Los miembros de GTTS: Germán Bordel, Mireia Díez, Mikel Peñagarikano, Luis Javier Rodríguez, Amparo Varona y Silvia Nieto.

En busca de la voz autoescrita

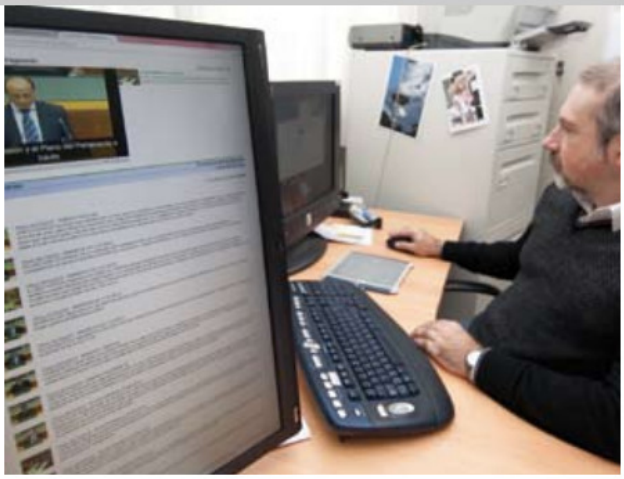
El grupo GTTS ha desarrollado un software con numerosas aplicaciones para el procesamiento automático de habla

El último poema del primer libro impreso en euskera se titula Sautrela. Ese es el nombre que el Grupo de Trabajo en Tecnologías Software, de la Facultad de Ciencia y Tecnología, eligió para el procesador que ha desarrollado para sus trabajos sobre procesamiento automático del habla.

Si existe un área dentro de las tecnologías de la información que se haya resistido desde el punto de vista de la industria, esa es la del reconocimiento del habla. La investigación comenzó en la segunda mitad de la década de los 60, pero el negocio no alcanzó los 1.000 millones de dólares hasta el año 2000. "En ningún otro campo se ha tardado tanto", precisa Germán Bordel, el director del Grupo de Trabajo en Tecnologías Software (GTTS), que investiga principalmente en transcripción de voz a texto, identificación de personas y reconocimiento de lenguas. Otros tres profesores, Amparo Varona, Luis Javier Rodríguez y Mikel Peñagarikano, la doctoranda Mireia Díez y la investigadora contratada Silvia Nieto completan el grupo. Todos ellos pertenecen al Departamento de Electricidad y Electrónica.

Quienes trabajan en reconocimiento del habla suelen utilizar algunos de los software existentes. Los más famosos son HTK de la Universidad de Cambridge, y ahora propiedad de Microsoft, y el SPHINX de la Carnegie Mellon University. GTTS utiliza Sautrela, una herramienta que ha creado desde cero y que continúa desarrollando a base de rutinas para procesar la señal, realizar modelización estadística… Sautrela ha probado su validez, por ejemplo, en las competiciones internacionales que organiza el National Institute of Standards and Technology (NIST).

GTTS se presenta desde hace cuatro años a las pruebas del NIST sobre reconocimiento de hablantes y de lenguas. Al principio participaba una docena de grupos, últimamente son unos treinta. Los resultados colocan a GTTS entre los más competitivos, sólo por detrás de algunos centros punteros entre los que destaca particularmente el Massachusetts Institute of Technology (MIT). "Andamos todos en porcentajes muy bajos de error, las diferencias entre unos y otros son pequeñas", indica Bordel, incluso cuando se trata de distinguir lenguas próximas como el castellano, catalán y gallego. Algo similar sucede con el



No sólo eso, al hablar dejamos frases inconclusas, equivocamos palabras, dudamos y reiniciamos, pero el discurso puede seguir sin problemas, porque quien escucha es capaz de completar lo que falta y corregir lo que falla. "Todo ese complejo sistema de entender lo hemos reducido a procesar una señal y convertirla en texto, obviando un montón de elementos, y se ha resuelto sólo en circunstancias muy determinadas", señala el profesor Mikel Peñagarikano. Por ejemplo, existen en el mercado aplicaciones que funcionan bien en entornos muy concretos como la medicina forense. Se graba al forense durante la autopsia y un reconocedor convierte sus frases en texto. Funciona porque el lenguaje que se utiliza es muy formal y protocolario, tampoco hay ruidos ni interferencias que afecten a la calidad del discurso, es decir, no se da ninguna de las circunstancias que en cualquier otro ámbito afectan a la locución. "El problema principal es la falta de robustez frente a los ruidos, las distintas pronunciaciones, los acentos diferentes, que dificultan el reconocimiento automático", explica Mikel Peñagarikano.

reconocimiento de locutores. ¿Superan, por tanto, las máquinas a las personas en estos campos? "No conozco ningún estudio al respecto, pero en mi opinión las máquinas llegan un poco más allá que nosotros", explica Bordel. "En lo que se trabaja ahora principalmente es en mejorar la robustez cuando la voz nos llega con diferentes condiciones acústicas: de locales cerrados, de exteriores, por vía telefónica…", añade el profesor Luis Javier Rodríguez.

La herramienta es muy competitiva en el reconocimiento de lengua y locutores

Falta de robustez

Las máquinas, por tanto, reconocen lenguas y personas de un modo muy competitivo, pero no son capaces de transcribir voz a texto al mismo nivel. "Aunque desde hace años se habla de ello y parece que está resuelto, la solución definitiva aún no existe", explica Amparo Varona. "La transcripción tiene su base en el reconocimiento de fonemas y ahí las máquinas creo que también son algo mejores que nosotros, pero luego hay que subir hacia arriba combinando los fonemas correctos, desechando los errores y, en definitiva, aplicando un conocimiento que los sistemas actuales son capaces de modelizar con bastantes limitaciones", añade Bordel.

Vídeos subtitulados

Aunque la tecnología para transcribir directamente aún no esté lista, GTTS ha dado forma a un proyecto pionero que ha puesto en marcha para el Parlamento Vasco y en el que ha volcado las técnicas de reconocimiento para detectar lenguas y personas y transcribir textos. Consiste en el alineado de la voz de los vídeos de las sesiones con los textos del diario de las mismas. La aplicación, que detecta automáticamente si el discurso se realiza en euskera o castellano, permite localizar cada palabra del diario en el vídeo y sincronizar la versión escrita y la oral. El resultado, el vídeo subtitulado, está listo un día después de que se les entregue la transcripción de cada sesión, normalmente al día siguiente de la jornada parlamentaria, y supone un paso más para favorecer la accesibilidad de las personas discapacitadas.

"El producto se puede mejorar, pero hay poco margen. La idea es que no haya errores, y a día de hoy los que se producen son prácticamente imperceptibles", explica Luis Javier Rodríguez. Aunque se está en periodo de pruebas, el Parlamento ya proporciona en su web los vídeos subtitulados, pero, en cuanto la herramienta se ajuste no requerirá supervisión, y el protocolo se acelerará. "El tiempo de ejecución será el mínimo y dependerá fundamentalmente de los transcriptores", añade. Sí que queda margen de mejora en el desarrollo de un buscador, que permitirá localizar palabras o locutores concretos en los archivos de vídeo. "Esta aplicación la hemos planteado como banco de pruebas de nuestras investigaciones, pero en breve el Parlamento dispondrá de una primera versión operativa", añade Bordel.