Brno University of Technology

Faculty of Information Technology

Speech@FIT

Jiří Kopecký, Igor Szöke, Michal Fapšo, Martin Karafiát, Lukáš Burget, Ilya Oparin, František Grézl, Petr Schwarz, Ondřej Glembek, Pavel Matějka, Honza Černocký

szoke@fit.vutbr.cz

http://www.fit.vutbr.cz/speech

NIST STD 2006 workshop, December 13.-14. 2006, Gaithersburg

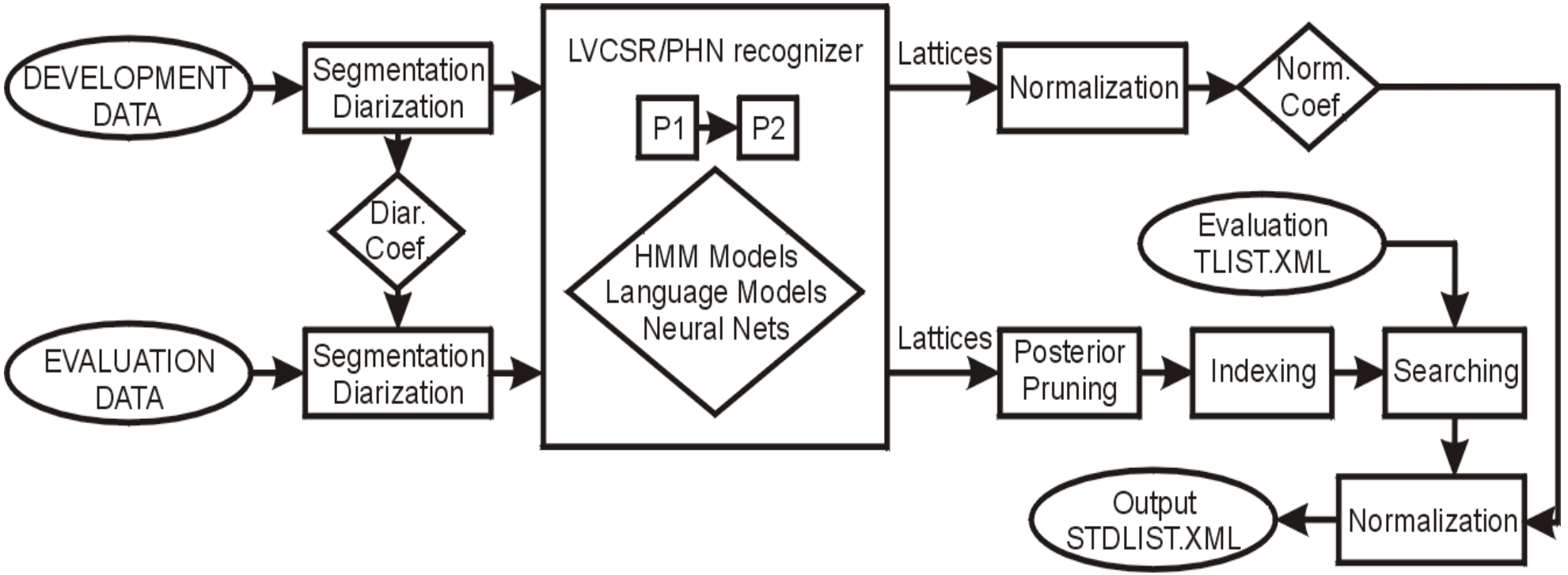
Outlines

This presentation contains only differences which were done to adapt a system on Arabic language. For full detail see the English system presentation or Arabic system description.

- **System overview**
- **LVCSR/Phoneme recognizer**
- **Indexing and searching**
- **Results and discussion**

