Crosslingual Adaptation of Semi-Continuous HMMs using Acoustic Sub-Simplex Projection

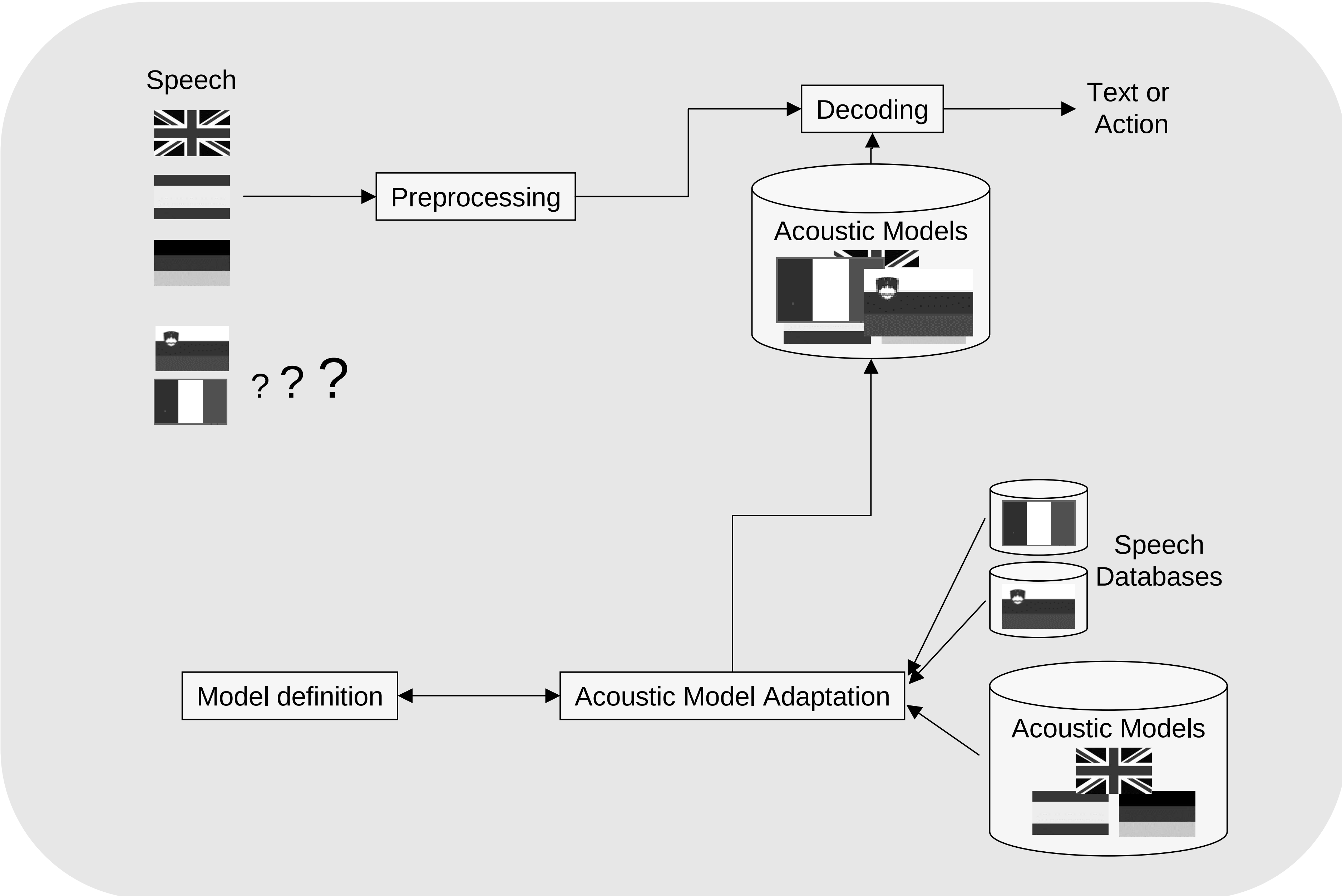
Frank Diehl, Asunción Moreno, Enric Monte

