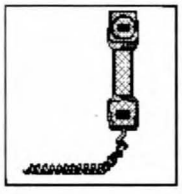


DESARROLLOS FUTUROS EN APLICACIONES COMANDADAS POR VOZ

Xavier Pérez González

El reconocimiento de lenguaje es un campo en constante evolución y desarrollo desde las últimas dos décadas. Tras varios años de estudio e investigación, las exigencias del mercado requieren la implementación de aplicaciones específicas en las que gracias a la interacción automática con los equipos informáticos las empresas pueden obtener una mayor competitividad.

Durante la década de los 80 el campo de reconocimiento de lenguaje estaba orientado al estudio y desarrollo de técnicas que fuesen capaces de procesar y reconocer la información vocal. Tras esta etapa, durante los 90, entramos en una fase de transición, en que surgen aplicaciones que interactúan con el usuario. Hoy día, el reconocimiento de lenguaje está en claro crecimiento, y moviéndose rápidamente hacia el mercado informático. En junio de 1996 IBM presentó en Nueva York el paquete VoiceType 3.0 para Windows 95, permitiendo a los usuarios trabajar con su ordenador personal mediante simplemente la palabra hablada. Sin siquiera tocar el teclado o ratón, los usuarios pueden abrir aplicaciones, dictar memos, enviar mensajes y editar documentos con un reconocimiento suficientemente preciso.

Otras grandes firmas, como Hewlett Packard y Philips, trabajan sobre reconocimiento para desarrollar productos en los que es necesario que se utilice el diálogo como interface de entrada, dado que por su tamaño reducido no es posible la interacción con un teclado. Al mismo tiempo, vemos cómo la telefonía celular reduce cada vez más y más el tamaño de sus terminales, hasta el punto en que el factor limitador es el teclado numérico.

El reconocimiento de lenguaje tiene también aplicaciones inmediatas en aquellos entornos en que el usuario tiene la vista ocupada, y no puede atender al teclado o dispositivo de entrada manual. Por poner un ejemplo, la Agencia Espacial Europea (ESA) planea utilizar el reconocimiento verbal en sus misiones espaciales y asimismo éste resulta de gran utilidad en los pilotajes de aviones militares y comerciales.

Sin embargo, uno de los principales objetivos del reconocimiento de lenguaje es el de facilitar la distribución de información. Frecuentemente tiene lugar una interacción con sistemas telefónicos y gestores de bases de datos que forman parte de un sistema de información y comunicaciones mucho mayor. Es aquí donde las operadoras telefónicas (PTT) juegan un papel importante en el campo del reconocimiento. Este es el caso de Telefónica, AT&T, British Telecom, etc.

Con la entrada en la *era de la información,* inmediatamente surge la necesidad de crear servicios automatizados capaces de interactuar con el usuario. La *Tecnología de la Información* (IT) revoluciona mercados y es con creces una de las más importantes fuentes de ingresos financieros. En este ambiente se crea en Europa el organismo I*MEurope (*Information Market Europe*), para hacer frente a la corriente de cambios que tiene lugar en la sociedad.

I*M-Europe actúa como intermediario para dar soporte a las acciones de la Dirección General de Telecomunicaciones, Mercado de la Información y Valorización de la Investigación (DG-XIII) de la Comisión Europea, cuyo objetivo es el de estimular el mercado europeo de servicios electrónicos de información y las industrias de contenido multimedia. Desde I*M-Europe se define el *Programa de Aplicaciones Telemáticas*, que queda dividido en

Figura 1.- El organismo I*M Europe concibe el Programa de Aplicaciones Telemáticas

XAVIER PÉREZ GONZÁLEZ *es estudiante de PFC en el Grupo de Procesado Digital de la Señal, Universidad Politécnica de Cataluña.*
E-mail: algonza@gps.tsc.upc.es

tres grandes campos: *Ingeniería de Información, Ingeniería de Lenguaje y Aplicaciones para Bibliotecas.*

Gracias al progreso alcanzado en la ingeniería de lenguaje, se pueden implementar teleservicios accionados por voz de gran aplicación en:

· **servicios de información** (información del horario de trenes,...)

· **servicios de transacción** (televentas,...)

· **servicios de procesado de llamada** (correo por voz,...)

Existen muchas empresas europeas activas en la creación de tales servicios y que ofrecen la tecnología de voz necesaria para ello. Para la implementación de la tecnología de procesado de voz y de aplicaciones flexibles de lenguaje (como el reconocimiento de lenguaje o la verificación de persona), hacen falta recursos específicos del lenguaje hablado, concretamente bases de datos de voz y léxico que aporten medios de flexibilidad y facilidad de adaptación a nuevo vocabulario, en contraposición a los métodos utilizados hasta ahora, en que la creación de bases de datos de vocabulario específico era un prerrequisito para la implementación de un reconocedor de lenguaje.

