

# JUSTIFICACIÓN TÉCNICA PARA LA ADJUDICACIÓN DIRECTA DEL EXPEDIENTE 2023-06-01-GEST-PPT-REC VOZ-PILOTO-06.23

Código:	2023-06-15-INF.JUSTIF-REC VOZ-PILOTO-06.23
Fecha:	30/06/2023

Este documento presenta las razones que requieren que se utilice el procedimiento descrito en el apartado 15.b de las Instrucciones Internas de Contratación de Crida A.I.E. y se envíe la actuación 2023-06-01-GEST-PPT-REC VOZ-PILOTO-06.23 únicamente al proveedor EML Speech Technology GmbH (formerly European Media Laboratory GmbH), EML.

Clasificación:	Confidencial
----------------	--------------

**Página dejada intencionalmente en blanco**

### *Hoja de Identificación del documento*

Título:	Justificación Técnica para la adjudicación directa del expediente 2023-06-01-GEST-PPT-REC VOZ-PILOTO-06.23
Código:	2023-06-15-INF.JUSTIF-REC VOZ-PILOTO-06.23
Fecha:	30/06/2023
Fichero:	2023-06-01-INF.JUSTIF-REC VOZ-PILOTO-06.23.DOCX

Autor:	Raquel García; Adrián Fabio Bracero
Revisor:	Ángel Martínez Mateo
Aprobado:	José Miguel de Pablo

Versiones:			
Numero	Fecha	Autor	Comentarios
0.1	28/06/2023	Raquel García	Primer borrador
1.0	30/06/2023	Raquel García, Adrián Fabio	Versión definitiva

## *Resumen Ejecutivo*

Este documento presenta las razones que requieren que se utilice el procedimiento descrito en el apartado 15.b de las Instrucciones Internas de Contratación de Crida A.I.E. y se envíe la actuación 2023-06-01-GEST-PPT-REC VOZ-PILOTO-06.23 únicamente al proveedor EML Speech Technology GmbH (formerly European Media Laboratory GmbH).

En él se detallan los antecedentes, objetivos del expediente para el próximo periodo y justificación para utilizar este procedimiento.



## *Índice de Contenidos*

1	Justificación Técnica .....	6
1.1	Antecedentes .....	6
1.2	Objetivos y metas .....	7
1.3	Justificación .....	7
1.4	Procedimiento de contratación y empresa propuesta para la adjudicación .....	9

# 1 Justificación Técnica

## 1.1 Antecedentes

CRIDA junto con ENAIRE (antes AENA) ha estado trabajando durante los últimos años en un sistema llamado VOICE capaz de detectar los eventos que realizan los controladores de tráfico aéreo a través de comunicaciones de voz reales en los centros de control aéreos. El prototipo VOICE se desarrolló en el marco de la licitación DNA 344/2009 AENA, proporcionado por Unysis. Un módulo clave de este sistema es el sistema de Reconocimiento Automático de Voz, que procesa audios de entrada y proporciona su transcripción como salida. Esta transcripción alimenta el algoritmo de detección de eventos que utiliza palabras clave para funcionar, por tanto la precisión de la transcripción es un factor determinante en el índice de detección de eventos obtenidos por el conjunto del sistema.

En el prototipo de VOICE que se creó, el Reconocedor Automático de Voz (EML Transcription Server) consistente en un servidor de transcripción adaptable a cada empresa que incluye el reconocedor de voz y los correspondientes modelos acústicos y de lenguaje fue desarrollado por EML Speech Technology GmbH (formerly European Media Laboratory GmbH), EML. Esta pieza clave del prototipo es un factor determinante en el rendimiento final del sistema y se ha ido evolucionando a lo largo de los años y a través de sucesivos entrenamientos.

El primer prototipo se entrenó con grabaciones del centro de control de área de Madrid, creándose un modelo acústico y de lenguaje optimizado para este ambiente. En años sucesivos se crearon nuevos modelos correspondientes al centro de control de área de Barcelona, entorno de Torre y comunicaciones de piloto. EML desarrolló estos modelos y como consecuencia de varios análisis y pruebas realizadas junto con CRIDA se decidió evolucionar el Reconocedor Automático de Voz:

- creando un único modelo para el entorno de centro de control de área, ya que la combinación de los modelos de Madrid y Barcelona enriquecían el modelo acústico mejorando la tasa de reconocimiento de palabras y disminuyendo la tasa de errores.
- desarrollando un modelo reconocedor de voz basado en clases para mejorar la tasa de transcripción correcta del identificativo de los vuelos (callsign) y los nombres de los "waypoints". El identificativo de un vuelo es clave para poder asignar el evento de control que el controlador aéreo ha realizado sobre él, por otro lado, el reconocimiento de los "waypoints" es clave para detectar eventos como directos o esperas.

El uso de un modelo basado en clases implicó además la evolución del prototipo para ser capaz de actualizar dinámicamente la tabla de posibles indicativos y "Waypoints".

- Incluyendo nuevas tecnologías de separación e identificación de ruidos VAD ("Voice Activity Detection") y de identificación de hablantes SD ("Speaker Diarization").
- Desarrollando un modelo para su uso en tiempo real - streaming.
- Creando un modelo de piloto centrado en el reconocimiento en tiempo real de identificativos o callsign del avión. Este modelo se diferencia del de controlador debido en la calidad de los audios y en la variedad de acentos.

## 1.2 Objetivos y metas

CRIDA desea continuar la evolución del sistema VOICE para su aplicación en el entorno piloto, ya que ENAIRE está interesado en la identificación de vuelos que contactan con el controlador.