TALP Research Center
Universitat Politècnica de Catalunya

- Motivation
- The adaptation procedure
- Tests
- Conclusions

- **Motivation**
- The adaptation procedure
- Tests
- Conclusions

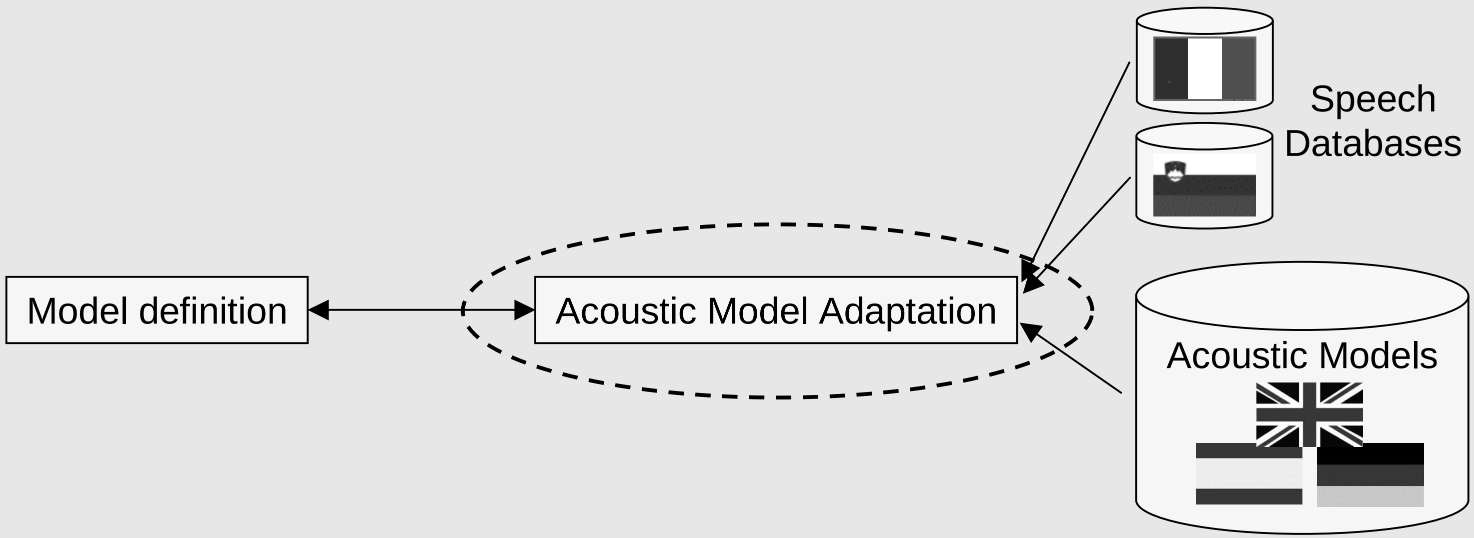
Motivation


Speech
? ? ?
Preprocessing
Decoding
Text or Action
Acoustic Models
Speech Databases
Model definition
Acoustic Model Adaptation
Acoustic Models

