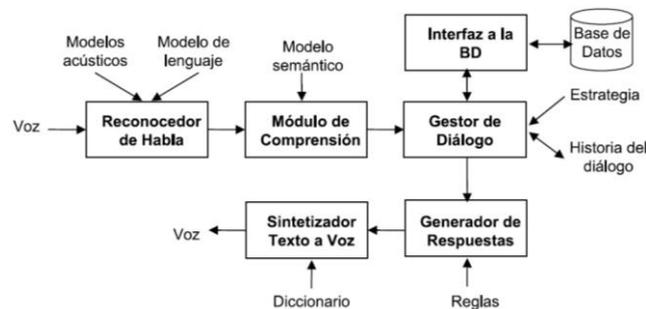


Figura 1. Arquitectura modular de un sistema de diálogo hablado

### 3. Sistemas de diálogo y educación

El diseño, implementación y conjunto de funcionalidades de los sistemas de diálogo empleados en aplicaciones de e-learning es muy diverso. La interacción con el usuario suele llevarse a cabo a través de sencillos formularios en formato texto [Heffernan 03], utilizados por los usuarios para proporcionar las respuestas a las preguntas escribiendo directamente en un teclado. Algunos sistemas utilizan interfaces personificados [Graesser et al. 01], capaces de transmitir emociones y gestos, mientras que otros utilizan un avatares más sencillos

[Kerly et al. 08b]. La incorporación de la interacción oral, mediante el uso de reconocedores de voz para las entradas del usuario y de sintetizadores de texto a voz para las salidas, es cada vez más viable [Graesser et al. 01] [Litman y Silliman 04] [Bos et al. 02].

De acuerdo con [Roda et al. 01a], se espera que los sistemas de e-learning avanzados i) aceleren el proceso de aprendizaje, ii) faciliten el acceso a la educación, iii) personalicen el proceso de aprendizaje, y iv) provean un entorno de aprendizaje más significativo. Además, se describen tres aproximaciones principales para integrar los sistemas de diálogo en entornos educativos tanto a nivel individual como del grupo:

1. Herramientas para facilitar los procesos de aprendizaje y una ayuda avanzada;
2. Entrenadores personales provistos del conocimiento específico del dominio de interacción;
3. Actores en entornos de aprendizaje simulados y juegos de roles.

Los sistemas de diálogo para facilitar los procesos de aprendizaje y una ayuda avanzada integran modelos de conocimiento estructurados sobre el dominio de la aplicación y el entorno de interacción. Aunque este tipo de sistemas puede ser muy útil, por lo general se consideran molestos y no inteligentes. Sus principales limitaciones se deben a su incapacidad para contextualizar las acciones de los usuarios en el conjunto de posibles usos de la aplicación informática. De esta manera, el sistema de diálogo no tiene ningún conocimiento relacionado con el usuario y no se puede adaptar de acuerdo a sus preferencias y motivaciones.

Además, por lo general, este tipo de sistemas no incluyen funcionalidades de diálogo y sólo pueden utilizarse para resolver cuestiones aisladas, proporcionando información que suele estar ya presente en el manual de ayuda de la aplicación. Por ello, estos sistemas no incorporan ningún modelo de aprendizaje. Ejemplos de este tipo de sistemas son los agentes conversacionales de Microsoft integrados en las aplicaciones de escritorio del paquete Office, así como agentes similares que están disponibles en muchas aplicaciones de base de datos.

Los sistemas de diálogo como entrenadores personales integran información sobre el dominio de

la aplicación. Los sistemas de este tipo se caracterizan por la posibilidad de representar y actualizar continuamente la información que representa el estado cognitivo y social de los usuarios. El objetivo principal es orientar a los usuarios en el proceso de aprendizaje, ofreciéndoles sugerencias y otras funcionalidades no sólo con la aplicación desarrollada, sino también con el resto de estudiantes. Para ello, estas aplicaciones suelen integrar interfaces realistas e interactivas.

K-Inca es un agente conversacional artificial diseñado para ayudar a la asimilación de actividades orientadas a la gestión del conocimiento [Angehrn et al. 06] [Roda et al. 01b]. El sistema utiliza perfiles de los usuarios para calcular su similitud con un conjunto de perfiles predefinidos, proporcionando un servicio personalizado, por ejemplo, mediante ejercicios específicos, sugerencias y estímulos que modifican la forma con la que el sistema interactúa con los estudiantes.

Los sistemas de diálogo como actores en entornos de aprendizaje simulados y juegos de roles son capaces de llevar a cabo una función específica de una manera muy realista dentro de un entorno simulado que emula el entorno de aprendizaje real. Estos sistemas integran los conocimientos sobre este entorno y el dominio de aplicación (tareas, comportamientos, objetos, relaciones, etc.) y son capaces de mantener un diálogo adaptando este conocimiento al considerar el estado social y cognitivo de los usuarios.

El sistema *Change VIBE* (C-VIBE) [Angehrn 01] se desarrolló para interactuar en el entorno simulado EIS [Manzoni y Angehrn 97], que se utiliza actualmente para formar a directivos en la teoría y práctica de la necesidad de cambio y transformación de las organizaciones frente a la resistencia natural a la innovación. Los estudiantes interactúan en esta plataforma a través de avatares. El objetivo principal es gestionar con éxito, individualmente o en grupo, una misión específica que se propone en una sala de juntas virtual.

### 3.1 Sistemas aplicados a la tutorización

La tutoría es una de las áreas de investigación más importantes para la incorporación del diálogo en lenguaje natural en aplicaciones de e-learning. Trabajos como los desarrollados por [Kumar et al. 11] han demostrado que los agentes que emulan el papel de un tutor en un ambiente de aprendizaje

colaborativo pueden proporcionar una mejora significativa del proceso de aprendizaje. Trabajos adicionales, como el desarrollado por [Liu y Chee, 04], han explorado una variedad de modelos de interacción y tácticas que podrían utilizarse en entornos educativos con múltiples participantes.

La mayor parte de la investigación existente sobre estrategias de interacción para sistemas de diálogo utilizados en este campo se centra en aquellas relacionadas con aspectos de la tarea. En el caso de los tutores conversacionales, la definición de esta estrategia incluye aspectos como instruir a los estudiantes acerca de la tarea, la elección de la forma más adecuada con la que proporcionar la respuesta (por ejemplo, mediante pistas), la realimentación y otras estrategias [Graesser et al. 05].

Algunos estudios [Rose et al. 01] [Wang y Johnson, 08] han evaluado el efecto de estas estrategias en relación con el comportamiento conversacional. Asimismo, la computación afectiva y su aplicación al diálogo tutorial se centra en la identificación de los estados emocionales de los estudiantes [D'Mello et al. 08] y su utilización para mejorar la elección de acciones de los tutores.

El proyecto AutoTutor [Graesser et al. 99] proporciona diálogos tutoriales sobre temáticas como la alfabetización informática y la física. Las técnicas de tutorización empleadas por este sistema ayudan a los estudiantes en la construcción activa del conocimiento, y se basan en un análisis detallado de sesiones de tutores reales. La evaluación de este sistema mostró que proporcionaba una mejora importante en tareas relacionadas con la memoria y el tiempo requerido para el aprendizaje.

