



Resumen

Durante la última década, los medios de comunicación han experimentado una revolución, alejándose de la televisión convencional hacia las plataformas de contenido bajo demanda. Además, esta revolución no ha cambiado solamente la manera en la que nos entretenemos, si no también la manera en la que aprendemos. En este sentido, las plataformas de contenido educativo bajo demanda también han proliferado para proporcionar recursos educativos de diversos tipos. Estas nuevas vías de distribución de contenido han llegado con nuevos requisitos para mejorar la accesibilidad, en particular las relacionadas con las dificultades de audición y las barreras lingüísticas.

Aquí radica la oportunidad para el reconocimiento automático del habla (RAH) para cumplir estos requisitos, proporcionando subtítulo automático de alta calidad. Este subtítulo proporciona una base sólida para reducir esta brecha de accesibilidad, especialmente para contenido en directo o *streaming*. Estos sistemas de *streaming* deben trabajar bajo estrictas condiciones de tiempo real, proporcionando la subtitulación tan rápido como sea posible, trabajando con un contexto limitado. Sin embargo, esta limitación puede conllevar una degradación de la calidad cuando se compara con los sistemas para contenido en diferido u *offline*.

Esta tesis propone un sistema de RAH en *streaming* con baja latencia, con una calidad similar a un sistema *offline*. Concretamente, este trabajo describe el camino seguido desde el sistema *offline* híbrido inicial hasta el eficiente sistema final de reconocimiento en *streaming*. El primer paso es la adaptación del sistema para efectuar una sola iteración de reconocimiento haciendo uso de modelos de lenguaje estado del arte basados en redes neuronales. En los sistemas basados en múltiples iteraciones estos modelos son relegados a una segunda (o posterior) iteración por su gran coste computacional. Tras adaptar el modelo de lenguaje, el modelo acústico basado en redes neuronales también tiene que adaptarse para trabajar con un contexto limitado. La integración y la adaptación de estos modelos es ampliamente descrita en esta tesis, evaluando el sistema RAH resultante, completamente adaptado para *streaming*, en conjuntos de datos académicos extensamente utilizados y desafiantes tareas basadas en contenidos audiovisuales reales. Como resultado, el sistema proporciona bajas tasas de error con un reducido tiempo de respuesta, comparables al sistema *offline*.