Motivation

CDHMM

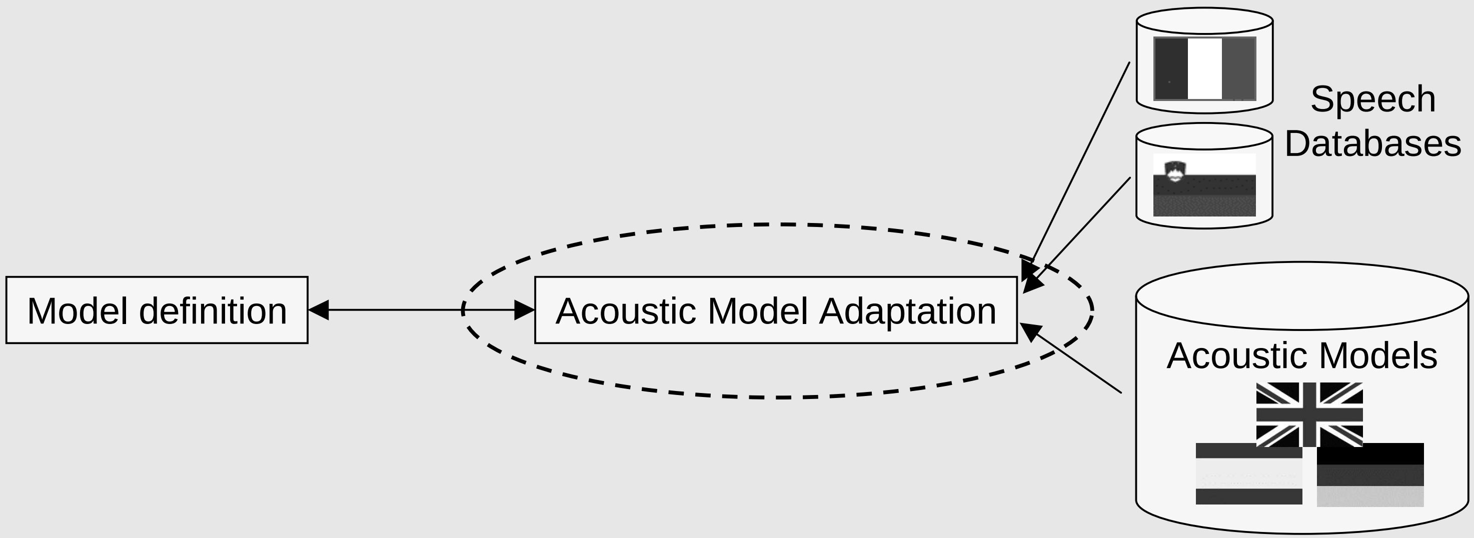
- MLLR / MAP
- Usually Gaussian mean adaptation
- MLLR favored for little adaptation data, regression classes
- MAP: more data needed, prior definition



Motivation

SCHMM

- One common codebook → no regression classes
  → MLLR makes little sense
  → Mean adaptation questionable
- MAP is possible but more data needed, priors needed.
- Prototype weights should be adapted.
  → Solution need to stay in the probabilistic simplex.
- Transformation based solution desired
  → little data necessary

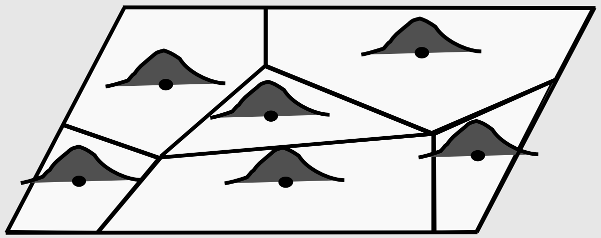


- Motivation
- **The adaptation procedure**
- Tests
- Conclusions

- Motivation
- The adaptation procedure
  - **The data model**
  - Maximum Likelihood Convex Regression (MLCR)
  - Model prediction and regression classes
  - Probabilistic Latent Semantic Analysis (PLSA)
  - Maximum a Posteriori Convex Regression (MAPCR)
- Tests
- Conclusions

The data model

Codebook with K prototypes



Output density

$$P(x\,|\,s)=\sum_{k=1}^{K} c_{sk}\,G_k(x,\cdot)$$

State index

$$\forall s\in\{1,\ldots,S\}$$

For each state s the weights vector $\underline{c}_s=\left[c_{s1},\ldots,c_{sK}\right]^T$ need to be adapted / estimate.

Convex combination of prototype weights vectors

$$\underline{c}_s=\left[\underline{u}_{s1},\underline{u}_{s2},\ldots,\underline{u}_{sL}\right]\begin{bmatrix}\alpha_{s1}\\ \alpha_{s2}\\ \cdot\\ \cdot\\ \alpha_{sL}\end{bmatrix}=\underline{U}_s\,\underline{\alpha}_s$$

$$K\times L,\quad K>>L$$

Prototype weights vectors

Combination weights

Acoustic regression classes + PLSA

ML/MAP estimation

- Motivation
- The adaptation procedure
  - The data model
  - **Maximum Likelihood Convex Regression (MLCR)**
  - Model prediction and regression classes
  - Probabilistic Latent Semantic Analysis (PLSA)
  - Maximum a Posteriori Convex Regression (MAPCR)
- Tests
- Conclusions

Maximum Likelihood Convex Regression (MLCR)

