

# Resumen del trabajo de tesis

## “Confidence Measures for Automatic and Interactive Speech Transcription”

### Medidas de Confianza para el Reconocimiento del Habla Automático e Interactivo

por Isaias Sanchez-Cortina.

Supervisores: Alfons Juan Ciscar y J. Alberto Sanchis Navarro.

Este trabajo contribuye en el campo del *reconocimiento automático del habla* (RAH). Y en especial, en el de la *transcripción interactiva del habla* (TIH) y el de las *medidas de confianza* (MC) para RAH. Los objetivos principales son los siguientes:

1. Diseño de métodos y herramientas TIH para mejorar las transcripciones automáticas.
2. Evaluar los métodos y herramientas TIH empleando tareas de transcripción realistas extraídas de grandes repositorios de vídeos educativos.
3. Mejorar la fiabilidad del TIH mediante la mejora de las MC.

## Resumen

El *reconocimiento automático del habla* (RAH) es una tarea crucial en una amplia gama de aplicaciones importantes que no podrían realizarse mediante transcripción manual. El RAH puede proporcionar transcripciones rentables en escenarios de creciente impacto social como el de los  *cursos abiertos en línea masivos* (MOOC), para el que la disponibilidad de transcripciones es crucial, incluso cuando no son completamente perfectas. Las transcripciones permiten la automatización de procesos como buscar, resumir, recomendar, traducir; hacen que los contenidos sean más accesibles para hablantes no nativos y usuarios con discapacidades, etc. Incluso se ha comprobado que mejora el rendimiento de los estudiantes que aprenden de videos con subtítulos incluso cuando estos no son completamente perfectos. Desafortunadamente, la tecnología RAH actual aún está lejos de la precisión necesaria.

Las transcripciones imperfectas resultantes del RAH pueden ser supervisadas y corregidas manualmente, pero el esfuerzo puede ser incluso superior al de la transcripción manual. Con el fin de aliviar este problema, esta tesis presenta un novedoso sistema de *transcripción interactiva del habla* (TIH). Este método TIH consigue reducir el esfuerzo de semi-supervisión siempre que sea aceptable una pequeña cantidad de errores; además mejora a la par los modelos RAH subyacentes.

Con objeto de transportar el marco propuesto para MOOCs, también se investigaron otros métodos de interacción inteligentes que involucran esfuerzo limitado por parte del usuario. Además, se introdujo un nuevo método que aprovecha las interacciones para mejorar aún más las partes no supervisadas (ASR con *búsqueda restringida*).

La investigación en TIH llevada a cabo se desplegó en una plataforma web con el que fue posible producir un número masivo de transcripciones de videos de dos conocidos repositorios, videoLectures.net y poliMedia.

Por último, el rendimiento de la TIH y los sistemas de RAH se puede aumentar directamente mediante la mejora de la estimación de la *medida de confianza* (MC) de las palabras transcritas. Por este motivo se desarrollaron dos contribuciones: un nuevo modelo discriminativo *logístico* (LR); y la adaptación al locutor de la MC para los casos en que es posible, como por ejemplo en MOOCs.