

Sistema de comprensión de comunicaciones habladas para el control de tráfico aéreo del proyecto INVOCA

V. Sama Rojo, F. Fernández Martínez, J. Ferreiros López,
J. Macias-Guarasa, R. De Córdoba, J. M. Montero Martínez,
J. Colas Pasamontes*, E. Campos Palarea*, J. M. Pardo Muñoz

Grupo de Tecnología del Habla, Universidad Politécnica Madrid, Ciudad Universitaria s/n

*Laboratorio de Tecnología Hombre-Computador, Universidad Autónoma Madrid
{vsama, efhes, jfl, macias, cordoba, juancho, pardo}@die.upm.es; jose.colas@ii.uam.es

Resumen: El proyecto INVOCA consiste en un sistema de comprensión de comunicaciones habladas tierra-aire para el control de tráfico aéreo. Este sistema emplea una arquitectura de dos reconocedores de habla en paralelo con un módulo de detección de idioma (para inglés y castellano). La salida del reconocedor pasa a un módulo de comprensión basado en reglas dependientes de contexto que extrae los conceptos clave de cada tarea.

Palabras clave: reconocimiento, habla, control aéreo, comprensión, modelo del lenguaje

1 Introducción

El proyecto INVOCA, tiene como objetivos primarios analizar las posibilidades que ofrece el estado del arte en las tecnologías de reconocimiento del habla para su aplicación a los sistemas de control de tráfico aéreo en torre, y el análisis de viabilidad de su incorporación en el Sistema Automatizado de Control de Tránsito Aéreo (SACTA), se trata por lo tanto de una evaluación de tecnología.

La funcionalidad del sistema desarrollado es la detección de datos clave en canales tierra-aire, sobre habla espontánea en inglés y castellano, siguiendo una fraseología oficial. Principalmente se ha trabajado en las posiciones de control aéreo disponibles en el Aeropuerto de Madrid-Barajas: **Arribadas** (aterrizajes), **Autorizaciones** (autorización del plan de vuelo), **Despegues** (autoriza los despegues, como su propio nombre indica), **Norte** (control de tráfico rodado en la parte norte del aeropuerto) y **Sur** (control de la zona sur del aeropuerto).

El mayor esfuerzo del proyecto se centró en la tarea de Autorizaciones ya que es la que tiene más contenido en cuanto a datos relevantes, como pueden ser la altura de vuelo, tipo de salida, y otros datos de los planes de vuelo.

2 Descripción del sistema

2.1 Arquitectura

2.1.1 Front-End

El sistema consta de un primer módulo o *front-end* que convierte la señal acústica en un conjunto de vectores. Este *font-end* está compuesto a su vez por un detector, cuyo cometido es la detección de voz / no voz a la entrada del sistema, y el parametrizador, que lleva a cabo la parametrización de la información acústica segmentada.

2.1.2 Módulo de reconocimiento

La salida del *front-end* pasa al módulo de reconocimiento compuesto por dos reconocedores, uno para inglés y otro para castellano. Se obtienen las frases reconocidas en el idioma correspondiente en ambos reconocedores y se pasan como entradas al módulo de detección de idioma.

2.1.3 Módulo de detección de idioma

Este módulo decide el idioma al que corresponde la frase que está siendo procesada.

La decisión puede tomarse aplicando modelos de lenguaje en base a medidas de perplejidad sobre los resultados de ambos reconocedores. En nuestro caso, debido al gran peso dado al modelado de lenguaje durante el

proceso de reconocimiento¹, la decisión puede tomarse directamente en base a la diferencia de *scores* entre ambos reconocedores.

2.1.4 Módulo de comprensión

El módulo de comprensión extrae los conceptos clave de la tarea a partir de la salida del reconocedor. Esta comprensión está basada en reglas dependientes de contexto, y se utilizan diccionarios específicos para cada tarea, y para cada idioma. Estos diccionarios están etiquetados semánticamente en función de la información que se desea extraer. A todas las palabras que no aportan información a la tarea, se les asigna una categoría “basura”, al igual que a las palabras fuera de vocabulario que pudiesen aparecer, aunque esto pueda suponer una posible pérdida de información.

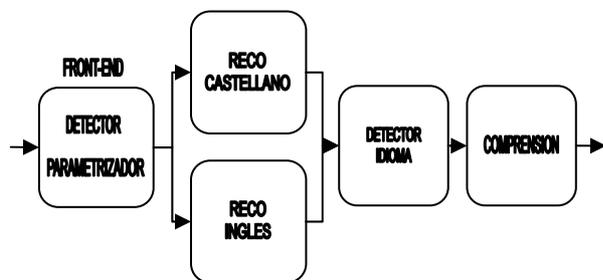


Figura 1: Diagrama de bloques

2.2 Modelos de lenguaje

Se emplean modelos de lenguaje de tipo estocástico basados en bigramas. Para la tarea de Autorizaciones tenemos un vocabulario de 793 palabras para inglés y 1104 para castellano. Contamos con 36 palabras sin modelo gramatical en inglés, y en castellano encontramos 86. La no disponibilidad de modelo gramatical para esas palabras es debida a la no aparición de las mismas en entrenamiento. Las perplejidades para cada idioma respectivamente en entrenamiento son 10,4 y 10,5.

3 Base de Datos

La Base de Datos se compone de los archivos de sonido de cada frase o intervención del controlador en formato WAV con una frecuencia de muestreo de 8 KHZ y 16 bits por muestra, así como la transcripción de cada frase en archivos con formato SAM.

¹ Los pesos para el modelo de lenguaje son 9.5 para castellano y 11 para inglés

FRECUENCIA	FICHEROS		
	SP	EN	TOTAL
Arribadas	213	63	276
Autorizaciones	5091	3256	8347
Despegues	234	75	309
Norte	352	72	424
Sur	235	127	362
TOTAL			9718

Figura 2: Composición Base de Datos

4 Resultados obtenidos

En lo que respecta a la evaluación, los datos más relevantes son los de *word accuracy* y de *concept accuracy*. También consideramos importantes las frases procesadas perfectamente.

	CASTELLANO		INGLÉS	
	%wa	% perf	%wa	% perf
AUTORIZACIONES				
W.A	86'26	33'6	73'26	17'4
C.A	77'73	53'1	47'9	35'5
MEDIA RESTO FRECUENCIAS				
W.A	75'52	21'27	74'42	15'92
C.A	56'72	45'03	46'83	29'26

En la tabla superior se ofrecen los resultados de Autorizaciones y la media de los resultados del resto de frecuencias (Arribadas, Despegues, Norte y Sur).

En la actualidad está en proceso la evaluación del sistema en la torre de control de Madrid-Barajas.

4.1 Agradecimientos

Este proyecto ha sido financiado por AENA, y las pruebas han sido realizadas en SENASA y las instalaciones del Aeropuerto de Madrid-Barajas.

Agradecemos especialmente a Germán González Antequera, y Myriam Santamaría González, su disponibilidad y colaboración en todos los aspectos del proyectos.

También ha participado personal del Grupo Tecnológico del Habla, a los que agradecemos igualmente toda su aportación, son: Sergio Díaz, María José Pozuelo Varo, Gregoire Prime.