Otro sistema de tutorización que emplea el diálogo es Ms. Lindquist [Heffernan 03]. Este sistema proporciona un aprendizaje basado en ejemplos en lugar de ofrecer una instrucción explícita. Los primeros trabajos con el sistema sugirieron que el diálogo era beneficioso para mantener la motivación del estudiante. Los autores concluyeron que Ms. Lindquist era un ejemplo de aproximación del tipo "Menos es Más", donde los alumnos abordan menos problemas, pero aprenden más en cada uno de ellos al estar motivados en un diálogo inteligente [Heffernan 03].

CycleTalk [Forbus et al. 99] es un sistema tutorial inteligente que ayuda a los estudiantes a aprender los

principios de ciclos termodinámicos en el contexto del diseño de una planta de energía. Como parte de esta tarea de aprendizaje, los estudiantes participaron de forma colaborativa durante 30-45 minutos en ConcertChat, un entorno de colaboración basado en texto [Mühlpfordt y Wessner 05].

ITSPOKE es un sistema de diálogo hablado aplicado a la tutorización y enfocado en proporcionar una retroalimentación y corregir ideas erróneas [Litman y Silliman 04]. Utiliza una interacción basada en texto para la tutoría de conceptos en el campo de la Física<sup>1</sup>.

Otro ejemplo de tutoría en lenguaje natural es el sistema Geometry Explanation Tutor [Alevan et al. 04], donde los alumnos explican sus respuestas a problemas de Geometría utilizando sus propias palabras. El sistema utiliza un enfoque basado en el conocimiento para reconocer explicaciones correctas o parcialmente correctas, y un clasificador estadístico cuando el método basado en el conocimiento falla. Los estudios realizados con este sistema probaron que los estudiantes que utilizaban el diálogo para la resolución de problemas paso a paso (relativos a teoremas y definiciones en Geometría) aprendían mejor que aquellos que utilizaban menús predefinidos.

El agente conversacional Oscar CITS (*Conversational Intelligent Tutoring System*) [Latham et al. 12] tiene como objetivo simular un tutor humano que modela implícitamente el estilo de aprendizaje durante la tutoría, personalizando este proceso para aumentar la confianza y mejorar la eficacia de las actividades de aprendizaje. El sistema utiliza el lenguaje natural para proporcionar una comunicación sobre temas específicos, prediciendo de forma dinámica y adaptándose a la evolución del aprendizaje de los estudiantes. Se basa en el modelo de Índice de Estilos de Aprendizaje (ILS, *Index of Learning Styles*) [Felder y Silverman 88] para ofrecer un tutorial de SQL.

Kim (2007) describe un sistema de diálogo educativo para actividades de e-learning también en el campo de la Geometría. El conocimiento del sistema se representa mediante el lenguaje AIML (*Artificial Intelligence Markup Language*), basado en XML. El sistema puede responder a las preguntas de los

<sup>1</sup> Puede consultarse un listado de proyectos adicionales desarrollados en la Universidad de Pittsburgh en <http://www.cs.pitt.edu/~litman/itspoke.html>

estudiantes haciendo referencia al conocimiento previo que dispone de él/ella gracias a interacciones anteriores. Para ello, considera información contextual relacionada con el estudiante mediante la definición de modelos de comportamiento.

Gustafson et al., (1998) describen un entorno educativo desarrollado para un sistema de diálogo hablado modular. El objetivo del entorno es instruir a los estudiantes, con diferentes niveles de conocimiento, sobre el comportamiento de los sistemas de diálogo hablado. El diálogo se modela mediante un árbol cuyos nodos describen las acciones del sistema. El sistema educativo se ha utilizado en cursos en varias universidades de Suecia.

Otros sistemas de tipo pregunta-respuesta incluyen un panel de discusión [Feng et al. 06], donde el sistema de diálogo proporciona la siguiente respuesta comparando la situación actual del diálogo con aquellas que dispone en un corpus de aprendizaje, seleccionando aquella que es más similar a la actual de acuerdo con un criterio predefinido.

### 3.2. Compañeros de aprendizaje y sistemas personificados

Diferentes estudios han concluido que el desarrollo de sistemas con interfaces parecidos a los seres humanos mejoraba las relaciones que se establecían con los estudiantes [Dehn y van Mulken 00]. Los compañeros de aprendizaje (*learning companions*) son personajes simulados que actúan como un compañero del estudiante dentro de un entorno de aprendizaje social [Chou et al. 03]. Por este motivo, se han desarrollado un gran número de sistemas de diálogo personificados (ECA, *Embodied Conversational Agents*) para ayudar a los estudiantes durante el proceso de aprendizaje [Gratch et al. 02], actuando como consejeros [Marsella et al. 03], entrenadores personales [Bickmore 03], asesores para una vida saludable [de Rosis et al. 03], etc.

En efecto, los ECAs tienen el potencial de involucrar a los usuarios en una conversación similar a la humana, incluyendo señales verbales y no verbales para proporcionar retroalimentación, mostrar empatía e incluso un comportamiento emocional [Cassell et al. 01]. Debido a estas características, este tipo de sistemas proporcionan ventajas en el ámbito pedagógico especialmente cuando es importante establecer una relación con el estudiante en un

periodo largo de interacción [Johnson et al. 04] [Bickmore y Picard, 05].

Estos agentes se han aplicado en dominios tan diversos como ayudar a niños a aprender Biología [Lester et al. 99], aplicaciones médicas [Shaw et al. 99], o el entrenamiento en tácticas navales [Rickel y Johnson 99]. Baylor y Kim (2005) describen un estudio sobre los diferentes roles que pueden representar este tipo de agentes en esta variedad de aplicaciones prácticas.

En algunos sistemas, es el estudiante quien debe enseñar al agente [Chan y Baskin 98], o interactuar además con compañeros virtuales [Dillenbourg y Self 92], que pueden incluso tomar el papel de alborotadores destinados a provocar la disonancia cognitiva para impulsar el aprendizaje [Aimeur et al. 97]. Los investigadores también han estudiado las preferencias del usuario con respecto al nivel de experiencia del agente. Los resultados sugieren que, en general, las similitudes de competencia entre el agente y el alumno tienen impactos positivos en los logros de los estudiantes, de tal manera que los estudiantes académicamente fuertes mostraron mayores creencias de eficacia después de trabajar con un agente de alta capacidad, mientras que los estudiantes académicamente débiles mostraron una mayor eficacia después de trabajar con un agente de menor capacidad [Kim 07].

Por otra parte, aunque los ECAs han demostrado tener buenas prestaciones para establecer una relación empática con el usuario [de Rosis et al. 05] [Cassell et al. 00] [Ai et al. 06] [Bailly et al. 10] [Edlund et al. 08], es difícil comunicarse con estos agentes cuando el estudiante no está delante de un ordenador, pero requiere sugerencias y consejos de una forma continuada. Por ejemplo, el sistema DESIA [Johnson et al. 04] es un paso en esta dirección. Este agente ha sido adaptado para funcionar con éxito en dispositivos móviles y proporcionar un apoyo continuo al estudiante [Marsella et al. 03].

