

RECONOCIMIENTO DEL LOCUTOR EN TELEFONIA: ACTIVIDADES DEL PROYECTO EUROPEO COST 250*

Javier Hernando¹, Carmen García-Mateo², Leandro Rodríguez-Liñares²
Joaquín González-Rodríguez³, Javier Ortega-García³

¹Centro TALP, Depto. TSC, UPC, Barcelona, javier@gps.tsc.upc.es

²Depto. TC, UV, Vigo, ³Depto. DIAC, UPM, Madrid

RESUMEN

El objetivo de esta comunicación es presentar las actividades realizadas desde noviembre de 1994 dentro del proyecto "Speaker Recognition in Telephony", financiado por la Comunidad Europea en el marco del programa "European Cooperation in the Field of Scientific and Technical Research" (COST), con referencia COST250. Los países participantes en el proyecto COST250 son Bélgica, Dinamarca, Eslovenia, España, Francia, Grecia, Inglaterra, Irlanda, Italia, Holanda, Portugal, Suecia, Suiza y Turquía.

1. INTRODUCCIÓN

El objetivo principal del proyecto europeo COST 250 "Speaker Recognition in Telephony" [1] es mejorar la tecnología del reconocimiento del locutor mediante:

- la mejora del conocimiento de los problemas
- la evaluación de las técnicas y los productos actuales
- la investigación de nuevas técnicas
- la mejora de los procedimientos de evaluación

El interés se centra en el reconocimiento del locutor sobre línea telefónica, pero contando con que la experiencia obtenida pueda aplicarse a otras áreas como la estandarización de las técnicas de medida y las pruebas forenses. Para abordar estos objetivos, se han adoptado los principios científicos:

- identificación de un conjunto de cuestiones apropiadas

- definición de experimentos diseñados para contestar a estas cuestiones

- publicación del conocimiento obtenido

El trabajo se ha coordinado en cuatro grupos de trabajo:

- Aplicaciones
- Bases de datos
- Algoritmos y Parámetros
- Valoración de Sistemas y Difusión de Resultados

2. APLICACIONES

En el grupo de trabajo de "Aplicaciones" se definieron las siguientes áreas clave:

- revisión de las áreas potenciales de aplicación de las técnicas de verificación de locutores
- definición de los criterios de prestaciones para cada área basados en tasas de error, cantidad de datos de entrenamiento y prueba, y dependencia/independencia del texto

* Este trabajo ha sido subvencionado por los proyectos TIC 98-0685 y TIC 98-0423-C06-01

- definición de una aplicación estándar para comparar diferentes algoritmos de verificación
- estimación del mercado de la verificación de locutores

El objetivo inicial de la acción COST, en principio el reconocimiento de locutores por línea telefónica, no impidió que se consideraran todo tipo de aplicaciones de reconocimiento de locutores, desde aplicaciones bancarias remotas al área forense.

Inicialmente, el grupo de trabajo hizo circular una propuesta preliminar entre los miembros de la acción para obtener una perspectiva del estado del arte actual, así como las diversas opiniones particulares. También se llevó a cabo una revisión de las aplicaciones de reconocimiento de locutores actuales y potenciales.

Finalmente, el grupo consideró las reacciones del usuario ante la interacción con los sistemas automáticos de reconocimiento de locutores. Los resultados del grupo proporcionan información útil sobre el diseño del diálogo para los diferentes subsistemas y recomendaciones para el caso de que el proceso de verificación se encuentre con dificultades.

Una conclusión de los trabajos realizados por el grupo es que los límites de la tecnología de reconocimiento de locutores son la aceptabilidad y la flexibilidad del usuario. Sus puntos débiles son los tiempos de entrenamiento, las tasas de aciertos en general y la sensibilidad a la variabilidad, tanto debida al usuario (por ejemplo, por un resfriado) como al proceso de captura de la voz (por ejemplo, debida al micrófono, a la tarjeta de sonido o al canal en aplicaciones telefónicas).

Varias de las comunicaciones presentadas durante la acción han destacado la necesidad de un diseño cuidadoso de la aplicación, en particular del diálogo con el usuario y la provisión de mecanismos de "escape" para desviar el problema a un operador humano si la autenticación falla

repetidamente. Es especialmente importante no desanimar al público a usar sistemas de reconocimiento del locutor con interfaces lentas y complicadas.