Estas mejoras se realizarán mediante:

- La mejora del modelo centrado en conversaciones de piloto. Este nuevo modelo utilizará el modelo desarrollado en la campaña 2019-2020. La tasa de detección deseada para este nuevo modelo es de un 80% en *Word Correct Rate* (WCR) y 65% en *callsigns* (CDR). El modelo de la campaña 2021 no mejoró respecto a este debido a la mala calidad de las transcripciones, por lo que se descartó.
- Continuar la exploración iniciada en la campaña 2019-2020 sobre la aplicación de nuevas tecnologías como redes neuronales para el procesado de lenguaje natural y al detector de eventos.
- Prolongar la licencia de la instalación de un entorno de Transcripción en las oficinas de CRIDA que tenga las mismas características que el que se ha estado usando hasta ahora online (online EML Transcription Workplace).
- Mejorar el modelo en tiempo real ("streaming"), debido al creciente interés de ENAIRE por dicho modelo. Se buscará reducir el tiempo en el que se provee el primer resultado parcial menos de 2 segundos.

## 1.3 Justificación

El procedimiento de adjudicación a la empresa EML se justifica por la continuidad del proyecto VOICE, como resultado del contrato DNA 344/2009, dado que EML participó en él como empresa externa para el suministro del servidor y modelo de reconocimiento de voz. Relación que continuó a través del contrato CRIDA GEST-PPT-ASR-EVOLUTION-11.12 de 2012 a 2014, dónde EML fue el contratista principal y donde se obtuvieron resultados satisfactorios en mejoras concretas del sistema; en GEST-PPT-ASR-EXPANSION-10.14 donde se analizaron, probaron, planificaron y desarrollaron los primeros pasos para la evolución hacia un modelo de clases, y se desarrolló un primer modelo de piloto; en CRIDA GEST-PPT-ASR CLASS-11.15 donde se completó el desarrollo del modelo de clases, y se creó un modelo acústico y de lenguaje multi-dominio; en GEST-PPT-REC VOZ-12.2016 donde se amplió el modelo acústico y de lenguaje al entorno TMA y se adaptó el servidor para dar servicio en *streaming*; en GEST-PPT-ASR IMPROVEMENT-12.2017 dónde se incorporaron nuevas tecnologías para la segmentación de las transcripciones por interlocutor y para hacer que las transcripciones sean menos sensibles a los ruidos de fondo; en GEST-PPT-ASR-EVOLUTION-12.2018 dónde se empezó la aplicación de redes neuronales en el modelo acústico del controlador y se desarrolló la tecnología en tiempo real, *streaming*, para que los tiempos de reacción fueran más funcionales; en GEST-PPT-ASR-EVENT-12.2019 donde se creó un modelo de piloto funcional; 2021-05-21-GEST-PPT-VOICE-05.21 donde se desarrolló un modelo de detección de eventos basado en redes neuronales.

Además, de manera periódica, CRIDA realiza una exploración del estado del arte, es decir, realiza un análisis de las tasas de reconocimiento de modelos de transcripción comerciales (e.g., *Amazon WS Transcribe*) en el entorno ATM. El último de estos estudios se llevó a cabo en 2019 para los modelos de AWS y Google Cloud arrojando tasas de WCR del 48% y del 52%, respectivamente. Si bien es cierto, son varios los factores que se consideraron un impedimento a la hora de decidir usar este tipo de transcritores comerciales.

En primer lugar, estas soluciones no pueden ser desplegadas en local, condición hasta ahora que se ha de cumplir teniendo en cuenta el futuro despliegue de estas soluciones en, por ejemplo, centros de control. En segundo lugar, la provisión del servicio de control en España requiere que el transcriptor sea capaz de trabajar en dos idiomas. Sin embargo, el uso de los servicios comerciales testeados hasta el momento obliga a definir el idioma e incluso el acento de la comunicación, haciendo inviable el uso operacional en la provisión de servicio de control en España. En última instancia, en el caso del servicio de *Amazon WS Transcribe*, no resulta posible desarrollar un modelo propio – i.e., entrenar el ya existente modelo comercial con comunicaciones del entorno ATM – lo que dificulta en gran medida, la mejora de las tasas de reconocimiento.

Por todo ello, alcanzar el estado actual del sistema Voice con otro sistema de reconocimiento de voz y los correspondientes modelos acústicos y de lenguaje y fonético requeriría un gran esfuerzo técnico, económico y de personal, además de la incertidumbre asociada a la posible futura implementación de cualquiera de estos modelos. Por este motivo, se plantea la necesidad de seguir con la misma licencia. EML es la única empresa que posee la licencia.

## **1.4 Procedimiento de contratación y empresa propuesta para la adjudicación**

De acuerdo con los requerimientos y circunstancias planteadas, el responsable del expediente VOICE requiere que por razones técnicas se utilice el procedimiento descrito en el apartado 15.b de las Instrucciones Internas de Contratación de Crida A.I.E. y se envíe la actuación 2023-06-01-GEST-PPT-REC VOZ-PILOTO-06-23 únicamente al proveedor EML Speech Technology GmbH, EML, anteriormente denominada "European Media Laboratory" por no ser posible que otras empresas cumplan con los requisitos.

En Madrid, a 30 de junio de 2023,

<p>Revisado por:</p>  <p>Adrián Fabio Bracero Miembro del proyecto</p>	<p>Revisado por:</p>  <p>Ángel Martínez Mateo Miembro del proyecto</p>	<p>Aprobado por:</p>  <p>José Miguel de Pablo Guerrero Director de CRIDA</p>
--	---	---