Summary of the Ph.D thesis
"Confidence Measures for Automatic and Interactive Speech Transcription"

by Isaias Sanchez-Cortina.
Advisors: Alfons Juan i Ciscar and J. Alberto Sanchis Navarro.

This thesis work contributes to the field of the *Automatic Speech Recognition* (ASR). And particularly to the *Interactive Speech Transcription* and *Confidence Measures* (CM) for ASR. The main goals of this thesis work can be summarised as follows:

1. To design IST methods and tools to tackle the problem of improving automatically generated transcripts.

2. To assess the designed IST methods and tools on real-life tasks of transcription in large educational repositories of video lectures.

3. To improve the reliability of the IST by improving the underlying (CM).

Abstracts

The *Automatic Speech Recognition* (ASR) is a crucial task in a broad range of important applications which could not accomplished by means of manual transcription. The ASR can provide cost-effective transcripts in scenarios of increasing social impact such as the *Massive Open Online Courses* (MOOC), for which the availability of accurate enough is crucial even if they are not flawless. The transcripts enable search-ability, summarisation, recommendation, translation; they make the contents accessible to non-native speakers and users with impairments, etc. The usefulness is such that students improve their academic performance when learning from subtitled video lectures even when transcript is not perfect. Unfortunately, the current ASR technology is still far from the necessary accuracy.

The imperfect transcripts resulting from ASR can be manually supervised and corrected, but the effort can be even higher than manual transcription. For the purpose of alleviating this issue, a novel *Interactive Transcription of Speech* (IST) system is presented in this thesis. This IST succeeded in reducing the effort if a small quantity of errors can be allowed; and also in improving the underlying ASR models in a cost-effective way.

In other to adequate the proposed framework into real-life MOOCs, another intelligent interaction methods involving limited user effort were investigated. And also, it was introduced a new method which benefit from the user interactions to improve automatically the unsupervised parts (the *Constrained Search* (CS)).

The conducted research was deployed into a web-based IST platform with which it was possible to produce a massive number of semi-supervised lectures from two different well-known repositories, videoLectures.net and poliMedia.

Finally, the performance of the IST and ASR systems can be easily increased by improving the computation of the *Confidence Measure* (CM) of transcribed words. As so, two contributions were developed: a new particular *Logistic Regresion* (LR) model; and the speaker adaption of the CM for cases in which it is possible, such with MOOCs.