

# Tecnologías del habla: nuevas oportunidades para los archivos de televisión.

Virginia Bazán<sup>1</sup>, Eduardo Lleida<sup>2</sup>, Carmen Pérez<sup>3</sup>, Manuel Gómez<sup>4</sup>, Alberto de Prada<sup>5</sup>,

<sup>1</sup> <https://orcid.org/0000-0003-4920-2212> + Fondo Documental RTVE, Madrid, España. [virginia.bazan@rtve.es](mailto:virginia.bazan@rtve.es).

<sup>2</sup> <https://orcid.org/0000-0001-9137-4013> + Instituto de Investigación en Ingeniería de Aragón. Universidad de Zaragoza, Zaragoza, España. [lleida@unizar.es](mailto:lleida@unizar.es).

<sup>3</sup> <https://orcid.org/0000-0002-3325-0546> + Innovación y Estrategia Tecnológica RTVE, Madrid, España. [carmen.perez@rtve.es](mailto:carmen.perez@rtve.es).

<sup>4</sup> <https://orcid.org/0000-0001-5128-5905> + Área de Desarrollo Digital RTVE, Madrid, España. [manuel.gomez@rtve.es](mailto:manuel.gomez@rtve.es)

<sup>5</sup> <https://orcid.org/0000-0002-9140-5745> + Fondo Documental RTVE, Madrid, España. [alberto.deprada@rtve.es](mailto:alberto.deprada@rtve.es).

## Resumen

La sobreabundancia de contenidos audiovisuales y la dificultad creciente para identificar y describir esos contenidos de forma eficiente ha convertido la Inteligencia Artificial en un objeto de deseo para los archivos de televisión. Los futuros procesos de generación automática de metadatos en los archivos se fundamentarán en tres tecnologías complementarias: visión artificial, tecnologías del habla y procesamiento del lenguaje natural. La aplicación de estas tecnologías no solo facilitará el acceso a un volumen creciente de contenidos audiovisuales, sino que además permitirá alcanzar un nivel de detalle en el análisis hasta ahora impensable en los archivos de televisión. En este nuevo horizonte, las funciones esenciales de los documentalistas se verán una vez más alteradas, ya que los procesos automatizados requerirán de su implicación en las fases de entrenamiento de los algoritmos en los que se basan estas tecnologías, así como en el control de la calidad de los datos generados de forma automática.

En este contexto, en el que la Inteligencia Artificial brinda a los archivos de televisión nuevas oportunidades, la Corporación RTVE y la Universidad de Zaragoza formalizaron en julio de 2017 la creación de una Cátedra cuyo principal objetivo es la realización de actividades de formación, investigación, estudio y divulgación de las Tecnologías de la Información y de las Comunicaciones relacionadas con el Big Data y su aplicación al análisis de contenidos audiovisuales y sonoros. En la Cátedra, junto al equipo de la Universidad de Zaragoza colaboran distintas áreas de RTVE como son: el Fondo Documental, el área de Innovación y Estrategia Tecnológica y el Área de Desarrollo Digital. En 2018, como parte de sus actividades, y con el objetivo de impulsar la investigación en tecnologías del habla en español, la Catedra lanzó el RTVE Iberspeech Challenge 2018. Un reto tecnológico que puso a disposición de la comunidad científica más de 500 horas de contenidos audiovisuales y que permitió a los distintos grupos de investigación, nacionales e internacionales, poner a prueba sus algoritmos en tres tareas diferenciadas: transcripción de habla a texto (*speech to text*), diarización de hablantes y diarización multimodal. Los resultados obtenidos han demostrado las dificultades tecnológicas que todavía deben superarse tanto en la transcripción voz a texto como en la diarización, tanto de hablante como multimodal. Estos resultados deben abordarse además desde la perspectiva del usuario. La pregunta ¿Cuál es el grado de tolerancia de los usuarios antes los errores de

transcripción automática? debe responderse con estudios de usuarios en tres contextos diferenciados: la accesibilidad, la edición y la preservación de los contenidos.