Starting point: Baum-Welsh auxiliary function for the weights $\overline{c}_{sk}$

$$Q(\lambda, \overline{c}_{sk}) = \sum_{k=1}^{K} \sum_{t=1}^{T} \gamma_{sk}(t) \log(\overline{c}_{sk})$$

$$\cdot \Big/ \sum_{t=1}^{T} \sum_{k=1}^{K} \gamma_{sk}(t)$$

$$\tilde{Q}(\lambda, \overline{c}_{sk}) = \sum_{k=1}^{K} \frac{\sum_{t=1}^{T} \gamma_{sk}(t)}{\sum_{t=1}^{T} \sum_{k=1}^{K} \gamma_{sk}(t)} \log(\overline{c}_{sk})$$

$$c_{sk} = \frac{\sum_{t=1}^{T} \gamma_{sk}(t)}{\sum_{t=1}^{T} \sum_{k=1}^{K} \gamma_{sk}(t)}$$

$$\tilde{Q}(c_{sk}, \overline{c}_{sk}) = \sum_{k=1}^{K} c_{sk} \log(\overline{c}_{sk})$$

$$\underline{\overline{c}}_{s} = \underline{U}_{s} \underline{\alpha}_{s}$$

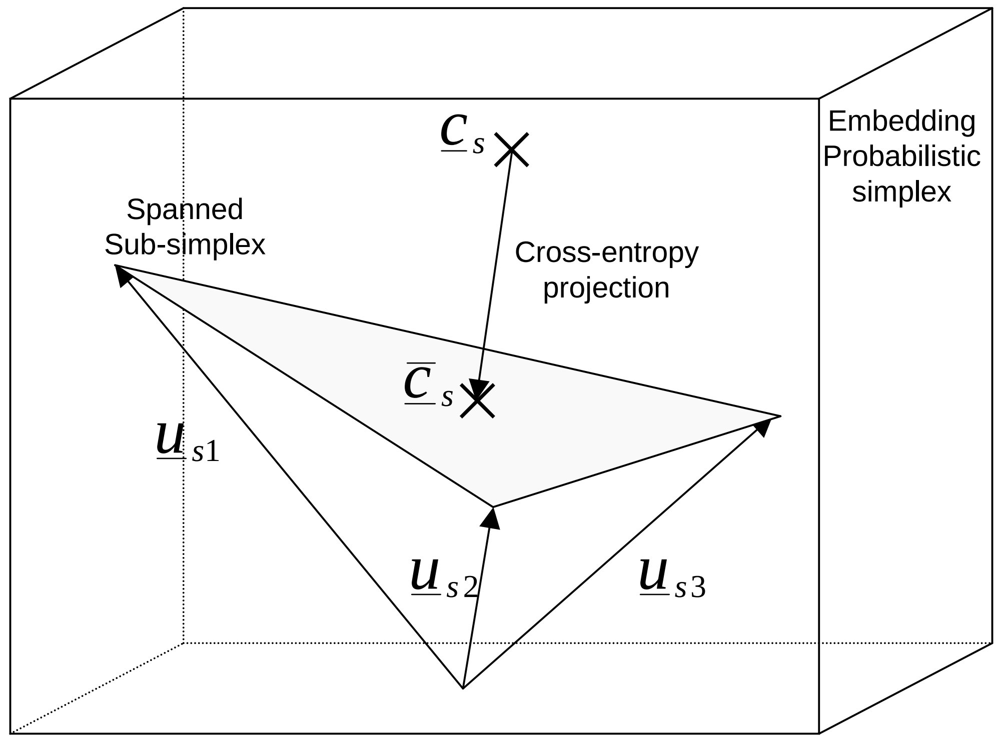
$$\tilde{Q}(c_{sk}, \underline{\alpha}_{s}) = \underline{c}_{s}^{T} \log(\underline{U}_{s} \underline{\alpha}_{s})$$

Maximum Likelihood Convex Regression (MLCR)

$$\arg\min_{\underline{\alpha}_s} -\underline{c}_s^T \log\left(\underline{U}_s \underline{\alpha}_s\right) \qquad \textit{subject to} \quad \sum_{l=1}^{L} \alpha_{sl} = 1$$