La investigación dirigida a aplicaciones que requieren un alto nivel de seguridad ha revelado que, además de la disminución de los errores, las organizaciones involucradas prefieren configuraciones que puedan encajar en sus actuales sistemas de control de accesos.

Para satisfacer estas restricciones, se ha propuesto que la verificación del locutor debe combinarse con la verificación de la información verbal.

De cara al futuro, la llegada de sistemas multimodales de interfaz hombre-máquina puede impulsar el desarrollo de los sistemas de reconocimiento de locutores. En este tipo de entornos donde la voz habrá reemplazado al teclado como principal método de comunicación, el reconocimiento del locutor será la opción natural de garantizar y restringir el acceso a la información

3. BASES DE DATOS

La voz es el material básico de investigación en las tecnologías del habla. La disponibilidad de buenas bases de datos de voz es crucial para el desarrollo y la valoración de sistemas de procesado de voz. Existen diversas bases de datos de dígitos, palabras y frases, pero la gran mayoría han sido diseñadas para aplicaciones de reconocimiento del mensaje (habla) y no del locutor (hablante) y, por tanto, no son adecuadas para la investigación dentro de este contexto.

Los objetivos del grupo "Bases de datos" son:

- revisión de las bases de datos disponibles que se han usado en reconocimiento del locutor
- caracterización de bases de datos de locutores
- diseño, grabación y producción de una

nueva base de datos, llamada POLYCOST

La base POLYCOST ha sido grabada sobre línea telefónica y diseñada para tareas de reconocimiento de locutores. Se grabó a través de la red telefónica europea entre enero y marzo de 1996. Se utilizó una frecuencia de muestreo de 8 kHz. La base contiene unas 10 sesiones de 134 locutores de 14 países. Cada sesión consta de 15 elocuciones: una para detección DTMF, 10 de dígitos conectados en inglés, 2 con frases fijas en inglés y 2 libres en lengua materna. La lengua materna de la mayoría de locutores no es el inglés, lo cual posibilita experimentos sobre variabilidad intra/inter-locutor/lengua.

4. ALGORITMOS Y PARÁMETROS

Las diferencias entre locutores tienen un doble origen: diferencias en la anatomía de los órganos de producción del habla y diferencias en la forma de producción del habla adquirida por el locutor. El progreso en el modelado de estas diferencias entre locutores y la comparación de las mismas es necesario para la mejora de los sistemas actuales de reconocimiento. El objetivo de del grupo "Algoritmos y Parámetros" de la acción es la identificación de los parámetros que mejor caracterizan la identidad del locutor y el desarrollo de algoritmos de comparación eficientes.

Respecto a la difusión de resultados, se procedió a

- la presentación de comunicaciones en congresos internacionales
- el mantenimiento de una página web
- la promoción de la base POLYCOST
- organización de workshops [3] [4] [5] [6]

[6] Proc. COST Workshop Speaker Recognition in Telephony, pendiente.

6. CONCLUSIONES

La acción COST 250 ha servido de foro de discusión de ideas sobre el diseño y la implementación de bases de datos, algoritmos, sistemas y aplicaciones de reconocimiento de locutores. Los participantes han profundizado en los problemas de esta tecnología y en posibles aproximaciones para abordarlos. También la acción ha servido para proporcionar no sólo a los participantes sino a una comunidad investigadora más amplia una profunda revisión del estado actual del arte, así como una base de datos diseñada para el reconocimiento de locutores con una guía de experimentos de referencia.

REFERENCIAS

- [1] COST 250 “Speaker Recognition in Telephony”, <http://www.fub.it/cost250>
- [2] D. Gibbon, R. Moore, R. Winsky (editores), “Handbook of Standard and Resources for Spoken Language Systems”, Mouton de Gruyter, 1997.
- [3] Proc. COST Workshop Application of Speaker Recognition Techniques in Telephony, Vigo, noviembre 1996.
- [4] Proc. COST Workshop Speech Technology in Public Telephonic Network: Where Are We Today?, Rodas (Grecia), septiembre 1997.
- [5] Proc. COST Workshop Speaker Recognition by Man and Machine: Directions for Forensic Applications, Ankara (Turquía), abril 1998.