PHONOTACTIC LANGUAGE RECOGNITION USING I-VECTORS AND PHONEME POSTERIOGRAM COUNTS

Luis Fernando D'Haro,
Ondřej Glembek, Oldřich Plchot,
Pavel Matejka, Mehdi Soufifar,
Ricardo Cordoba, Jan Černocký

GTH

BRNO UNIVERSITY OF TECHNOLOGY
FIT
FACULTY OF INFORMATION TECHNOLOGY

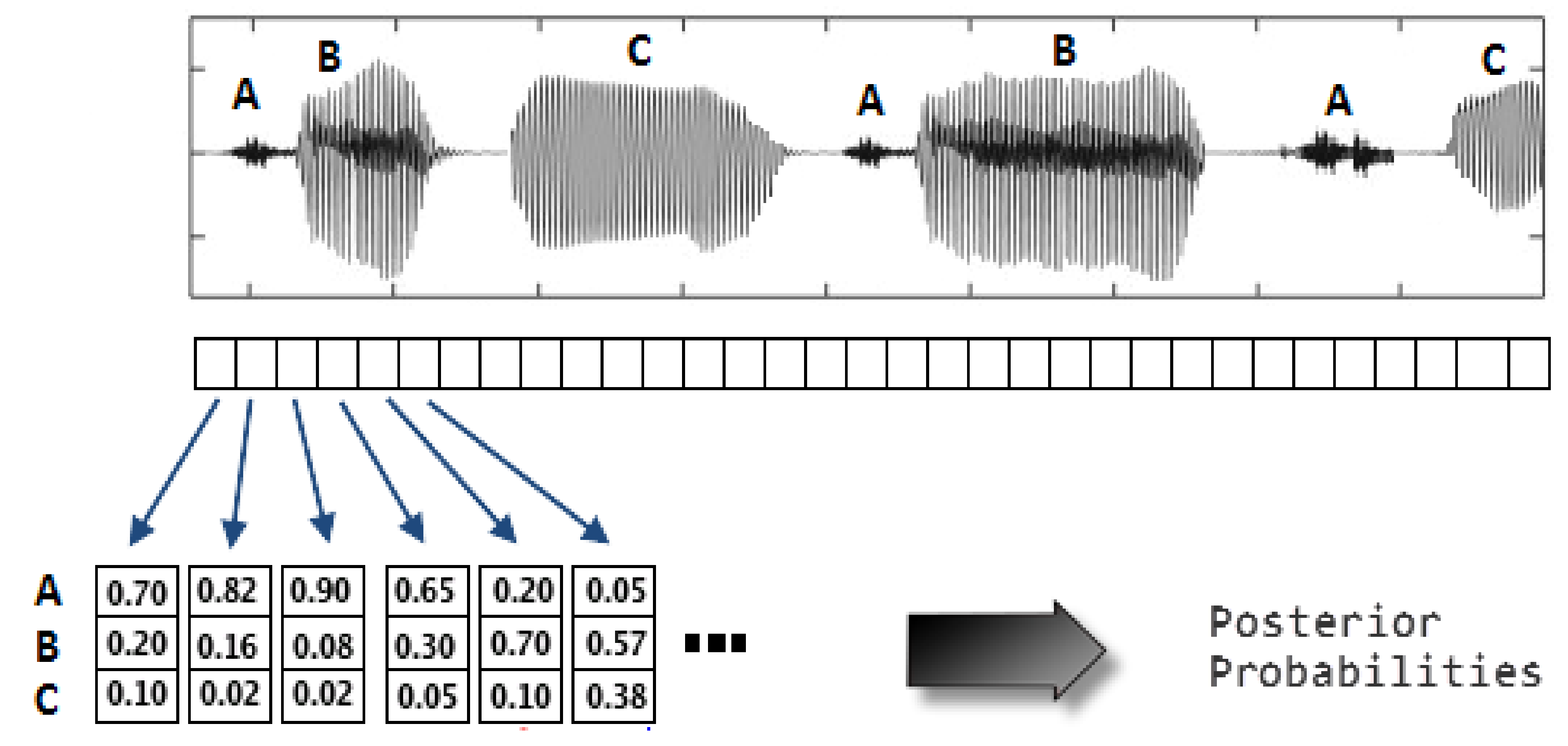
Content

- Introduction
- Steps to create Joint-Posteriogram n-gram counts
- Subspace Multinomial Models
  - i-vectors
- Results on LRE-2009
- Conclusions and Future work

