SD-TEAM: Tecnologías de aprendizaje interactivo, autoevaluación y multimodalidad en sistemas de diálogo hablado multidominio

SD-TEAM: Interactive Learning, Self-Evaluation and Multimodal Technologies for Multidomain Spoken Dialogue Systems

R. Justo y M. I. Torres
PR & ST Group.
Univ. del País Vasco
(EHU), España

E. Lleida
CTG
Univ. de Zaragoza
(UZ), España

E. Sanchis
PR&AI Group.
Univ. Politécnica de Valencia
(UPV), España

R. de Córdoba
Speech Technology Group.
Universidad Politécnica de Madrid
(UPM), España

J. Macías-Guarasa
Intelligent spaces and Transport Group.
Universidad de Alcalá de Henares
(UAH), España

1. Datos del Proyecto

Director: María Inés Torres Barañano (manes.torres@ehu.es). Grupo Pattern Recognition & Speech Technologies. Universidad del País Vasco. Sarriena s/n. 48940 Leioa, España.

Patrocinadores: Este proyecto está financiado por el Ministerio de Ciencia e Innovación del Gobierno de España, mediante el programa TIN2008-06856-C05-01.

URL: <http://www.sd-team.es>.

2. Contexto

Los avances tecnológicos han introducido un gran número de dispositivos con los que interaccionamos continuamente y que han modificado enormemente nuestros hábitos cotidianos. Dentro del ámbito en el que se enmarca este proyecto, las tecnologías del habla, es importante destacar el desarrollo y la investigación en sistemas de diálogo hablado, robustos, intuitivos y transparentes como modo de interacción entre el usuario humano y el servicio automático. En las últimas décadas, se ha logrado un aumento importante de las prestaciones de los sistemas de reconocimiento automático del habla (RAH), produciendo, a su vez, un aumento en la exigencia de funcionalidad de los sistemas interactivos basados en voz que deberán abordar tareas cada vez más complejas. Tal es el caso de los sistemas de diálogo hablado, que en

la actualidad están en una fase de desarrollo preliminar, aún no consolidado en forma de productos comerciales usables por la sociedad; o los sistemas de procesamiento automático de información oral como la traducción automática, la generación de resúmenes o la clasificación/extracción/recuperación de información.

Estos nuevos escenarios suponen retos tecnológicos que requieren avances científicos importantes. En este contexto se sitúa el proyecto SD-TEAM que supone una evolución natural del proyecto EDECÁN [TIN2005-08660-C04] que fue desarrollado previamente por el mismo equipo. El proyecto EDECÁN proponía incidir en los componentes de un sistema de diálogo clásico para mantener sus prestaciones frente a cambios en el contexto acústico (tanto físicos como debidos al usuario); minimizar el esfuerzo de rediseño frente a cambios en el contexto de aplicación y explotar la información del usuario para personalizar el sistema. El proyecto SD-TEAM pretende la evolución de estos sistemas de diálogo cerrados, hacia sistemas de interacción humana dinámicos y flexibles, con una clara capacidad de evolución autónoma mediante el aprendizaje, evaluación y cooperación a través de la experiencia de uso. Estos nuevos sistemas “vivos” serán una evolución de sus predecesores, capaces de adaptarse, interactuar con usuarios y con otros sis-

temas y de aprender “sobre la marcha” para reconfigurarse. En este contexto, una arquitectura distribuida y flexible (evolución a su vez de la propuesta en el proyecto EDECÁN) permitirá la cooperación entre diferentes sistemas (módulos de comprensión, reconocedores, gestores de diálogo, etc. desarrollados para diferentes entornos de funcionamiento) así como la interacción de naturaleza multimodal con usuarios, ambas basadas en un cierto funcionamiento autónomo que permitirá una evolución constante e independiente. Esta evolución debe realizarse teniendo como objetivos claros el aumento de las prestaciones de los sistemas automáticos y el aumento de la satisfacción del usuario cuando los utiliza e interactúa con ellos.

3. *Objetivos*

A lo largo de los años se han desarrollado sistemas de diálogo “estáticos”, donde el usuario se adapta hacia las necesidades del sistema. Por el contrario, cada día más, los sistemas tienden a ser “orgánicos”, capaces de aprender, crecer, reconfigurarse y auto repararse ofreciendo de esta forma una mayor robustez frente a cualquier cambio en las condiciones de trabajo. En este contexto se hace imprescindible estudiar en profundidad los aspectos metodológicos y tecnológicos que faciliten la extensión de sistemas de diálogo hablado estáticos hacia sistemas “orgánicos” con capacidad de aprendizaje, evaluación y cooperación, en definitiva, sistemas que evolucionen de forma autónoma hasta lograr la plena satisfacción del usuario. Entre los usuarios potenciales se deben incluir aquellos colectivos para los cuales un sistema que no tenga en cuenta una adecuada adaptación y personalización no funcionaría. Nos referimos, entre otros, a las personas con algún tipo de discapacidad que lleven aparejados problemas de comunicación (habla alterada, sordera,...) y de movilidad o usuarios de habla no nativa.

3.1. *Objetivos Científicos*

- Explorar mecanismos eficaces de **aprendizaje interactivo**.
- Explorar metodologías eficientes para la **detección del entorno**.
- Desarrollar aproximaciones que consideren la **interacción** del usuario como un

elemento o módulo más del sistema que **coopera** con los demás.

- Desarrollar igualmente aproximaciones que contemplen la **interacción entre sistemas** de diferente complejidad.
- Desarrollar **arquitecturas distribuidas**.
- Definir medidas de **autoevaluación**.
- Explorar metodologías eficientes para el **tratamiento conjunto de entradas multimodales**: voz, pantalla táctil, caracteres, video, etc.

3.2. *Objetivos Tecnológicos*

- Ampliar de forma importante las **capacidades tecnológicas del consorcio** (reconocedores para grandes vocabularios, dominios más amplios para los sistemas de diálogo, tratamiento de palabras no vistas, arquitecturas acopladas y desacopladas, integración de la interacción del usuario, etc.).
- Construir **prototipos demostradores**.

4. *Grupos Participantes*

Los grupos de investigación solicitantes tienen larga experiencia de trabajo conjunto en el desarrollo de sistemas de RAH, y en el dominio del desarrollo de sistemas de diálogo hablado. Las diferencias en el origen y especialización de cada uno de los grupos hacen especialmente atractivo el trabajo en conjunto. El grupo CTG de la UZ proviene del ámbito de teoría de la señal y las comunicaciones, aportando un conocimiento fundamental para el desarrollo de técnicas de robustez de sistemas de reconocimiento. El grupo de PR&AI de la UPV y el grupo PR&ST de la EHU poseen conocimiento esencial sobre el aprendizaje automático de los modelos, permitiendo al consorcio profundizar en el desarrollo de metodologías de aprendizaje automático a partir de muestras, para el diseño de modelos de lenguaje, comprensión y diálogo. El grupo de Tecnología del Habla de la UPM tiene una extensa experiencia en el diseño y evaluación de sistemas de diálogo persona-máquina basados en tecnología del habla. Finalmente el Grupo de ingeniería electrónica aplicada a espacios inteligentes y transporte de la UAH tiene una extensa experiencia en sistemas de posicionamiento de móviles en entornos inteligentes.