Otro ejemplo es la arquitectura VU-MAS (*Virtual University Multi-Agent System*) [De Carolis et al. 06]. Cada alumno puede interactuar con VU-MAS con un agente personal, llamado *MyCoach*, y representado mediante un ECA. El objetivo principal de este agente es monitorizar las actividades de los estudiantes, siguiendo sus mejoras y seleccionando materiales útiles de acuerdo a los objetivos y

necesidades que se hayan reconocido. El agente también es capaz de proporcionar sugerencias útiles de forma proactiva. Al estar diseñado para funcionar en teléfonos inteligentes o en PDAs, este agente supone un ejemplo muy interesante de actividades del tipo m-learning<sup>2</sup>.

### 3.3 Aprendizaje social y otras aplicaciones

El aprendizaje social permite a los estudiantes interactuar con otras personas, como maestros o compañeros de clase (tanto reales como simulados) durante las actividades de aprendizaje. Diferentes trabajos sugieren la posibilidad de integrar los sistemas de diálogo en comunidades de aprendizaje 2.0 mediante la creación de asistentes, moderadores o compañeros virtuales dentro de la comunidad [Chou et al. 03]. Las características de diálogo y antropomorfias de los sistemas de diálogo pedagógicos pueden, de este modo, enfatizar la dimensión social de las actividades educativas e incrementar la motivación por el conocimiento [Chou et al. 03].

Los sistemas de diálogo también pueden reforzar el aprendizaje en actividades específicas, en lugar de aplicarse a la enseñanza de nuevos conceptos. De este modo, pueden ayudar a los estudiantes a estimar su conocimiento acumulado mediante preguntas relativas a interacciones anteriores con la herramienta. También se pueden llevar a cabo evaluaciones formativas, que puedan utilizarse para generar una retroalimentación tanto a los estudiantes como al personal docente.

También se han desarrollado sistemas de diálogo para ofrecer la práctica de conversación para estudiantes de idiomas. En este ámbito de aplicación, Jia (2002) concluyó que los usuarios estaban satisfechos con las respuestas proporcionadas por un sistema básico del tipo ALICEbot<sup>3</sup>, aunque el reconocimiento de las posibles acciones de los usuarios se consideró insuficiente para su uso generalizado como entorno de práctica de un nuevo idioma. No obstante, el sistema Jabberwacky<sup>4</sup> [Fryer y Carpenter 06] utiliza una tecnología muy diferente a ALICEbots, aprendiendo de forma interactiva a partir de sus conversaciones anteriores y sugiriéndose para la

práctica avanzada de idiomas. Fryer y Carpenter destacaron que los sistemas tipo *chatbot* ofrecen a los estudiantes la posibilidad de utilizar un gran variedad de estructuras gramaticales y vocabulario del idioma [Fryer y Carpenter 06].

También se han propuesto los sistemas de diálogo hablado para mejorar las habilidades relacionadas con la Fonética y la Lingüística. El sistema Vocaliza [Vaquero et al. 06] se utiliza para la terapia del habla asistida por ordenador en el idioma español, de forma que sirva de ayuda en el trabajo diario de los terapeutas del habla que enseñan habilidades lingüísticas relacionadas con diferentes patologías del lenguaje. Del mismo modo, el sistema Litem (Literacy Innovation that Speech Technology Enables) es un tutor de lectura automatizada que muestra historias en una pantalla y escucha a los niños mientras las leen en voz alta [Mostow 12].

Por último, el diálogo también se utiliza como un indicador para la reflexión. Grigoriadou et al. (2003) describen un sistema en el que el alumno lee un texto sobre un acontecimiento histórico antes de afirmar su posición acerca de la temática del mismo y su justificación de esta opinión. Las respuestas de los estudiantes se clasifican como científicas, casi-científicas o no-científicas. Mediante esta clasificación el sistema intenta ofrecer un aprendizaje reflexivo adaptado a cada alumno.

CALMsystem [Kerly et al. 08c] promueve la reflexión de forma diferente. Los usuarios responden preguntas sobre el dominio del sistema y manifiestan el grado de confianza en su capacidad para responderlas correctamente. El sistema deduce un nivel de conocimiento del estudiante considerando sus respuestas. A continuación, alienta al alumno a participar en un diálogo para reflexionar sobre su autoevaluación y las diferencias entre su estimación del conocimiento y la realizada por el sistema. Los estudios realizados demostraron que este sistema mejoraba la precisión de las actividades de autoevaluación significativamente [Kerly et al. 08c]. La mejora del compromiso y motivación de los estudiantes también se destaca frecuentemente [Baylor y Kim 05], [Heffernan 03], [Fryer y Carpenter 06]. En algunos casos la motivación puede ser reforzada activamente a través de técnicas de motivación aplicadas a la tutorización.

<sup>2</sup> <http://www.m-learning.org/>

<sup>3</sup> [www.alicebot.org](http://www.alicebot.org)

<sup>4</sup> [www.jabberwacky.com](http://www.jabberwacky.com)

Una característica clave de la aplicación de los sistemas de diálogo en el e-learning es el uso de un método de comunicación natural, con el que los estudiantes están familiarizados a la hora de expresarse [Beun et al. 03]. La creciente alfabetización digital y la familiaridad con las tecnologías con las que se desarrollan estos sistemas, permite además incrementar sus aplicaciones y el número de posibles entornos de interacción.

En esta sección se ha mostrado la gran variedad de aplicaciones de los sistemas conversacionales en el ámbito del e-learning. Los sistemas descritos muestran la incorporación de características avanzadas como, por ejemplo, los interfaces basados en agentes personificados, con reconocimiento de gestos e incluso del estado emocional del usuario, así como el reconocimiento del habla y la síntesis de voz. Tal y como se ha descrito, estos sistemas se han integrado aplicaciones educativas de muy distinto tipo y con objetivos muy diversos.

#### 4. Ejemplos de sistemas de diálogo educativos desarrollados

En esta sección se describen dos interfaces multimodales interactivos que hemos desarrollado y que engloban tecnologías y aplicaciones descritas en la sección anterior.

##### 4.1 El sistema de diálogo multimodal *VoiceApp*

El sistema Voice Applications (*VoiceApp*) [Griol et al. 11] se ha desarrollado con el principal objetivo de servir como un marco de referencia para el estudio de la tecnología XHTML+Voice (X+V)<sup>5</sup> en la creación de sistemas de diálogo multimodal que facilitan la accesibilidad a la información en la red. Este lenguaje combina las funcionalidades ofrecidas por XHTML para la interacción utilizando modalidades visuales y las ofrecidas por el lenguaje VoiceXML<sup>6</sup> para la interacción mediante el habla.

El sistema integra diferentes aplicaciones, centrándonos en esta comunicación en la descripción de dos de ellas. *Voice Dictionary* recibe del usuario el contenido que se desea buscar en la enciclopedia Wikipedia, recoge el resultado de la búsqueda, lo procesa y lo muestra/narra al usuario, permitiéndole

también realizar una nueva búsqueda o seleccionar mediante la voz cualquiera de los enlaces que aparecen en el resultado. *Voice Search Engine* realiza búsquedas en el navegador web, remitiendo la información al buscador, procesando la información resultante y mostrarla al usuario, facilitando del mismo modo el acceso multimodal a los enlaces obtenidos como resultado de la búsqueda.