Introduction

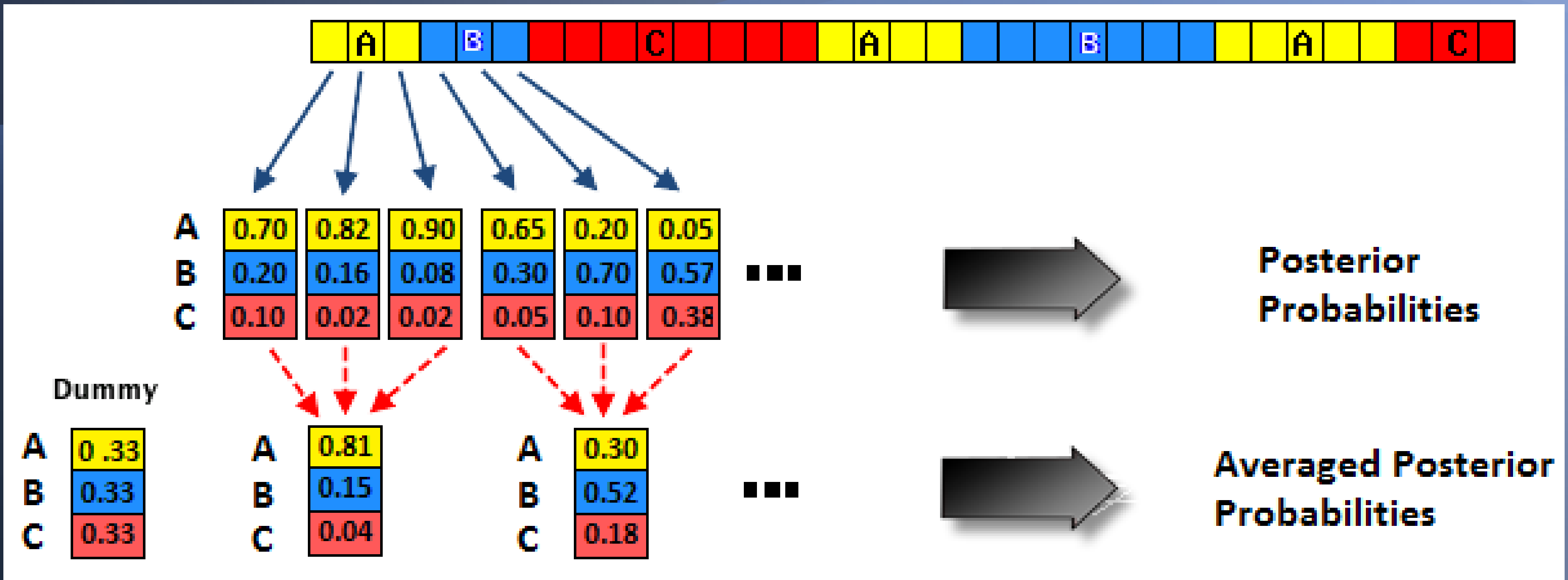
- Current main approaches to LID
  - Acoustic-based: i-vectors, JFA, SVMs, or GMMs
  - Phonotactic
- Phonotactic systems:
  - PRLM, PPRLM: LMs created using different phonetic ASR
  - Lattice-based soft-counts: Created from phone lattices
    - Zero counts (i.e. data sparseness)
    - Limited by the number of phonemes and n-gram order
    - Dimensionality reduction: PCA or n-gram selection
  - Using i-vectors through Subspace Multinomial Models (SMMs)
    - We propose a new feature vector that performs better than soft-counts

