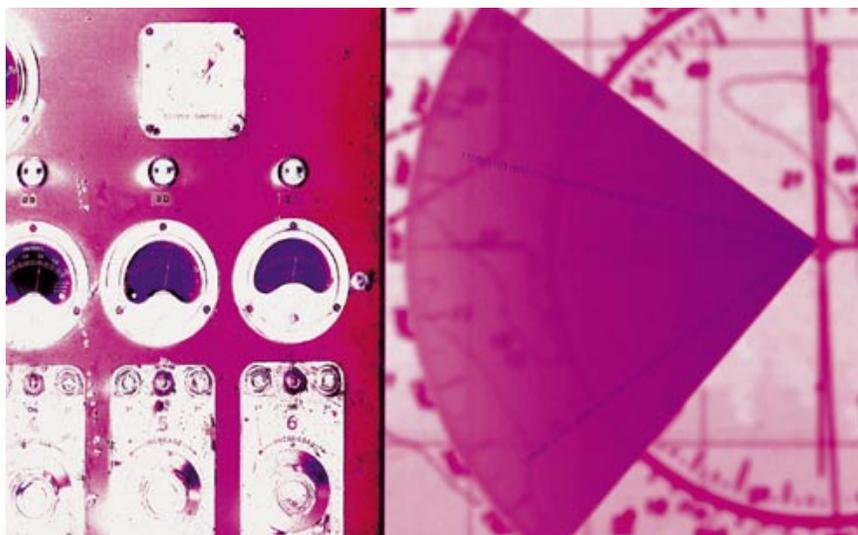


LOS SISTEMAS DE RECONOCIMIENTO Y SÍNTESIS DE VOZ, EN AUGE

# ¡Mira quién habla!

Esther Macías.-

**L**o lograr una interactividad real con las máquinas a través de la voz. Éste, además de ser el sueño de muchos profesionales y una necesidad manifiesta para el colectivo de discapacitados, es la actividad a la que se dedican cada vez más compañías. Si bien tecnologías como el dictado digital ya están muy conseguidas y proporcionan importantes ventajas como el ahorro de tiempo a sus usuarios, el reto final de los investigadores de estos ámbitos es lograr que los dispositivos electrónicos puedan mantener conversaciones en lenguaje natural.



No es ciencia ficción. Las tecnologías de tratamiento del lenguaje, reconocimiento y síntesis de voz son un hecho. Si bien su implantación aún es escasa y se trata de un mercado al que le queda mucho camino por recorrer, cada vez hay más empresas y profesionales con sus ojos puestos en estos programas que aportan importantes ventajas a sus usuarios, en términos de dinero y tiempo. En definitiva: estos sistemas pueden revolucionar los negocios. Claro que, según datos de la consultora tecnológica IDC, aunque estos sistemas interactivos avanzados con posibilidades de conversación están emergiendo comercialmente y existe ya una demanda de éstos en el mercado, aún estamos lejos de una interacción real de estos sistemas en los negocios. Rafael Achaerandio, analista de IDC España, asegura que estas tecnologías de reconocimiento de voz están ligadas a campos de aplicación como los entornos CTI (Computer Telephony Integration), que están asociados a los CRM (sistemas de gestión de la relación con clientes) y a los Call Centers, donde equipos como los IVR (Interactive Voice Response) están muy implantados. "Éste en concreto —señala— es un entorno donde esta tecnología está empezando a estar madura". El analista destaca también los desarrollos en el ámbito de la seguridad, donde el reconocimiento de voz se utiliza en aplicaciones de seguridad biométrica, así como en el sector de las telecomunicaciones, sobre todo en móviles, donde ya se están lanzando servicios basados en reconocimiento de voz. Asimismo, se están realizando progresos en el mundo de la ofimática: "Discapacitados y profesionales con necesidades de realizar dictado digital han utilizado este tipo de tecnologías en sus PCs y existen diversos fabricantes de software de reconocimiento de voz para ordenadores personales, como es el caso de IBM". Achaerandio vaticina que estos sistemas irán incrementando su relevancia a medida que progrese la implantación de VoIP, sobre todo en lo relacionado con las herramientas colaborativas (conferencias, mensajería instantánea, etc.), y en aplicaciones empresariales como los ERP (software de gestión empresarial) y los SCM (gestión de la cadena de suministro), etc.: "La voz se irá integrando e interrelacionando cada vez más con los datos, dando lugar a procesos de negocio optimizados".