El sistema está implementado mediante un conjunto de documentos X+V. Algunos de ellos se encuentran almacenados desde el inicio en el servidor de documentos de la aplicación, mientras que otros se generan dinámicamente utilizando programación en PHP y tienen en cuenta las características y preferencias de los usuarios, así como información extraída de diversas bases de datos MySQL.

Para visualizar los documentos X+V que componen la aplicación, los usuarios deben disponer en su PC o dispositivos móviles de un navegador que soporte la interacción oral y la especificación X+V. En nuestra implementación hemos utilizado el navegador Opera<sup>7</sup>, que permite además la navegación mediante la voz (como recargar una página, volver a la página anterior, etc.).

A través del acceso a la página principal del entorno, mostrada en la Figura 2, encontramos un portal amigable que alberga las aplicaciones desarrolladas. El contenido de la página principal (a la cual los usuarios pueden acceder desde cualquier página del entorno pronunciando la palabra “Home”) consta de una sencilla presentación de cada una de los módulos del sistema. Cada módulo se muestra con su imagen representativa a la izquierda, y alineado a la derecha un texto que describe brevemente el contenido de la aplicación.

La totalidad de contenidos de la aplicación son leídos al usuario a través de la interfaz oral de la página, incluyendo la posibilidad de interrumpir este diálogo en cualquier momento, así como la utilización del resto de posibilidades de navegación multimodal ofrecidas por el entorno.

El desarrollo de interfaces orales implementados en X+V implica la definición de gramáticas, que establecen los límites de recepción de información del motor oral de la aplicación. Para poder cubrir el mayor rango de posibilidades de búsqueda en los

<sup>5</sup> [www.w3.org/TR/xhtml+voice/](http://www.w3.org/TR/xhtml+voice/)

<sup>6</sup> [www.w3.org/TR/voicexml20/](http://www.w3.org/TR/voicexml20/)

<sup>7</sup> [www.opera.com](http://www.opera.com)

diferentes módulos de la aplicación se estableció una estrategia que favoreciese esta maximización. Esta estrategia se basa en aspectos como la generación automática de gramáticas una vez se dispone de los resultados generados por la aplicación, la utilización de gramáticas con frases completas para favorecer la naturalidad de la interacción con el sistema, y la utilización incluso el deletreado de las palabras en los casos en los que requiere no acotar la búsqueda o en las situaciones en las que se haya detectado fallos continuados de reconocimiento.

o acceder a cualquiera de los enlaces destacados en el texto. Esta funcionalidad se consigue mediante la generación dinámica de las gramáticas correspondientes, en las que se incorporan directamente los vínculos encontrados en el contenido resultante de la búsqueda.

La aplicación *Voice Search Engine* se ha desarrollado con el objetivo fundamental de posibilitar la búsqueda de información a través del buscador Google utilizando la voz, tal y como muestra la Figura 3. La interfaz de la aplicación recibe el contenido facilitado por el usuario y muestra los resultados de la búsqueda visual y oralmente. Además, la aplicación permite seleccionar cualquiera de los enlaces mostrados en la página de resultados (Figura 3).



Figura 2. Pantalla inicial del sistema *VoiceApp*

Mediante la aplicación *Voice Dictionary* accedemos a un entorno sencillo mediante el cual pueden realizarse búsquedas en la enciclopedia Wikipedia, con la particularidad de que el contenido resultante de la búsqueda es narrado íntegramente al usuario.

Una vez que el contenido de la búsqueda inicial se muestra por pantalla y la interfaz oral comienza a narrarlo, el usuario puede perfectamente visitar cualquier otra aplicación interrumpiendo la narración,



Figura 3. Aplicación *Voice Search Engine* para la búsqueda de información

## 4.2 El sistema *Test Your English*

El sistema de diálogo multimodal *Test Your English* [Baena 13] tiene como principal objetivo posibilitar un aprendizaje autónomo y personalizado de idiomas. La aplicación web está desarrollada combinando los lenguajes XHTML+Voice y PHP. Se encuentra hospedada en un servidor web Apache, que contiene además las bases de datos MySQL, así como el conjunto de imágenes, textos, ficheros de audio, gramáticas y estilos.

La Figura 4 muestra la pantalla inicial del entorno web desarrollado. Tal y como puede observarse, los usuarios deben registrarse previamente en la

aplicación, de tal forma que el sistema pueda tener en cuenta las interacciones anteriores de cada usuario y ofrecerle un aprendizaje personalizado.



Figura 4. Página principal del sistema *Test Your English*

Las funciones de la aplicación desarrollada pueden detallarse en función de los siguientes módulos: Módulo de Práctica, Módulo de Evaluación de Resultados y Módulo de Gestión de Contenidos.

El Módulo de Práctica está enfocado al usuario final, permitiendo el acceso a tres tipos de ejercicios: ejercicios de gramática, ejercicios de vocabulario y ejercicios de escucha. Los estudiantes acceden a este módulo a través de un formulario simple en el que se selecciona la categoría (tiempo verbal en los ejercicios de gramática, temática en los ejercicios de vocabulario y título del texto en los ejercicios de escucha) y el nivel de los ejercicios.

Los ejercicios de gramática (Figura 5) consisten en completar frases con el correspondiente verbo, adjetivo, adverbio, artículo o cualquier otro elemento gramatical perteneciente al tema que inicialmente haya sido seleccionado, junto con el nivel de dificultad, por el usuario.

Los ejercicios de vocabulario (Figura 6) permiten al usuario dictar o escribir las palabras que definen un conjunto de imágenes que irán apareciendo de manera sucesiva relacionadas con el tema seleccionado.

En los ejercicios de escucha (Figura 7), el sistema de diálogo lee un texto seleccionado por el usuario. A continuación le son formuladas una serie de preguntas tipo test relacionadas con lo que ha escuchado, cuyas respuestas permiten analizar la capacidad de comprensión del texto.

El Módulo de Evaluación de Resultados (Figura 8) permite repasar las respuestas dadas por el usuario a cualquiera de los ejercicios realizados previamente,

conjuntamente con la corrección correspondiente en caso de haber fallado. El sistema realiza además una evaluación de resultados ofreciendo porcentajes, puntuaciones y comparativas al usuario con el fin de guiarle en el proceso de aprendizaje.



Figura 5. Ejemplo de ejercicios de gramática

La personalización de la herramienta se lleva a cabo a través de la opción "*choose for me*", que tiene en cuenta el número de ejercicios de cada categoría y nivel resueltos correctamente por cada usuario, de tal forma que el sistema pueda incidir en la corrección de los errores detectados en interacciones previas [Baena 13].



Figura 6. Ejemplo de ejercicios de vocabulario

Esta opción es un complemento de uso opcional que se presenta en la página de selección de categoría y nivel, dentro del módulo de realización de ejercicios, para cualquiera de las familias (gramática, vocabulario y escucha). El objetivo principal es que el propio sistema decida automáticamente la categoría y nivel de los ejercicios estudiando una serie de criterios que consideran la evolución específica de cada usuario.

Los criterios de selección son similares para los ejercicios de gramática y vocabulario, sin embargo presentan ligeras variaciones en el caso de la selección de ejercicios de escucha, por ese motivo han sido separados en el siguiente apartado donde se detallan.