1. Compute Posterior Probabilities



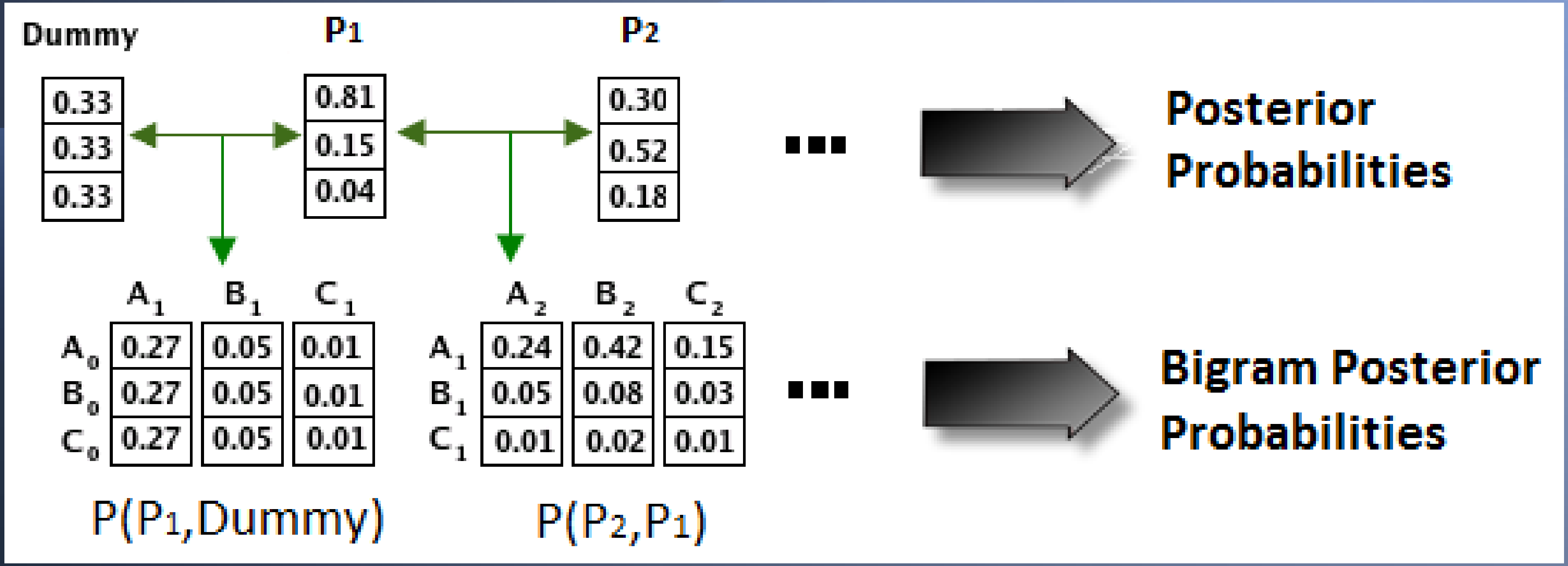
- In the example, we consider only three phonemes and bigrams. In our experiments, they were 33 and we used trigrams.