## Productividad y ahorro de tiempo

Las ventajas que conlleva para los profesionales la adopción de los sistemas de reconocimiento vocal y dictado digital, que permiten

### HÍBRIDOS ELEGANTES: LOS SISTEMAS INTERACTIVOS DEL FUTURO

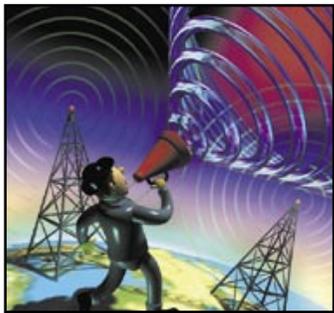
Según IDC, la próxima generación de sistemas interactivos serán "híbridos elegantes" que combinarán tecnologías avanzadas, herramientas metalingüísticas, estadísticas y de múltiples interpretaciones, reconocimiento de voz y diálogos y herramientas visuales, para entender e interactuar con datos y texto. La consultora asegura que actualmente nos encontramos en un primer paso hacia la interactividad real en los sistemas de conversación.

dictar textos y hacer que el ordenador obedezca a sus órdenes hablándole, son amplias en términos de incremento de productividad y comodidad en el ámbito laboral, la reducción de papeleo y costes, así como la prevención de lesiones físicas por el esfuerzo repetitivo que se realiza al trabajar con un periférico que acaba de cumplir 136 años: el teclado. Así se desprende de un informe elaborado por Speech Recognition Ware, una consultora con sede en Bruselas (Bélgica) creada por el jurista Jesús María Boccio y que acaba de lanzar al mercado Transcription Aid, una aplicación informática que perfecciona las prestaciones que ofrece el programa de reconocimiento vocal Dragon Naturally Speaking Profesional, propiedad de la multinacional americana ScanSoft, para adaptarlo a los entornos profesionales europeos. Como asegura Boccio, gracias a su tecnología española, cualquier usuario de este motor de reconocimiento vocal puede delegar la corrección de sus transcripciones, así como la tarea de mejorar su perfil de usuario para conseguir más precisión, en una/o secretaria/o. "Esta característica es esencial para que el reconocimiento vocal sea un éxito en Europa, como lo es en EE.UU., ya que la medicina y otras profesiones candidatas a utilizar nuestra tecnología tienen un carácter público en el Viejo Continente y dichos profesionales no están tan dispuestos como sus homólogos americanos a invertir tiempo propio y esfuerzos no remunerados en dichas tareas...". El jurista apunta que para que el éxito de esta tecnología sea total, las organizaciones interesadas tendrán que integrarla con su sistema informático y modificar sus procesos organizativos y de trabajo. Médicos, abogados, procuradores, traductores, escritores, agentes comerciales y discapacitados son los usuarios que más se pueden beneficiar de estas tecnologías. Boccio revela que su compañía acaba de obtener una importante subvención a fondo perdido de la Fundación la Caixa para desarrollar un proyecto de inserción laboral de discapacitados mediante el reconocimiento vocal, con la empresa pública SINPROMI del Cabildo de Tenerife. El Gobierno Vasco, el Ministerio de Asuntos Exteriores, el Ministerio de Defensa, la Organización Mundial de Turismo y la propia Unión Europea son algunos de los clientes de Speech Recognition Ware.

Para Boccio, el futuro de estas tecnologías es espectacular, pues acaba de llegar a su madurez: "El propio Bill Gates, presidente de Microsoft, dice que el famoso Interfaz Gráfico de Usuario (Graphic User Interface, GUI) de su sistema operativo Windows será pronto sustituido por un Interfaz de Usuario Sonoro (Sound User Interface, SUI)". De hecho, en el informe antes citado se asegura que el mercado de reconocimiento vocal representó 897 millones de dólares a escala mundial en 2003 y se espera que alcance 5.000 millones en 2008. La puesta en marcha en la Unión Europea a finales de 2003 de la Directiva 2000/78/CE, que garantiza la igualdad de oportunidades para todos los trabajadores y que conlleva proporcionar a aquellos con problemas de discapacidad los medios tecnológicos necesarios para poder desarrollar su labor, impulsará estas tecnologías.