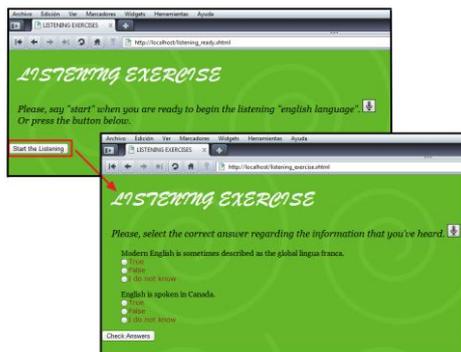


Figura 7. Ejemplo de ejercicios de escucha

El conjunto de criterios de selección incluye: i) El sistema buscará ejercicios que el usuario no haya hecho todavía y que pertenezcan al nivel que él mismo estimó e introdujo en el momento de hacer el registro de usuario. Para ello, se consulta la base de datos de la aplicación para conocer los ejercicios que el usuario ha completado ya de esa familia y nivel; ii) Se analiza automáticamente si los ejercicios que el usuario ha realizado de su nivel presentan un número de errores mayor que un umbral previamente especificado (por defecto, la mitad de los mismos). El objetivo es que el usuario repase ejercicios de su nivel, ya que ha cometido un número considerable de errores.

**MY RESULTS**

Hi, ishman  
 Here is the summary of results of all your activity in the application.  
 Please select one of the families to view: all the exercises done

Family	Level	Done/Total	Correct/Done	Link to Exercises
GRAMMAR EXERCISES	easy	0/0 (100%)	0/0 (100%)	Go to Exercises
	medium	0/0 (0%)	0/0 (0%)	Go to Exercises
	hard	0/0 (0%)	0/0 (0%)	Go to Exercises
VOCABULARY EXERCISES	easy	0/0 (100%)	1/0 (100%)	Go to Exercises
	medium	1/1 (100%)	1/1 (100%)	Go to Exercises
	hard	0/0 (0%)	0/0 (0%)	Go to Exercises
LISTENING EXERCISES	easy	0/0 (100%)	0/0 (0%)	Go to Exercises
	medium	0/0 (0%)	0/0 (0%)	Go to Exercises
	hard	0/0 (0%)	0/0 (0%)	Go to Exercises

PREVIOUS CONTENTS DONE: 87% (2013-11-12 18:50:26)  
 ACTUAL CONTENTS DONE: 88%  
 PREVIOUS SCORE: 89 (2013-11-12 18:50:26)  
 ACTUAL SCORE: 90

Figura 8. Interfaz gráfica del módulo de evaluación

La Figura 9 muestra un conjunto de casos de uso en los que puede observarse el funcionamiento de la opción de personalización de la herramienta. En el primer caso, el sistema recomienda al usuario la realización de ejercicios del nivel previamente seleccionado debido a que todavía no ha completado al menos la mitad de los mismos. En el segundo caso, el sistema realiza la misma recomendación debido a que se ha cometido un número de errores considerable. En el tercer caso, la recomendación es pasar al nivel inmediatamente superior dado que los resultados obtenidos en el nivel actual son satisfactorios. En el cuarto caso, la recomendación es seleccionar otra familia de ejercicios, ya que se han completado satisfactoriamente los requeridos para el nivel máximo de dificultad.

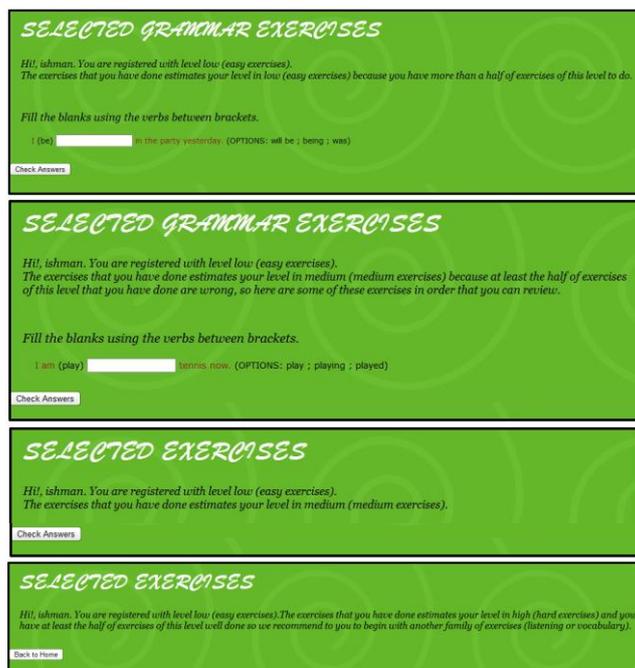


Figura 9. Casos de uso mostrando el funcionamiento de la opción *choose for me*

Finalmente, el Módulo de Gestión de Contenidos permite la administración de los contenidos prácticos de la aplicación y, por tanto, queda reservado a usuarios restringidos (profesorado, administradores de la aplicación, etc.). Este módulo hace uso de la herramienta phpMyAdmin<sup>8</sup> para la creación de nuevos ejercicios y la edición o borrado de ejercicios existentes de una forma sencilla e intuitiva.

<sup>8</sup> [www.phpmyadmin.net](http://www.phpmyadmin.net)

## 5. Conclusiones y trabajo futuro

En este artículo hemos explorado la variedad de aplicaciones de los sistemas de diálogo multimodales dentro del campo de la educación, identificando un gran número de finalidades y estrategias que incluyen la tutorización, el aprendizaje de idiomas, los compañeros de aprendizaje, los agentes pedagógicos, los sistemas de pregunta-respuesta, los sistemas que fomentan la reflexión del alumno, etc. Los beneficios reportados para los estudiantes incluyen las mejoras en las calificaciones, la motivación, el compromiso y las habilidades metacognitivas. El personal docente también puede beneficiarse de la capacidad de los sistemas conversacionales para realizar tareas de evaluación, reforzar los contenidos o desarrollar actividades de muy diferentes tipos.

También hemos explorado una variedad de temas relacionados con el desarrollo de sistemas de diálogo multimodal, incluyendo aspectos relacionados con el diseño y aplicaciones de estos sistemas. La principal conclusión es que los sistemas de diálogo desempeñan un papel muy importante en el futuro de los sistemas semipresenciales de e-learning, con lo que esperamos que la integración en estos sistemas sea cada vez más común.

Por último, también se han descrito dos sistemas desarrollados utilizando algunas de las aplicaciones innovadoras de los sistemas de diálogo multimodales descritas en el artículo. Las líneas de trabajo futuro incluyen la mejora de las capacidades de adaptación de dichos sistemas y la evaluación detallada de los diferentes módulos de los mismos, tanto de forma objetiva como mediante cuestionarios en los que se recoja la opinión subjetiva de los usuarios.

## Agradecimientos

Artículo desarrollado en el ámbito del proyecto “Guía y aplicación de recursos 2.0 para el desarrollo de entornos personales de aprendizaje” de la 10ª Convocatoria de Apoyo a Experiencias de Innovación Docente en estudios de Grado de la Universidad Carlos III de Madrid. Trabajo parcialmente financiado por los proyectos MINECO TEC2012-37832-C02-01, CICYT TEC2011-28626-C02-02, CAM CONTEXTS (S2009/TIC-1485) y TRA2010-20225-C03-01.