2. Compute Posterior Probs



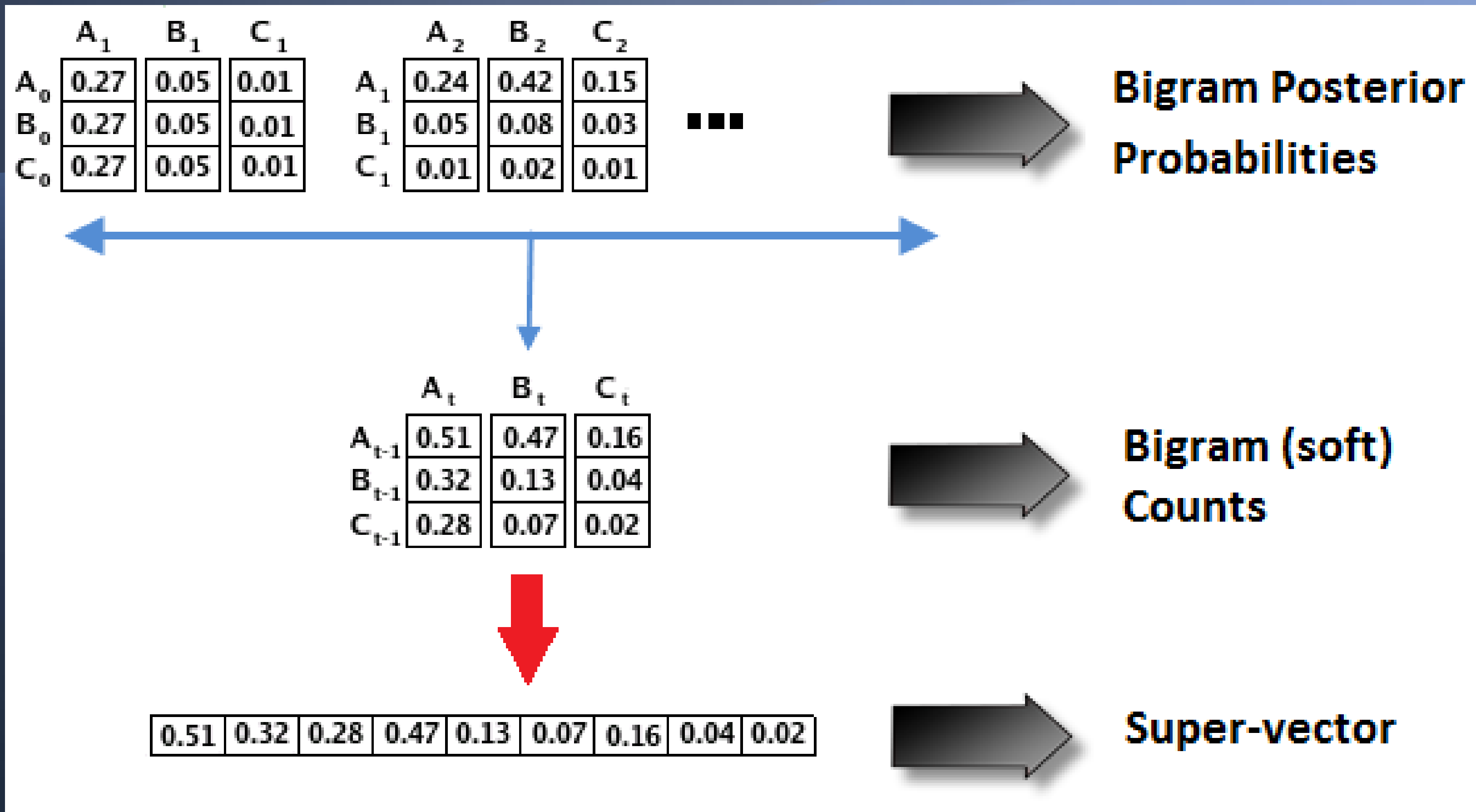
- Find the phoneme boundaries using Viterbi algorithm
  - Can be seen as incorporation of a-priori information
- Average the posterior probs over the phone boundaries
  - Smoothes the posterior probs and avoids the high-correlation of within-phoneme posteriors

3. Create N-gram Posterior Probs



- Outer product with the posteriogram of the previous phones
- Assume that the frames of the averaged posteriogram are statistically independent,
  - Therefore we have joint probabilities for sequences of phonemes

4. N-gram Counts via N-gram Posterior Probs



- Sum up all matrices to obtain n-gram soft counts
- Obtain feature super-vector for creating next the phonotactic i-vectors using SMMs

Subspace Multinomial Models

- Allows extraction of low-dimensional vectors of coordinates in total variability subspace (i.e. i-vectors)
- The log-likelihood of data *D* for a multinomial model with *C* discrete events is determined by

$$\log p(D) = \sum_{n=1}^{N} \sum_{c=1}^{C} \gamma_{nc} \log \varphi_{nc}$$

- Where γ_{nc} is the count for n-gram event *c* at utterance *n*, and φ_{nc} is the probability of a multinomial distribution defined by the subspace model

