



SEGMENTACIÓ DE LOCUTOR PER A L'INDEXACIÓ AUTOMÀTICA DE BASES DE DADES MULTIMÈDIA EN CATALÀ

**Xavier Anguera, Mireia Farrús,
Javier Hernando, Alberto Abad.**

Departament de Teoria del Senyal i Comunicacions (TSC)

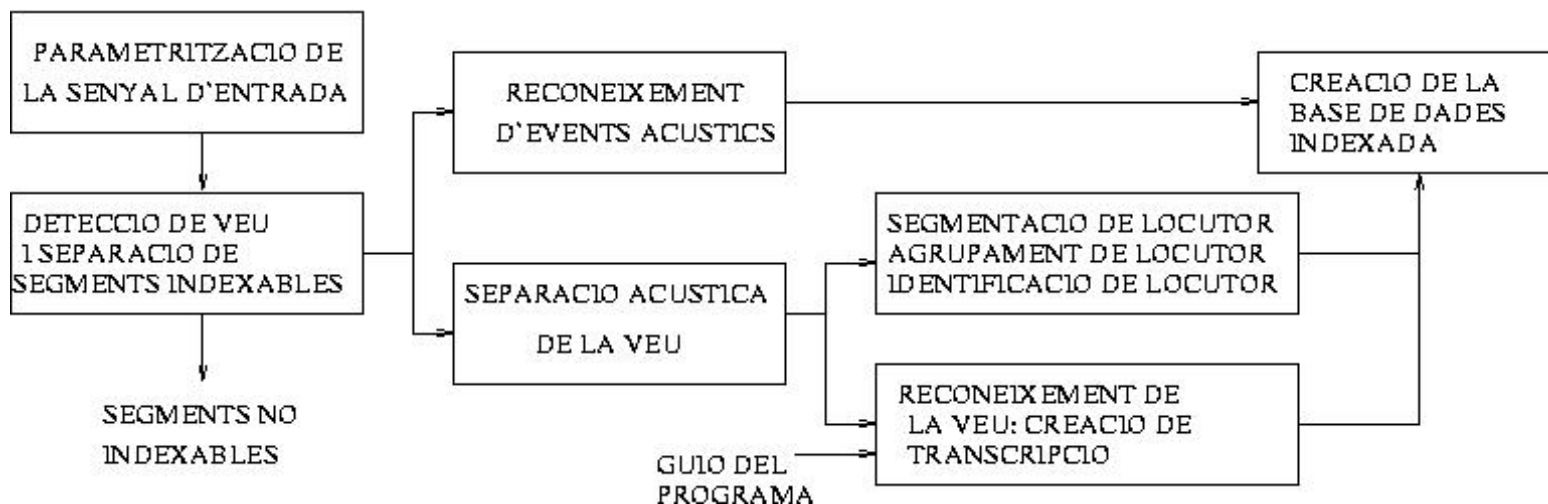
Centre de Recerca TALP

Universitat Politècnica de Catalunya (UPC)

Indexació automàtica de la senyal d'audio



- Gran quantitat de continguts audiovisuals es creen cada dia en català.
- Cal indexar aquesta informació per a un possible posterior ús.
- Les tècniques d'indexat manuals són costoses i poc eficients.
- L'indexació automàtica de la senyal de veu extrau característiques del senyal concatenant sistemes experts.





Segmentació de locutor

- Donat un segment d'audio, defineix on comença i acaba cada locutor.
- Útil per a: recuperar intervencions de locutors concrets, ajudar al reconeixement de veu, ...
- La mesura més frequentment usada es la Bayesian Information Criterion (BIC):

$$\Delta BIC = P(\Theta | \lambda) - P(\Theta_1 | \lambda_1) - P(\Theta_2 | \lambda_2) - \frac{1}{2} \Lambda K \log N$$

On

$$P(\Theta_i | \lambda_i) = \sum_{k=1}^N \log p(\theta_i(k) | \lambda_i)$$

Donats dos segments Θ_1 i Θ_2 , la mesura BIC compara si les dades pertanyen a dos locutors diferents o a un de sol.

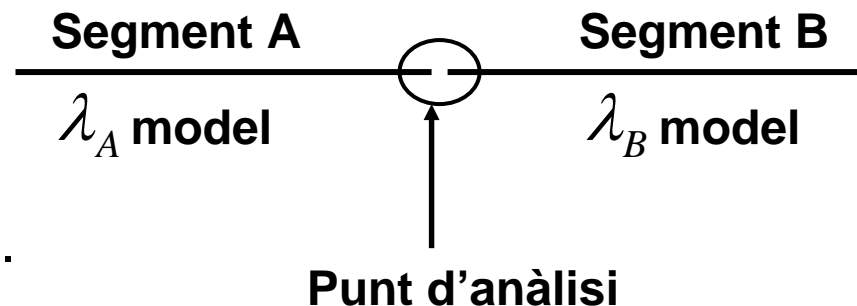
Mesura XBIC

- Presentem una mesura alternativa al BIC per a segmentació de locutor.

Donats dos segments A,B de longitud igual, creem dos models entrenats amb cadascun dels segments i comparem les probabilitats creuades.

$$XBIC = P(A | \lambda_B) + P(B | \lambda_A)$$

Per a un punt d'anàlisi, XBIC mesura la similitud de cadascun dels segments amb el model entrenat amb el segment contrari.



Experiments amb bases de dades en català



- Recollides i anotades a mà 2,5 hores de Telenotícies de TV3 i de 3/24.
- Mesures:
 - $PRC = \text{Tasa de canvis trobats correctament} / \text{nombre total de canvis trobats}$
 - $RCL = \text{Tasa de canvis trobats correctament} / \text{nombre total de canvis existents}$

<i>Telenotícies</i>	<i>Mesura RCL</i>	<i>Mesura PRC</i>
TV3 migdia 6-7-2004	50.00%	67.00%
TV3 migdia 7-7-2004	43.01%	63.40%
TV3 migdia 8-7-2004	53.04%	64.89%
3/24 TN matí 5-7-2004	46.47%	41.27%
Totals:	48.13%	59.14%



Conclusions i treball futur

- L'indexació automàtica de la veu aplicada als continguts audiovisuals en català presenta moltes aplicacions futures.
- Hem presentat una nova mesura per a realitzar la segmentació de locutor, anomenada XBIC.
- La mesura XBIC s'inspira en la mesura BIC i compara com de similars son dos segments adjacents.
- XBIC requereix menys computació que BIC i generalment redueix el error de segmentació de BIC.
- Proves fetes amb una base de dades recollides en català. Resultats preliminars mostren el potencial de la mesura presentada.