## Hacia una interacción hombre-máquina absoluta

El objetivo último de estas tecnologías va más allá de un mero reconocimiento vocal. Así lo manifiestan fuentes de la compañía española Vida Software, cuyos ingenieros desde hace dos años están desarrollando una tecnología de "Interacción Natural", que, aseguran, sólo está siendo desarrollada en el mundo por otra compañía en EE.UU. y que brinda la posibilidad de interactuar con máquinas y dispositivos móviles igual que lo hacen las personas entre sí. Actualmente, Vida Software está trabajando en una segunda generación de productos centrados en el reconocimiento de rasgos biométricos, de emociones, etc., basándose en el lenguaje natural. "Se trata —aseguran sus responsables— de posibilitar a los programas el reconocimiento de voz, huellas dactilares e incluso estados de ánimo del usuario con el objetivo final de eliminar barreras de usabilidad a la hora de interactuar con móviles y máquinas, lo que significa que más personas puedan hacer uso de más servicios más sencillos". Es decir, consiste en conseguir aplicaciones más convincentes y funcionales, de modo que se logre reducir la frustración del usuario en su relación con las aplicaciones informáticas con las que necesita interactuar. Además, en el caso de la telefonía móvil y teniendo en cuenta que el mercado de voz está saturado, esta tecnología tiene mucho que aportar a las operadoras, cuyas futuras ganancias dependen del despliegue de la tercera generación UMTS (Universal Mobile



incluso a la de un operador humano". El problema del lenguaje natural hombre-máquina radica en la diferencia de jergas, acentos y formas de hablar de cada individuo, que son únicas, así como el alto coste de estos desarrollos. No obstante, los progresos son significativos, según González López, sobre todo en EE.UU. "También trabajamos en la síntesis de voz, es decir, que el sistema pueda hablar

expresando emociones, pues no tiene sentido, por ejemplo, que informe a su interlocutor sobre una bajada de las acciones con una voz alegre". Unos desarrollos en línea con los de la compañía española Natural Vox, que destina el 41% de su presupuesto a I+D. Ésta, dedicada a la implantación de sistemas de telefonía interactiva, principalmente en el ámbito de la banca, el *ticketing* o venta

de entradas y en el sector de la administración pública, centra sus trabajos en conseguir que cuando el usuario habla con un sistema sienta el *feedback* que sentiría con una persona.

González López, de IBM, señala que su grupo también investiga la aplicación de estas tecnologías al ámbito de la traducción, con fuerte demanda en la Unión Europea. "Desde hace un mes, IBM parti-

cipa con la Universidad Politécnica de Cataluña, universidades alemanas, el centro de investigación francés Limsi, Nokia, Siemens y Sony, en el desarrollo de un proyecto de traducción de "habla a habla", es decir, hablar a una máquina y que ésta traduzca lo que decimos. El proyecto, de tres años de duración, cubre las traducciones de inglés a chino, de español a inglés y de inglés a español.

Telecommunications System), con la que mejorarán sus servicios. De momento, la compañía española ha realizado tres pilotos de pruebas con Vodafone y actualmente está realizando otro con Amena. Telefónica Móviles también ha mostrado interés por el producto. Además, recientemente ha firmado un acuerdo con Symbian (compañía propiedad de Ericsson, Nokia, Panasonic, Psion, Samsung Electronics, Siemens y Sony Ericsson) y su tecnología de interacción natural ya está disponible para los teléfonos basados en este sistema operativo.

La relevancia de los servicios móviles han provocado que también la compañía de Boccio haya adaptado su tecnología de reconocimiento vocal y dictado digital a cualquier dispositivo, ya que ésta soporta la transcripción automática de cualquier grabación de voz generada con otra aplicación informática, una grabadora digital, una agenda personal digital e incluso los teléfonos móviles con sistema operativo Windows Mobile Smartphone. "Ya es posible dictar un documento desde el móvil, transmitirlo por la actual generación GPRS (dentro de poco, por UMTS) a la oficina y encontrarlo transcrito al volver a la misma", señala.

### Lenguaje natural

La comprensión del lenguaje natural es también el área en la que trabaja el grupo de tecnología de habla de IBM, creado hace treinta años y cuyo trabajo en España se desarrolla desde mediados de los noventa en Sevilla. El director de este centro, Jorge González López, explica que éste se coordina con otros grupos creados en la zona EMEA (Europa, Oriente Medio y África), concretamente en Alemania, Reino Unido, Francia, Italia y Egipto, que conforman 40 profesionales, así como con los existentes en EE.UU. y Japón, para llevar a cabo sus desarrollos. "Aunque nuestros trabajos se centran en la actualidad en proporcionar una tecnología de reconocimiento vocal a través de sistemas telefónicos y de dispositivos embebidos en PDA y automóviles, el objetivo en el futuro va más allá —explica González López—. Se trata de lograr el llamado NLU (Natural Language Understanding) o comprensión del lenguaje natural, que es una tecnología que desarrollamos mediante procedimientos estadísticos y que pretende que un dispositivo pueda seguir un diálogo no forzado, sino natural, con un ser humano". Para ello, según el experto de IBM, se está investigando en la combinación de tecnologías de reconocimiento del habla con otras que detectan los movimientos del interlocutor. "Al utilizar una cámara que grabe los movimientos de los labios del interlocutor, la precisión supera