Measurement   Solution space

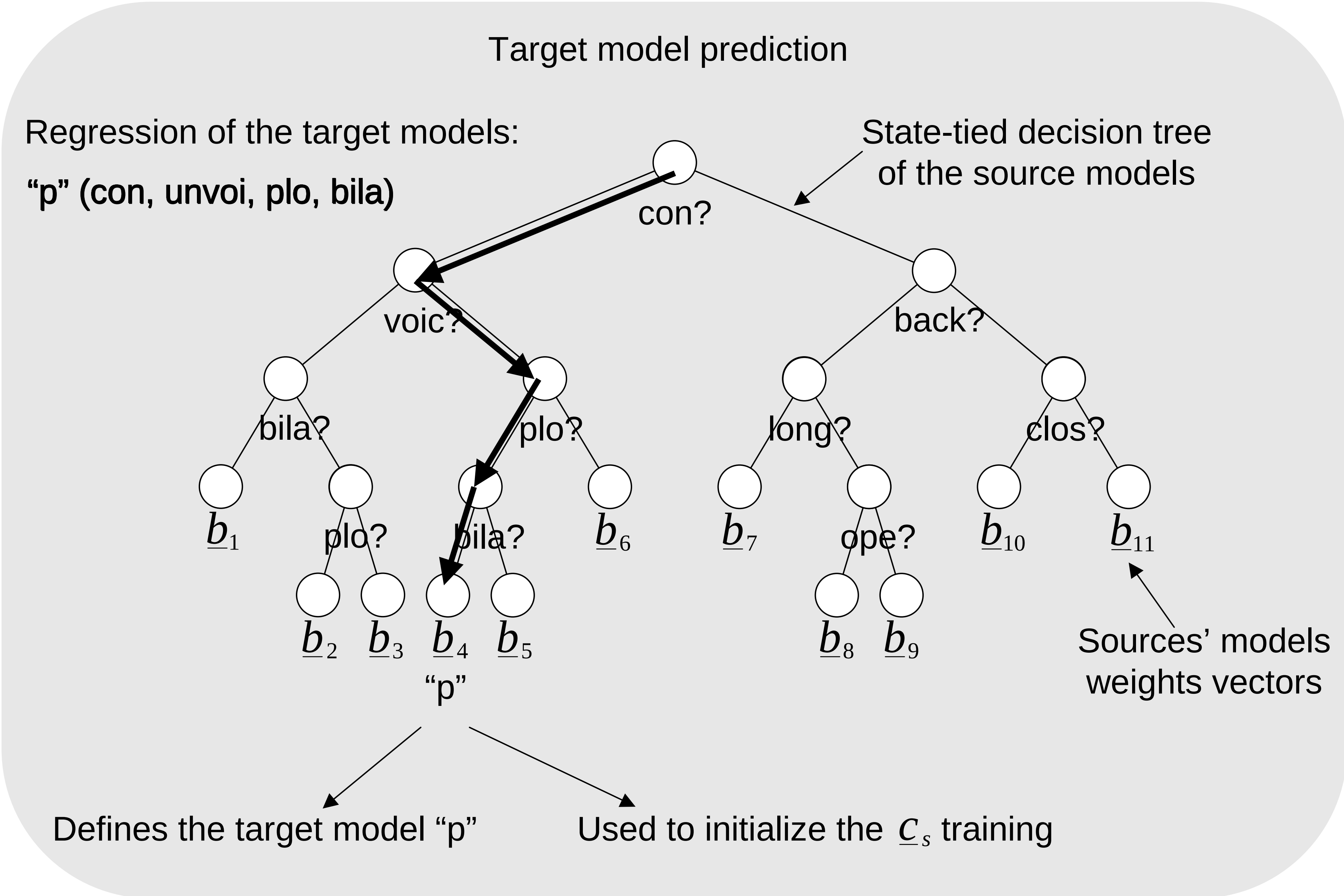
$$\textit{and} \quad \alpha_{sl} \geq 0 \quad \forall l \in \{1, \ldots, L\}$$