Arabic: Broadcast News, Conversational Telephone Speech

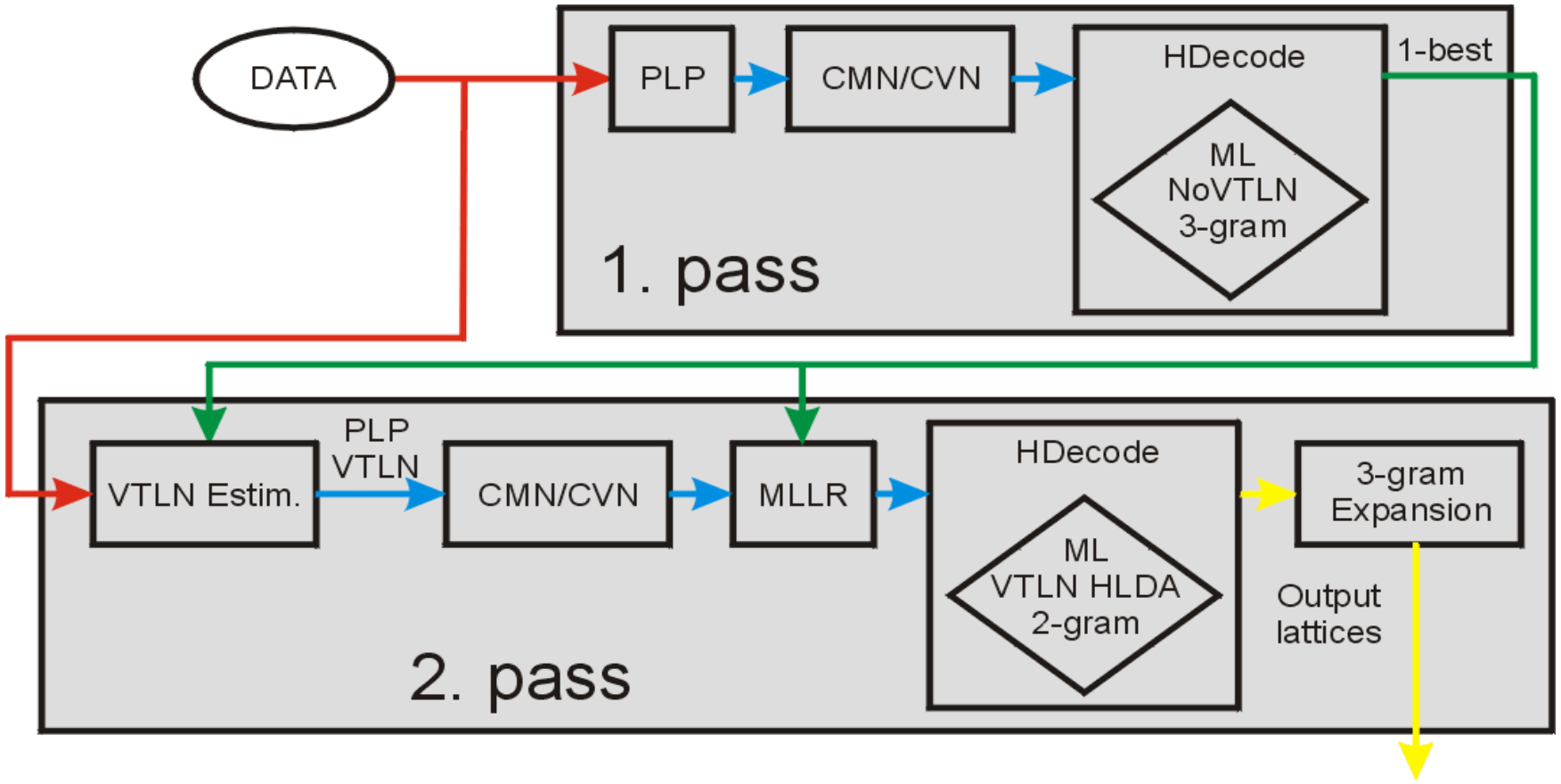
Spoken Term Detection System



Segmentation

- **Speech/nonspeech detection was done using LC/RC long temporal trajectory phoneme recognizer [1,2]**
- **Segments were separated by using silences longer than 0.5s.**
- **Segmentation for CTS was done using comparison of short time energy in both channels. Segment is labeled as silence if:**
  - **the average energy in 'speech' segment is 30 dB less than the maximum energy from the utterance**
  - **the energy in the other channel is bigger than maximum energy minus 3dB in the processed channel**
- **Diarization for BCN done by David van Leeuwen and Matěj Konečný at TNO. Diarization coefficients were used the same as for English BCN.**

Description of The LVCSR



Description of The LVCSR

- **LVCSR for English task was adapted for Arabic task.**
- **System uses 2-pass decoding:**
  1. **pass: PLP, CMN/CVN, ML models, 3-gram decoding, 1-best output**
  2. **pass: PLP, VTLN, CMN/CVN, HLDA, ML models, MLLR speaker adaptation, 2-gram decoding, expansion to 3-gram, lattices output**
- **Posterior pruning was applied on final lattices.**

For details see: System description and AMI LVCSR paper [Hain06]

LVCSR Training Data

- **CTS: acoustic models were trained on Levantine Arabic QT Training Data Set 5 - 55.6h, LM whole database 1.4Mw**
- **BCN: downsampled and used CTS acoustic models, LM TDT4 Multilingual Text and Annotations interpolated with CTS LM. 10Mw**
- **Only nodiacritized word labels were used in the whole system.**
- **Pronunciation dictionary: utf8 words -> cp1256 (BA expected this input) -> BA -> our_transliteration -> graphemes -> (g2p) phonemes (we got 45k of 70k words)**
- **Only one pronunciation was used for LVCSR**
- **Words from STD Devel. set were added to dictionary and LM.**

LVCSR WER and Oracle for STD Development set

	WER
BCN	45.67%
CTS	65.68%

Description of Phoneme System

- **Phoneme lattices were generated from P2 pass features and acoustic models. No phoneme LM was used.**
- **Posterior pruning was applied on final lattices**

Indexing and Searching

The same as for English, see English presentation please.

Normalization

The same as for English, see English presentation please.

Results

	EVAL ATWV Merged Dia	EVAL MTWV Merged Dia	EVAL ATWV Merged NoDia	EVAL MTWV Merged NoDia	DEVEL MTWV Merged Dia	DEVEL MTWV Merged NoDia
BCN	-0.0924	0.0661	0.0437	0.1098	0.179	0.384
CTS	0.0030	0.0342	0.0006	0.0285	0.010	0.047

Credit Outside BUT

- **Thomas Hain (Sheffield) for having coordinated the AMI LVCSR.**
- **David van Leeuwen and Matěj Konečný (TNO) for diarization.**
- **Funding agencies:**
  - **EC**
  - **Czech Ministry of Defence**
  - **CESNET (for the HW to burn)**

References

[Schwarz06] Schwarz P., Matejka P. and Cernocky J.: Hierarchical Structures of Neural Networks for Phoneme Recognition, In Proceedings of ICASSP 2006, May 2006, Toulouse, France

[Matejka06] Matejka P., Burget L., Schwarz P. and Cernocky J., Brno University of Technology System for NIST 2005 Language Recognition Evaluation. Odyssey: The Speaker and Language Recognition Workshop, San Juan, Puerto Rico,Jun 2006

[Hain06] Thomas Hain et al., The AMI Meeting Transcription System: Progress and Preformance, NIST RT06 evaluations, 2006

Thank You for Your attention.