Con el objetivo de fomentar la creación de esta serie de recursos y de alcanzar el grado de competitividad de las compañías norteamericanas, que parten de un extenso mercado monolingüe, se creó el consorcio *SpeechDat*, que pretende abrir el camino por el que se muevan las empresas europeas, que parten de un mercado multilingüe. El proyecto SpeechDat es una iniciativa financiada por la CEC (Comisión de las Comunidades Europeas), LRE-63314, que da cobertura a los campos de producción, estandarización, evaluación y diseminación de los SLR (*Spoken Language Resources*, Recursos de Lenguaje Hablado). Uno de sus principales objetivos es el de ayudar a crear Infraestructura Europea para la distribución y evaluación de los Recursos de Lenguaje.

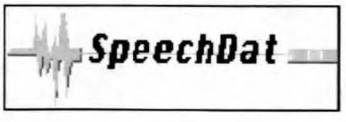


Figura 2.- logotipo de *SpeechDat*

Figura 3.- Consorcio *SpeechDat.* En negrita se marcan las instituciones participantes en la elaboración de *SpeechDat(I)* en castellano

Los SLR son bases de datos de voz que incluyen anotaciones adjuntas, léxico de pronunciación y materiales de modelado del lenguaje que son necesarios para el desarrollo y uso del reconocimiento de lenguaje y la tecnología de síntesis de voz. Los SLR son necesarios por una parte para desarrollar aplicaciones sistemas de diálogo comandados por voz y basados en la tecnología disponible de reconocimiento de lenguaje y por otra parte para desarrollar una tecnología de reconocimiento de voz que soporte lenguaje espontáneo y real, y lleve a productos de la década de los 2000. A corto plazo, las compañías europeas con actividad en el área de aplicaciones comandadas por voz tendrán una mayor participación en el sector de las telecomunicaciones, ya que en Europa existe una fuerte base de productos en este sector. Los teleservicios, que estarán parcial o totalmente automatizados gracias a la utilización de tecnología de voz actual, abarcarán un mercado de varios billones de Euros anuales sólo en Europa. Habrá una gran competencia con las compañías y operadoras estadounidenses, que se beneficiarán de su gran base económica monolingüe y además pueden tomar provecho de la liberalización del mercado de telecomunicaciones europeo.

SpeechDat surge de una iniciativa financiada por el IV Programa de Estructura de la Unión Europea, Aplicaciones Telemáticas, cuyo objetivo es el de crear bases de datos de voz con amplia cobertura y aplicación de las diferentes lenguas habladas en Europa. Para cada lengua se aporta información suficiente para dar pie a una gran variedad de:

·**aplicaciones:** palabras y comandos orientados a aplicaciones, oraciones ricas fonéticamente, pronunciaciones espontáneas.

·**estilos de diálogo:** órdenes y comandos, lenguaje espontáneo, lenguaje seleccionado.

·**influencias ambientales:** red de telefonía móvil y fija.

Se pretende que la base pueda ser utilizada para desarrollar, entrenar y evaluar sistemas robustos de reconocimiento del habla y de verificación de lenguaje para aplicaciones a medio plazo.

La duración del proyecto SpeechDat tal y como fue concebida va desde 1994 hasta 1998. Temporalmente se compone de dos fases (I y II), habiéndose iniciado ya la segunda de ellas. El proyecto *SpeechDat I* se inició en 1994, y consiste básicamente en una base de 1000 locutores para 7 idiomas europeos, que son danés, inglés, francés, alemán, italiano, portugués y español.

SpeechDat (I) en castellano

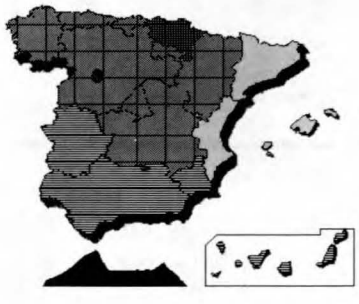
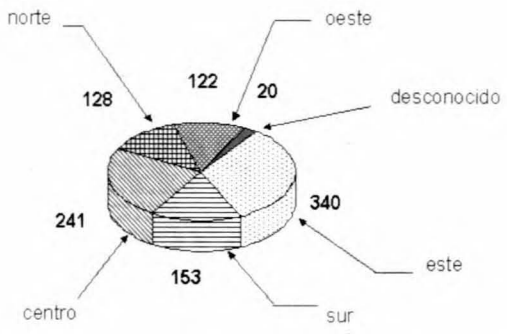


Figura 4: Mapa de distribución de población de locutores de SpeechDat

En la actualidad, *SpeechDat (I) en castellano* consta de una base de 1004 locutores grabada sobre la red de telefonía fija. Cada locutor pronuncia 43 elocuciones, entre las que están el nombre de la persona, el nombre de la provincia donde ha vivido más tiempo, números en diversos formatos, deletreos, fechas, horas, etc. Para la adquisición de la base, los locutores telefoneaban a un sistema de adquisición de voz, leyendo un texto que se les había asignado. La grabación fue realizada por Vocalis Ltd., en Inglaterra, donde ya se disponía de facilidades RDSI de banda estrecha.