Solved by convex optimization

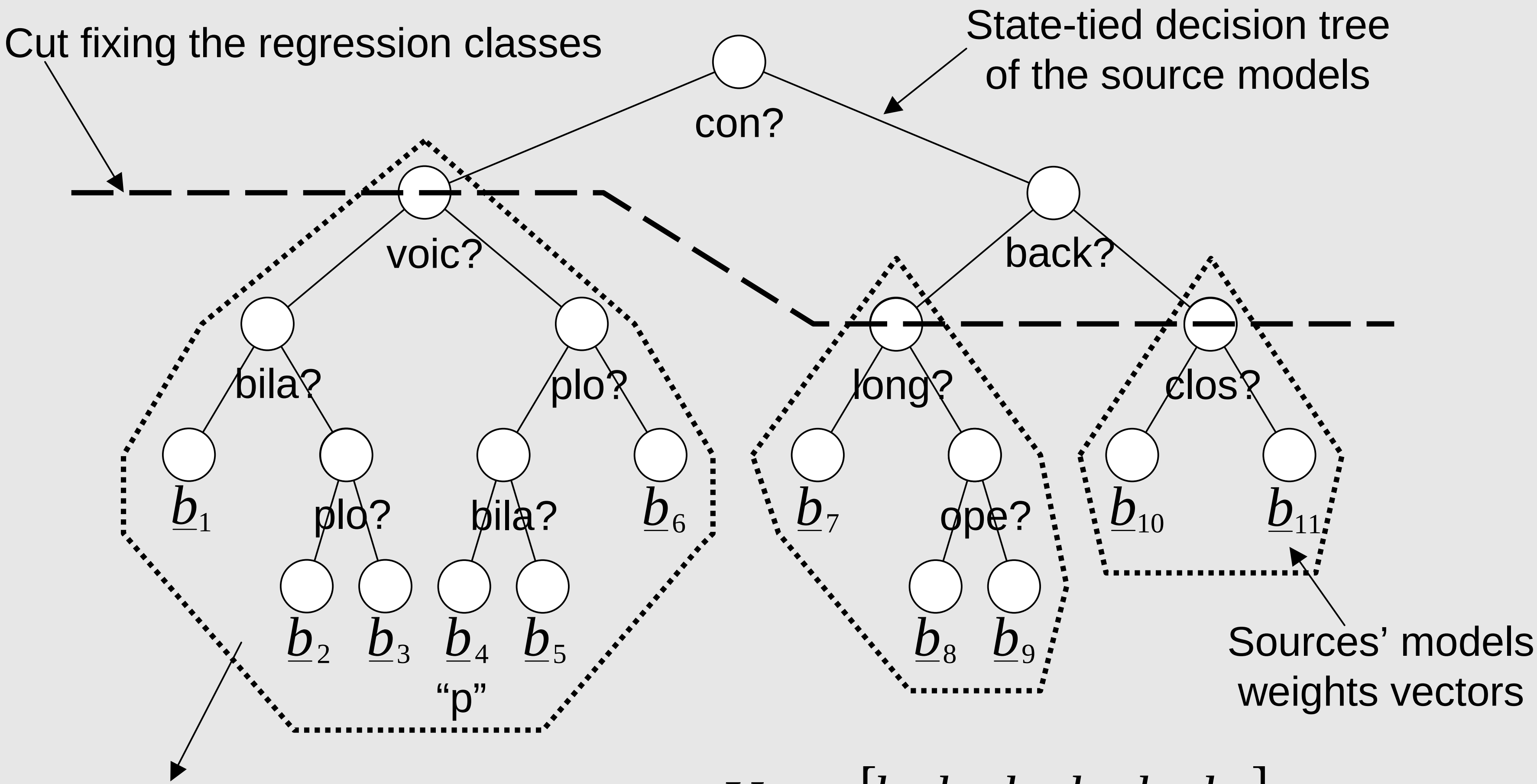
- Motivation
- The adaptation procedure
  - The data model
  - Maximum Likelihood Convex Regression (MLCR)
  - **Model prediction and regression classes**
  - Probabilistic Latent Semantic Analysis (PLSA)
  - Maximum a Posteriori Convex Regression (MAPCR)
- Tests
- Conclusions

Model prediction and regression classes



Model prediction and regression classes

Sub-simplex definition by acoustic regression classes



Acoustic neighbourhood of "p": $\underline{U}_{"p"} = [\underline{b}_1, \underline{b}_2, \underline{b}_3, \underline{b}_4, \underline{b}_5, \underline{b}_6]$