## Referencias

- [Ábalos et al. 10] Ábalos, N., Espejo, G., López-Cózar, R., Callejas, Z., Griol, D. (2010). A Multimodal Dialog System for an Ambient Intelligent Application in Home Environments. LNCS, 6231, 491-498.
- [Ai et al. 06] Ai, H., Litman, D., Forbes-Riley, K., Rotaru, M., Tetreault, J., Purandare, A. (2006). Using systems and user performance features to improve emotion detection in spoken tutoring dialogs. Proc. of Interspeech'06-ICSLP, pp. 797-800.
- [Aiemur et al. 97] Aiemur, E., Dufort, H., Leibu, D., Frasson, C. (1997). Some Justifications for the Learning by Disturbing Strategy. Proc. of Int. Conference on Artificial Intelligence in Education, pp. 119-126.
- [Aleven et al. 04] Aleven, V., Ogan, A., Popescu, O., Torrey, C., Koedinger, K. (2004). Evaluating the Effectiveness of a Tutorial Dialog System for Self-Explanation. Proc. of Int. Conference on Intelligent Tutoring Systems, pp. 443-454.
- [Andeani et al. 06] Andeani, G., Fabrizio, D. D., Gilbert, M., Gillick, D., Hakkani-Tur, D., Lemon, O. (2006). Let's DISCOH: Collecting an Annotated Open Corpus with Dialogue Acts and Reward Signals for Natural Language Helpdesks. Proc. of IEEE SLT'06, pp. 218-221.
- [Angehrn 01] Angehrn A. (2001). C-VIBE: A Virtual Interactive Business Environment addressing Change Management Learning. Proc. of ICALT'01, pp. 174-177
- [Angehrn et al. 06] Angehrn A., Nabeth T., Razmerita L., Roda C. (2001). K-InCA: Using Artificial Agents for Helping People to Learn New Behaviours. Proc. of ICALT'01, pp. 225-226.
- [Augusto 09] Augusto, J. (2009). Ambient Intelligence: Opportunities and Consequences of its Use in Smart Classrooms. *Italics* 8(2), 53-63.
- [Baena 13] Baena, I. (2013). *Test Your English: Desarrollo de un sistema multimodal mediante XHTML+Voice para un aprendizaje autónomo y personalizado de idiomas*. Trabajo Fin de Grado, Departamento de Informática, Universidad Carlos III de Madrid.

- [Bailly et al. 10] Bailly, G., Raidt, S., Elisei, F. (2010). Gaze, dialog systems and face-to-face communication. *Speech Communication*, 52(6), 598-612.
- [Baylor y Kim 05] Baylor, A.L. Kim, Y. (2005). Simulating Instructional Roles through Pedagogical Agents. *International Journal of Artificial Intelligence in Education*, 15(2), 95-115.
- [Beun et al. 10] Beun, R.-J., de Vos, E., Witteman, C. (2003). Embodied Dialog systems: Effects on Memory Performance and Anthropomorphisation. *Proc. of Int. Conference on Intelligent Virtual Agents*, pp. 315-319.
- [Bickmore 03] Bickmore, T. (2003). Relational Agents: Effecting Change through Human-Computer Relationships. PhD Thesis, Media Arts & Sciences, Massachusetts Institute of Technology.
- [Bickmore y Picard 05] Bickmore, T. W., Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer Human Interaction*, 12 (2), 293-327.
- [Bickmore et al. 13] Bickmore, T. W., Schulman, D., Sidner, C. (2013). Automated interventions for multiple health behaviors using conversational agents. *Patient Education and Counseling*, In Press.
- [Bohus y Rudnicky 05] Bohus, D., Rudnicky, A. (2005). LARRI: A Language-Based Maintenance and Repair Assistant. *Spoken Multimodal Human-Computer Dialogue in Mobile Environments*, 28, 203-218.
- [Bos et al. 99] Bos, J., Klein, E., Lemon, O., Oka, T. (1999). The verbmobil prototype system - A software engineering perspective. *Journal of Natural Language Engineering*, 5(1), 95-112.
- [Callejas y López-Cózar 10] Callejas, Z., López-Cózar, R., (2008). Influence of contextual information in emotion annotation for spoken dialogue systems. *Speech Communication*, 50 (5), 416-433.
- [Callejas et al. 07] Callejas, Z., Nouza, J., Cerva, P., López-Cózar, R., (2007). MyVoice goes Spanish. *Cross-lingual Adaptation of a Voice Controlled PC Tool for Handicapped People. Procesamiento del Lenguaje Natural*, 39, pp. 277-278.
- [Cassell et al. 01] Cassell, J., Sullivan, J., Prevost, S., Churchill, E.F. (2001). *Embodied Dialog systems*. The MIT Press.
- [Catizone et al. 03] Catizone, R., Setzer, A., Wilks, Y., (2003). Multimodal Dialog Management in the COMIC Project. *Proc. of EACL'03 Workshop on Dialog Systems: interaction, adaptation, and styles of management*, pp. 25-34.
- [Cavazza et al. 10] Cavazza, M., de la Camara, R.S., Turunen, M. (2010). How Was Your Day? a Companion ECA. *Proc. of AAMAS'10 Conference*, pp. 1629-1630.
- [Chan y Baskin 88] Chan, T.-W., Baskin, A.B. (1988). Studying with the Prince: The Computer as a Learning Companion. *Proc. of Int. Conference on Intelligent Tutoring Systems*, pp. 194-200.
- [Chou et al. 03] Chou, C.-Y., Chan, T.-W., Lin, C.-J. (2003). Redefining the Learning Companion: the Past, Present and Future of Educational Agents. *Computers & Education*, 40, 255-269.
- [Chu et al. 05] Chu, S.-W., O'Neill, I., Hanna, P., McTear, M. (2005). An approach to multistrategy dialog management. *Proc. of Interspeech'05-Eurospeech*, pp. 865-868.
- [Corradini et al. 05] Corradini, A., Mehta, M., Bernsen, N. O., Charfuelán, M. (2005). Animating an interactive conversational character for an educational game system. *Proc. of Int. Conference on Intelligent User Interfaces*, pp. 183-190.
- [D'Mello et al. 05] D'Mello, S. K., Craig, S. D., Gholson, B., Frankin, S., Picard, R., Graesser, A. C. (2005). Integrating Affect Sensors in an Intelligent Tutoring System. *Proc. of Workshop on Affective Interactions: The Computer in the Affective Loop at IUI*, pp. 7-13.
- [D'Mello et al. 12] D'Mello, S. K., Olney, A., Williams, C., Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies*, 70(5), 377-398.
- [De Carolis et al. 03] De Carolis, B., Pelachaud, C., Poggi, I., Steedman, M. (2003). *APML, a Markup*