i-vectors from SMM

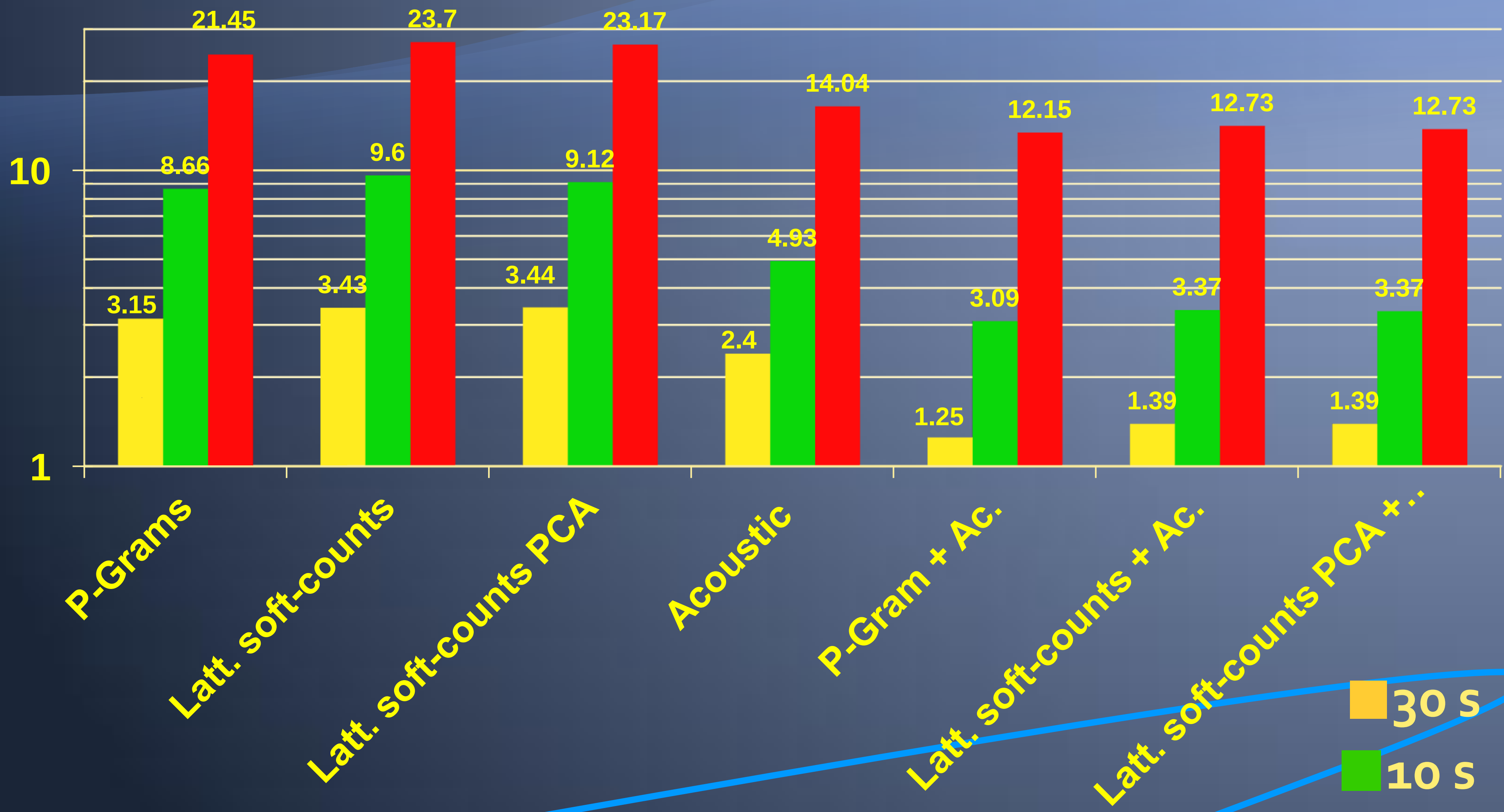
$$\varphi_{nc} = \frac{\exp\left(m_c + t_c w_n\right)}{\sum_i^C \exp\left(m_i + t_i w_n\right)}$$

- Where t_c is the c-th row of subspace matrix T (Extractor), and w_n is the i-vector
- An i-vector for a single utterance is estimated numerically by maximizing the likelihood (ML)
- Matrix T is trained numerically using ML by iteratively optimizing T and re-estimating the i-vectors for all training utterances
- Then we use these i-vectors as feature input for training a discriminative LID classifier
  - Multiclass logistic regression

Experimental Setup

- NIST LRE 2009 database
  - 50K segments for training (~119h), 38K segments for dev (~153h) and 31K sentences for test (~125h)
  - 23 languages, test on 3, 10, and 30 s conditions
  - Results given using C_{avg} metric
- Acoustic i-vector system
  - 7 MFCC + 49 SDCs, CMN/CVN, 2048 Gaussians -> i-vectors of 400 dimensions
- Comparisons with:
  - Lattice-based soft-counts with i-vectors (size 600)
  - Lattice-based soft-counts with PCA (reduction to 1000 dimensions)
- Fusion: Multiclass logistic regression
  - Acoustic and Phonotactic

Results on NIST LRE 2009



- Vertical-axes in logarithm

Conclusions and Future Work

- Advantages of the new features
  - Avoid data sparseness (i.e. robustness)
  - Results outperforms a similar system based on lattice soft-counts with i-vectors
    - 8,16% relative on 30 s condition
  - Fusion with acoustic i-vectors are also better
    - 10% relative on 30 s condition

- Future Work: Apply discriminative n-gram selection techniques to reduce the vector size
  - Avoids low frequency n-gram counts
  - Allows using high n-gram orders

…Thanks for your attention…

Comments or Questions?