"A solution sub-simplex $\underline{U}_s$ is given by the $\underline{b}_i$ of a regression class"

- Motivation
- The adaptation procedure
  - The data model
  - Maximum Likelihood Convex Regression (MLCR)
  - Model prediction and regression classes
  - **Probabilistic Latent Semantic Analysis (PLSA)**
  - Maximum a Posteriori Convex Regression (MAPCR)
- Tests
- Conclusions

Probabilistic Latent Semantic Analysis (PLSA)

Problem: - The sub-simplex dimension depends on the regression class
- Statistical dependencies within a sub-simplex

Remedy: Probabilistic latent semantic analysis

Probabilistic model: conditional independence given a ‘latent’ variable

$$P(k,s\,|\,z) = P(k\,|\,z)P(s\,|\,z)$$

$$\sum_{\forall z} \cdot P(z)$$

$$P(k\,|\,s)P(s) = \sum_{\forall z} P(k\,|\,z)P(z)P(s\,|\,z)$$

Matrix notation

$$[\underline{b}_1, \underline{b}_2, \underline{b}_3, \ldots]\,\underline{P}_s = \underline{P}_{k|z}\,\underline{P}_z\,\underline{P}^T_{s|z}$$

Regression class

State probabilities

Sub-simplex $\underline{U}_s$

- SVD-like matrix decomposition
- Definition of sub-simplex bases
- Free eligible order
- Solved by the EM algorithm

- Motivation
- The adaptation procedure
  - The data model
  - Maximum Likelihood Convex Regression (MLCR)
  - Model prediction and regression classes
  - Probabilistic Latent Semantic Analysis (PLSA)
  - **Maximum a Posteriori Convex Regression (MAPCR)**
- Tests
- Conclusions

Maximum a Posteriori Convex Regression (MAPCR)

Problem: - MLCR enforces a solution to lie on the solution sub-simplex $\underline{U}_s$.
- But, in case of plenty of adaptation data, a solution near to the measurement $\underline{c}_s$ would make more sense.

Remedy: Probabilistic weighting of the solution between the solution sub-simplex $\underline{U}_s$ and the measurement $\underline{c}_s$ → MAP solution

Extension of the solution sub-simplex by $\underline{c}_s$:

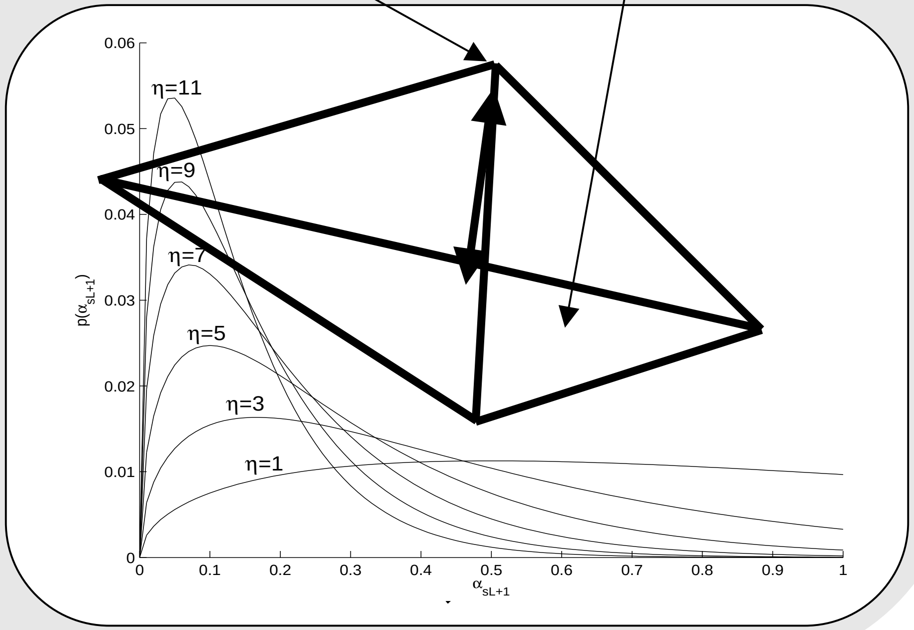
$$\overline{c}_s = \left[\underline{U}_s, \underline{c}_s\right]\left[\alpha_{s1},\ldots,\alpha_{sL},\alpha_{sL+1}\right]^T$$