- Language for Believable Behavior Generation. H. Prendinger and M. Ishizuka (Eds), *Life-like Characters. Tools, Affective Functions and Applications*.
- [de Rosis et al. 05] de Rosis, F., Cavalluzzi, A., Mazzotta, I., and Novielli, N. (2005). Can embodied dialog systems induce empathy in users? *Proc. of AISB'05 Virtual Social Characters Symposium*, pp. 1-8.
- [Dehn y van Mulken 00] Dehn, D. M., van Mulken, S. (2000). The impact of animated interface agents: a review of empirical research. *International Journal of Human Computer Studies*, 52 (1), 1-22.
- [Dillenbourg y Self 92] Dillenbourg, P., Self, J. (1992). People Power: A Human-Computer Collaborative Learning System in Intelligent Tutoring Systems, pp. 651-660.
- [Edlund et al. 08] Edlund, J., Gustafson, J., Heldner, M., Hjalmarsson A. (2008). Towards human-like spoken dialog systems. *Speech Communication*, 50 (8-9), 630-645.
- [Felder y Silverman 88] Felder, R.M., Silverman, L.K. (1988). Learning and Teaching Styles in Engineering Education, *Engr. Education*, 78(7), 674-681.
- [Feng et al. 06] Feng, D., Shaw, E., Kim, J., Hovy, E. (2006) An Intelligent Discussion-Bot for Answering Student Queries in Threaded Discussions. *Proc. of Int. Conference on Intelligent User Interfaces*, pp. 171-177.
- [Forbus et al. 99] Forbus, K. D., Whalley, P. B., Evrett, J. O., Ureel, L., Brokowski, M., Baher, J., Kuehne, S. E. (1999). CyclePad: An articulate virtual laboratory for engineering thermodynamics. *Artificial Intelligence*, 114(1-2), 297-347.
- [Fryer y Carpenter 06] Fryer, L., Carpenter, R. (2006). Bots as Language Learning Tools. *Language Learning and Technology*. *Language Learning and Technology*, 10(3), 8-14.
- [Graesser et al. 05] Graesser, A. C., Chipman, P., Haynes, B. C., Olney, A. (2005). AutoTutor: An Intelligent Tutoring System with Mixed-initiative Dialog. *IEEE Trans. in Education*, 48, 612-618.
- [Graesser et al. 01] Graesser, A.C., Person, N.K., Harter, D. (2001). Teaching Tactics and Dialog in AutoTutor. *International Journal of Artificial Intelligence in Education*, 12, 23-39.
- [Graesser et al. 99] Graesser, A.C., Wiemer-Hastings, K., Wiemer-Hastings, P., Kreuz, R., and the TRG. (1999) AutoTutor: A Simulation of a Human Tutor. *Journal of Cognitive Systems Research*, 1, 35-51.
- [Gratch et al. 02] Gratch, J., Rickel, J., Andre, J., Badler, N., Cassell, J., Petajan, E. (2002). Creating Interactive Virtual Humans: Some Assembly Required. *IEEE Intelligent Systems*, 17(4), 54-63.
- [Grigoriadou et al. 03] Grigoriadou, M., Tsaganou, G., Cavoura, T. (2003). Dialog-Based Reflective System for Historical Text Comprehension. *Proc. of Workshop on Learner Modelling for Reflection at Artificial Intelligence in Education*.
- [Griol y García-Jiménez 12] Griol, D., García-Jiménez, M. (2012) Development of Interactive Virtual Voice Portals to Provide Municipal Information. *Advances in Soft Computing*, 151, DCAI 2012, pp. 161-172.
- [Griol et al. 11] Griol, D., Molina, J.M., Corrales, V. (2011) The *VoiceApp* System: Speech Technologies to Access the Semantic Web. *LNCS*, 7023, CAEPIA 2011, pp. 393-402.
- [Griol et al. 08] Griol, D., Hurtado, L. F., Segarra, E., Sanchis, E. (2008). A statistical approach to spoken dialog systems design and evaluation. *Speech Communication*, 50 (8-9), 666-682.
- [Gustafson et al. 98] Gustafson, J., Elmberg, P., Carlson, R., Jönsson, A. (1998). An educational dialogue system with a user controllable dialogue manager. *Proc. of ICSLP'98*, pp. 33-37.
- [Heffernan 03] Heffernan, N.T. (2003) Web-Based Evaluations Showing both Cognitive and Motivational Benefits of the Ms. Lindquist Tutor, *Artificial Intelligence in Education*, pp. 115-122.
- [Hofmann et al. 12] Hofmann, N.T., Silberstein, A., Ehrlich, U., Berton, A., Mahr, A. (2012) Development of Speech-based In-car HMI Concepts for Information Exchange Internet Apps. *Proc. of International Workshop on Spoken Dialogue Systems*.

- [Hubal et al. 00] Hubal, R.C., Frank, G.A., Guinn, C.I. (2000). AVATALK Virtual Humans for Training with Computer Generated Forces. Proc. of 9th Conference on Computer Generated Forces and Behavioral Representation, pp. 617–623.
- [Jia 02] Jia, J. (2002) The Study of the Application of a Keywords-Based Chatbot System on the Teaching of Foreign Languages, University of Augsburg.
- [Johnson et al. 04] Johnson, W. L., LaBore, L., Chiu, Y.C. (2004). A Pedagogical Agent for Psychosocial Intervention on a Handheld Computer. Proc. of AAAI Fall Symposium on Dialogue Systems for Health Communication, pp. 22-24.
- [Jokinen et al. 04] Jokinen, K., Kanto, K., Rissanen, J. (2004). Adaptative User Modelling in AthosMail. LNCS, 3196, 149–158.
- [Kerly et al. 08a] Kerly, A., Ellis, R., Bull, S. (2008). Dialog systems in E-Learning. Proc. of AI-08, pp. 169-182.
- [Kerly et al. 08b] Kerly, A., Ellis, R., Bull, S. (2008). Children's Interactions with Inspectable and Negotiated Learner Models. Proc. of Int. Conference on Intelligent Tutoring Systems, pp. 132-141.
- [Kerly et al. 08c] Kerly, A., Ellis, R., Bull, S. (2008). CALMsystem: A Dialog system for Learner Modelling. Knowledge Based Systems, 21(3), 238-246.
- [Kim 07] Kim, Y. (2007) Desirable Characteristics of Learning Companions. International Journal of Artificial Intelligence in Education, 17(4), 371-388.
- [Kopp et al. 12] Kopp, K.J., Britta, M.A., Millisa, K., Graesser, A.C. (2012). Improving the efficiency of dialogue in tutoring. Learning and Instruction, 22(5), 320-330.
- [Kumar y Rose 11] Kumar R., Rose, C.P. (2011). Architecture for Building Dialog systems that Support Collaborative Learning. IEEE Trans. Learn. Technol., 4(1), 21-34.
- [Latham et al. 12] Latham, A., Crockett, K.A., McLean, D., Edmonds, B.A. (2012). A conversational intelligent tutoring system to automatically predict learning styles. Computers & Education, 59(1), 95-109.
- [Lester et al. 99] Lester, J.C., Stone, B.A., Stelling, G.D. (1999). Lifelike Pedagogical Agents for Mixed-Initiative Problem Solving in Constructivist Learning Environments. User Modeling and User-Adapted Interaction, 9, 1-44.
- [Litman y Silliman 04] Litman, D.J., Silliman, S. (2004). ITSPOKE: An Intelligent Tutoring Spoken Dialog System. Proc. of HLT Conference: North American Chapter of the ACL, pp. 5-8.
- [Liu y Chee 04] Liu, Y., Chee, Y. S. (2004). Designing Interaction Models in a Multiparty 3D learning environment. Proc. of 12th Int. Conference ICCE'04, pp. 293–302.
- [López-Cózar y Araki 05] López-Cózar, R., Araki, M. (2005). Spoken, Multilingual and Multimodal Dialog Systems. Development and Assessment. John Wiley and Sons.
- [Lutfi et al. 13] Lutfi, S.L., Fernández, F., Lucas, J.M., López, L., Montero, J.M. (2013). A satisfaction-based model for affect recognition from conversational features in spoken dialog systems. Speech Communication, 55(7-8), 825-840.
- [Mairesse et al. 09] Mairesse, F., Gasic, M., Jurcicek, F., Keizer, S., Thomson, B., Yu, K., Young, S. (2009). Spoken language understanding from unaligned data using discriminative classification models. Proc. of ICASSP'09, pp. 4749-4752
- [Manzoni y Angehrn 97] Manzoni J. F., Angehrn A., (1997). Understanding Organizational Dynamics of IT-Enabled Change: A Multimedia Simulation Approach. Journal of Management Information Systems, 14(3), 109-140.
- [Marsella et al. 03] Marsella, S.C., Johnson, W.L., LaBore, C.M. (2003). Interactive pedagogical drama for health interventions. IU. Hoppe et al. (Eds.), Artificial Intelligence in Education: Shaping the Future of Learning through Intelligent Technologies, pp. 341-348.
- [Mattasoni 02] Mattasoni, M., Omologo, M., Santarelli, A., Svaizer, P., (2002). On the Joint Use of Noise Reduction and MLLR Adaptation for In-Car Hands-Free Speech Recognition. Proc. of ICASSP'02, pp. 289–292.