Por el tamaño de la base, se consigue que haya una cierta abundancia de variantes dialectales del territorio español, diferentes edades de los locutores, y un equilibrio entre hombres y mujeres. Los locutores de la base en castellano se distribuyen según la estadística de la figura 5.



centro: Ávila, Burgos, Cantabria, Ciudad Real, Guadalajara, Huesca, La Rioja, León, Madrid, Segovia, Soria, Teruel, Toledo, Valladolid, Zaragoza.
norte: Álava, Guipúzcoa, Navarra, Vizcaya.
sur: Almería, Archirona, Badajoz, Buenos Aires, Cádiz, Córdoba, Ceuta, Granada, Jaén, Las Palmas GC, Málaga, Murcia, SC Tenerife, Tánger, Tetuán.
este: Alicante, Baleares, Barcelona, Girona , Lleida, Tarragona, Valencia.
oeste: Asturias, La Coruña, Lugo, Orense, Oviedo, Pontevedra.

Figura 5: Distribución de la procedencia de locutores de la base SpeechDat.

Si realizamos el mismo estudio para conocer los grupos de edad se obtiene la estadística de la figura 6.

Como especificación, el consorcio exige que en el conjunto global de la base debe haber como mínimo un 20 % de locutores cuya edad se ubique en los grupos 17-30, 21-45 y 46-60 años. Además un máximo del 40 % de los locutores serán menores de 17 o mayores de 60 años.

Desarrollos futuros: SpeechDat II

La segunda fase, *SpeechDat II*, se puso en marcha en marzo de 1996 y durará hasta febrero de 1998. Con ella se dará cobertura de las 11 lenguas oficiales europeas (danés, holandés, inglés británico, finés, francés, alemán, griego, italiano, portugués, sueco y español), además de los idiomas noruego, esloveno, galés y variantes específicas del holandés, francés, alemán y sueco. Para cada lengua oficial se creará una base de datos de 5000 locutores grabada sobre la red telefónica fija y de 1000 locutores para las variantes, con el objetivo de ser utilizada en el entrenamiento y test de los sistemas de reconocimiento. Además, algunas de las lenguas (un total de 5) dispondrán de una base de 1000 locutores grabada sobre la red de telefonía móvil, para el desarrollo y test de reconocedores bajo una red móvil de datos, y de una base de verificación de locutor (3 lenguas), en la que un grupo de usuarios realizarán repetidas llamadas, para permitir el entrenamiento y test de los sistemas de verificación. En la actualidad existen propuestas de ampliación de SpeechDat hacia los países del este (en una extensión que se denominaría SpeechDat-E). Sin embargo razones presupuestarias y el hecho de que no existan organismos competentes de ayudas a la investigación en los países del Este hacen que la propuesta se retrase hasta nuevo acuerdo.

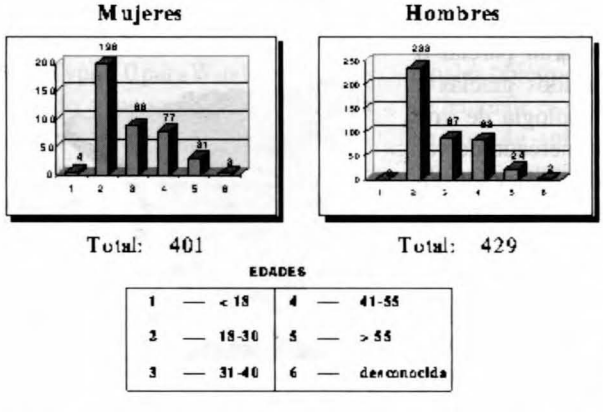


Figura 6: Distribución por edades de los locutores de SpeechDat.

Información y Bibliografía:

Profesor de contacto: Asunción Moreno (*asuncion@gps.tsc.upc.es*), Grupo de Procesado Digital de la Señal.

-***http://www2.echo.lu/dg13/en/dg13tasks.html:*** Dirección General de Telecomunicaciones, Mercado de la Información y Valorización de la Investigación (DG-XIII).

-***http://europa.eu.int:*** WWW de la Comisión Europea.

-***http://www2.echo.lu/telematics/home.html:*** Programa de Aplicaciones Telemáticas de la Comisión Europea.

-***http://www2.echo.lu:*** Institución I'M Europe definida por la Comisión Europea.

-***http://www2.echo.lu/langeng/en/le2/speechda.html:*** Página del proyecto SpeechDat.