Prior definition: *'ad hoc'*

- uniform for $\alpha_{s1},\ldots,\alpha_{sL}$
- gamma distribution for α_{sL+1}

$$p(\alpha_{sL+1}) = C\,\alpha_{sL+1}^{0.5}\exp(-\eta\,\alpha_{sL+1})$$

For all tests set to: $\eta = 7$



- Motivation
- The adaptation procedure
- **Tests**
- Conclusions

Tests

- System overview
  - SCHMM, mMFCC, Δ, $\Delta\Delta$, Δenergy
  - Gaussian mixtures with 256 / 32 prototypes
  - 3-state state-tied left-to-right demiphones
  - IPA-based phonetic questions

- Test setup
  - Multilingual Spanish-English-German source models
  - Training: 1000 speaker per language, phonetically rich sentences
  - Target languages: Slovenian, French (45/43 phonemes)
  - Adaptation material: 10/10 and 25/25 men/women, 170/425 phonetically rich sentences
  - Test setup: A list of phonetically rich words and application words, grammar size 372/445 (Slovenian/French)
  - Test material: Independent of the adaptation material, 50 men, 50 women, 614 and 670 sentences (Slovenian/French)

- All results are given in WER

Tests without PDTS

	Slovenian		French	
#Speaker	**20**	**50**	**20**	**50**
MONO	9.61		6.12	
PRED	50.49		45.37	
PRED-I1	26.71	20.68	27.91	22.84
MLLR	26.38	21.50	27.01	21.64
MLCR	32.41	32.08	31.19	31.79
MAPCR	20.03	18.89	22.84	19.40

- Conclusions
  - Model retraining is most effective
  - MLLR does not help
  - MLCR worses the situation
  - MAPCR improves the situation significantly
  - MAPCR is most effective for little adaptation data

Tests applying PDTS

- PDTS-5/10/15 → minimum model count in the newly generated leaves: 5/10/15

	Slovenian		French	
#Speaker	20	50	20	50
MAPCR	20.03	18.89	22.84	19.40
PDTS-5	32.57	26.06	21.19	14.03
PDTS-10	26.71	20.36	19.40	12.39
PDTS-15	25.57	19.22	19.25	11.94
MAPCR-PDTS-5	28.50	23.94	18.21	14.33
MAPCR-PDTS-10	23.13	19.71	16.12	11.79
MAPCR-PDTS-15	21.01	18.40	16.27	11.79

- Conclusions
  - French 20 speaker: performance boost due to PDTS and MAPCR
  - French 50 speaker: performance boost due to PDTS, MAPCR helps
  - Slovenian: Deterioration by PDST, MAPCR remedies the outcome somewhat
  - Robust measurements are favored over an improved context modeling

Tree size analysis (number of leaves)

	Slovenian		French	
#States	**1500/1017**		**1500/696**	
#Speaker	**20**	**50**	**20**	**50**
PDTS-5	1884	2672	1890	2516
PDTS-10	1468	2118	1516	2112
PDTS-15	1260	1828	1315	1834

- Slovenian seems to make better use of the initial not adapted tree than French (use of 1017 instead of 696 leaves out of 1500)
- The final tree sizes are comparable between Slovenian and French → PDTS generates more leaves for French

- Possible explanation of the bad Slovenian PDTS behavior
  - Predicting initial Slovenian models consumes useless-proven questions without improving the system performance
  - The wasted questions are missing during PDTS resulting in a badly adapted tree

- Motivation
- The adaptation procedure
- Tests
- **Conclusions**

Conclusions

- We have presented a novel adaptation scheme for the cross-lingual adaptation of SCHMM.
- The method is based on the projection of a measurement vector to an expected solution space (smoothing).
- The method makes use of prior information by incorporating acoustic regression classes derived form the decision tree of the source language/s.
- The method is proven to perform well in two cross-lingual test scenarios (reduction of WER of up to ca. 20%).
- Applying PDTS led to ambivalent results. Though substantial improvements are obtained for French, a performance degradation is observed for Slovenian.

Thank you for your attention