

Resum del treball de tesi

“Confidence Measures for Automatic and Interactive Speech Transcription”

Mesures de Confiança per al Reconeixement Automàtic i Interactiu de la Parla

per *Isaias Sanchez Cortina*.

Supervisors: Alfons Juan i Ciscar i J. Alberto Sanchis Navarro.

Aquest treball hi contribueix al camp del *reconeixement automàtic de la parla* (RAP). I en especial, al de la *transcripció interactiva de la parla* i el de *mesures de confiança* (MC) per a RAP. Els objectius principals són els següents:

1. Dissenyar mètodes i eines per a TIP per tal de millorar les transcripcions automàtiques.
2. Avaluar els mètodes i eines TIP per a tasques de transcripció realistes extretes de grans repositoris de vídeos educacionals.
3. Millorar la fiabilitat del TIP, mitjançant la millora de les MC.

Resum

El *reconeixement automàtic de la parla* (RAP) és una tasca crucial per una àmplia gamma d'aplicacions importants que no es poden dur a terme per mitjà de la transcripció manual. El RAP pot proporcionar transcripcions en escenaris de creixent impacte social com el *Cursos online oberts massius* (MOOC). Les transcripcions permeten automatitzar tasques com ara cercar, resumir, recomanar, traduir; a més a més, fa accessibles els continguts als parlants no nadius i els usuaris amb discapacitat, etc. Fins i tot, pot millorar el rendiment acadèmic de estudiants que aprenen de xerrades amb subtítols, encara que aquests subtítols no siguin perfectes. Malauradament, la tecnologia RAP actual encara està lluny de la precisió necessària.

Les transcripcions imperfectes resultants de RAP poden ser supervisades i corregides manualment, però aquest l'esforç pot acabar sent superior a la transcripció manual. Per tal de resoldre aquest problema, en aquest treball es presenta un sistema nou per a *transcripció interactiva de la parla* (TIP). Aquest sistema TIP va ser reeixit en la reducció de l'esforç per quan es pot permetre una certa quantitat d'errors; així com també en en la millora dels models RAP subjacents.

Per tal d'adequar el marc proposat per a MOOCs, també es van investigar altres mètodes d'interacció intel·ligents amb esforç d'usuari limitat. A més a més, es va introduir un nou mètode que aprofita les interaccions per tal de millorar encara més les parts no supervisades (RAP amb *cerca restringida*).

La investigació en TIP duta a terme es va desplegar en una plataforma web amb la qual va ser possible produir un nombre massiu de transcripcions semi-supervisades de xerrades de repositoris ben coneguts, videoLectures.net i poliMedia.

Finalment, el rendiment de la TIP i els sistemes de RAP es pot augmentar directament mitjançant la millora de l'estimació de la *Confiança Mesura* (MC) de les paraules transcrites. Per tant, es van desenvolupar dues contribucions: un nou model discriminatiu *logístic* (LR); i l'adaptació al locutor de la MC per casos en que és possible, per exemple amb MOOCs.