- [McTear 04] McTear, M. F. (2004). Spoken dialog technology. Springer.
- [Melin et al. 01] Melin, H., Sandell, A., Ihse, M. (2001). CTT-bank: A speech controlled telephone banking system - an initial evaluation. TMH-QPSR, KTH, 1:1-27
- [Meza et al. 08] Meza, I., Riedel, S., Lemon, O. (2008). Accurate Statistical Spoken Language Understanding from Limited Development Resources. Proc. of ICASSP'08, pp. 5021-5024.
- [Minker 98] Minker, W. (1998). Stochastic versus rule-based speech understanding for information retrieval. *Speech Communication*, 25 (4), 223–247.
- [Montoro et al. 06] Montoro, G., Haya, P. A., Alamán, X., López-Cózar, R., Callejas, Z. (2006). A proposal for an XML definition of a dynamic spoken interface for ambient intelligence. Proc. of ICIC'06, pp. 711–716.
- [Mostow 12] Mostow, J. (2012). Why and How Our Automated Reading Tutor Listens. Proc. of Int. Symposium on Automatic Detection of Errors in Pronunciation Training (ISADEPT), pp. 43-52.
- [Mühlpfordt y Wessner 05] Mühlpfordt, M., Wessner, M. (2005). Explicit Referencing in Chat Supports Collaborative Learning. Proc. of CSCL'05, pp. 662-671.
- [Nestorovič 12] Nestorovič, T. (2012). Creating a general collaborative dialogue agent with lounge strategy feature. *Expert Systems with Applications*, 39(2), 1607-1625.
- [Pareto et al. 12] Pareto, L., Haake, M., Lindstrom, P., Sjoden, B., Gulz, A. (2012). A teachable-agent-based game affording collaboration and competition: evaluating math comprehension and motivation. *Education Tech Research*, 60, 723-751.
- [Pieraccini 12] Pieraccini, R., (2012). *The Voice in the Machine: Building Computers That Understand Speech*. MIT Press.
- [Pon-Barry et al. 06] Pon-Barry, H., Schultz, K., Bratt, E.O., Clark, B., Peters, S. (2006). Responding to student uncertainty in spoken tutorial dialog systems. *International Journal of Artificial Intelligence in Education*, 16, 171-194
- [Rabiner et al. 96] Rabiner, L.R., Juang, B.H., Lee, C.H. (1996). An overview of automatic speech recognition. Kluwer Academic Publishers, *Automatic Speech and speaker Recognition: Advanced Topic*, pp. 1-30.
- [Reiter 95] Reiter, E. (1995). NLG vs. templates. Proc. of the Fifth European Workshop in Natural Language Generation, pp. 95–105.
- [Rickel y Johnson 99] Rickel, J., Johnson, W.L. (1999) Animated agents for procedural training in virtual reality: Perception, cognition, and motor control. *Applied Artificial Intelligence*, 13, 343-382.
- [Roda et al. 01a] Roda, C., Angehrn, A., Nabeth, T. (2001). Dialog systems for Advanced Learning: Applications and Research. Proc. of BotShow'01 Conference, pp. 1-7.
- [Roda et al. 01b] Roda C., Angehrn A., Nabeth T. (2001). Matching Competencies to Enhance Organizational Knowledge Sharing: An Intelligent Agents Approach. Proc. of 7th International Netties Conference, pp. 931-937.
- [Rosé et al. 01] Rosé, C.P., Moore, J.D., VanLehn, K., Allbritton, D. (2001). A Comparative Evaluation of Socratic versus Didactic Tutoring. Proc of Cognitive Sciences Society.
- [Segedy et al. 13] Segedy, J., Kinnebrew, J., Biswas, G. (2013). The effect of contextualized conversational feedback in a complex open-ended learning environment. *Education Tech Research*, 61, 71-89.
- [Shaw et al. 99] Shaw, E., Johnson, W.L., Ganeshan, R. (1999). Pedagogical Agents on the Web, Proc. of Int. Conference on Autonomous Agents, pp. 283-290.
- [Vaquero et al. 06] Vaquero, C., Saz, O., Lleida, E., Marcos, J., Canalís, C. (2006) VOCALIZA: An application for computer-aided speech therapy in Spanish language. Proc. of IV Jornadas en Tecnología del Habla, pp. 321–326.
- [Wahlster 06] Wahlster, W. (Ed.) (2006). *SmartKom: Foundations of Multimodal Dialog Systems*. Springer.
- [Wang et al. 07] Wang, Y., Wang, W., Huang, C. (2007). Enhanced Semantic Question Answering

- System for e-Learning Environment. Proc of AINAW'07 Conference, pp. 1023-1028.
- [Wang y Johnson 08] Wang, N., Johnson, L.W. (2008). The Politeness Effect in an intelligent foreign language tutoring system. Proc. of 9th Int. Conference on Intelligent Tutoring Systems, pp. 70-280.
- [Weng et al. 06] Weng, F., Varges, S., Raghunathan, B., Ratiu, F., Pon-Barry, H., Lathrop, B., Zhang, Q., Scheideck, T., Bratt, H., Xu, K., Purver, M., Mishra, R., Raya, M., Peters, S., Meng, Y., Cavedon, L., Shriberg, L., 2006. CHAT: A Conversational Helper for Automotive Tasks. Proc. of Interspeech-ICSLP'06, pp. 1061–1064.
- [Williams y Young 07] Williams, J., Young, S. (2007). Partially Observable Markov Decision Processes for Spoken Dialog Systems. *Computer Speech and Language*, 21(2), 393–422.
- [Zue y Glass 00] Zue, V.W., Glass, J.R. (2000). Conversational interfaces: Advances and challenges. Proc. of IEEE'88, pp. 1166–1180.