Brno University of Technology Demos@AMI-TT.Event

Jan "Honza" Černocký
cernocky@fit.vutbr.cz

http://www.fit.vutbr.cz/speech,
http://www.fit.vutbr.cz/research/groups/graph

BRNO
UNIVERSITY
OF TECHNOLOGY

FIT

FACULTY
OF INFORMATION
TECHNOLOGY

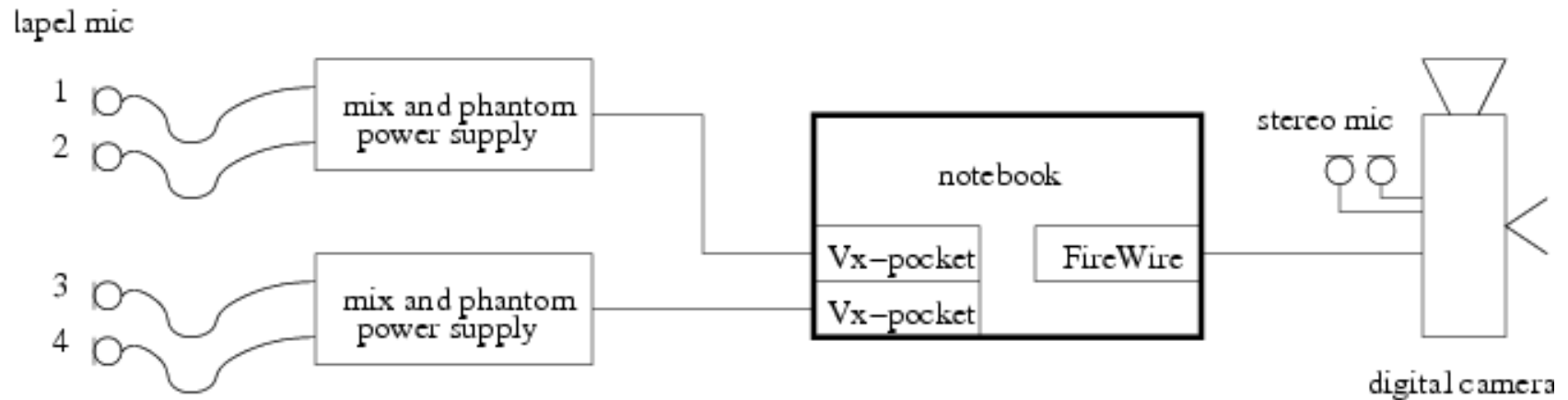AMI Technology Transfer Event, Brussels, March 7-8 2005

What is BUT ?

- Brno University of Technology – 2nd largest technical university in Czech republic (~2500 staff, ~10000 students).

- We are at the Faculty of Information Technology – FIT.

- Groups of speech and video processing.

- Partners in AMI.

You'll see …

1. Mobile meeting room with on-line speaker segmentation and keyword-spotting
2. Off-line keyword-spotting on the output of LVCSR recognition lattice.

1. Mobile meeting room

lapel mic

mix and phantom power supply

mix and phantom power supply

notebook

Vx–pocket

Vx–pocket
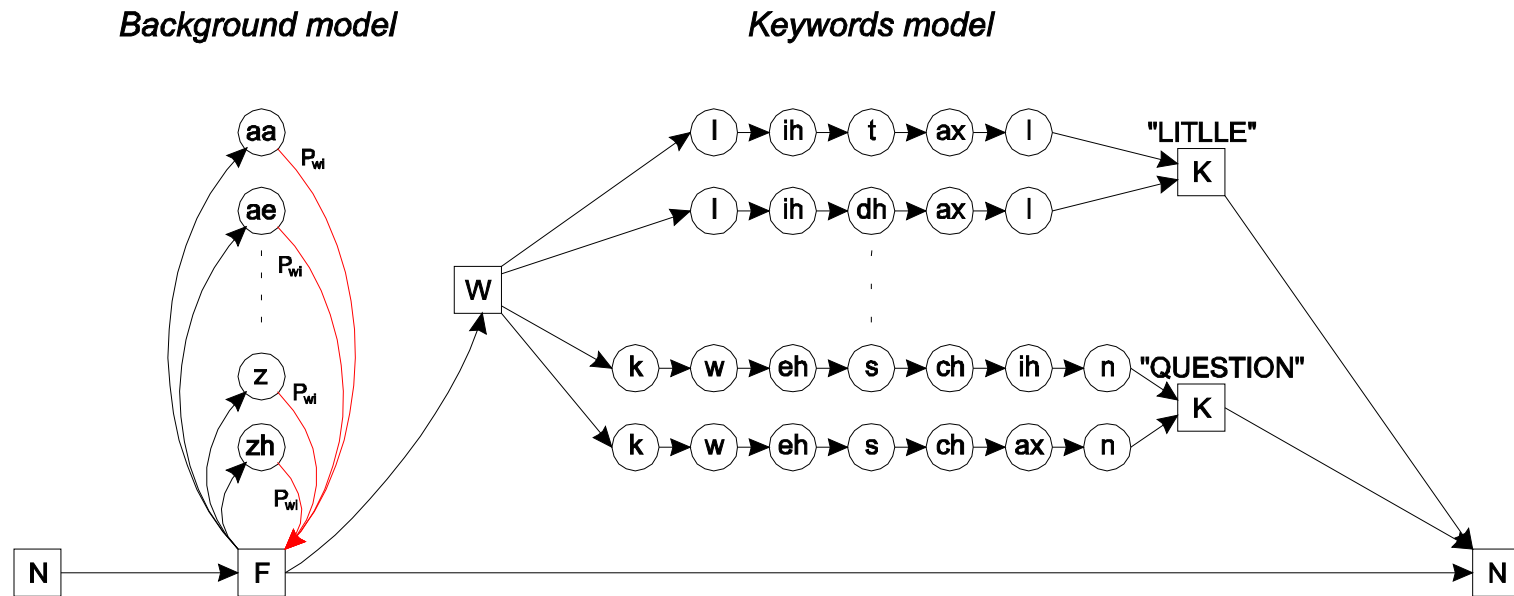
FireWire

stereo mic

digital camera

- recordings of up to 4 persons
- 4 lapel mikes
- one camera with hyperbolic mirror allowing for 360 degree view.
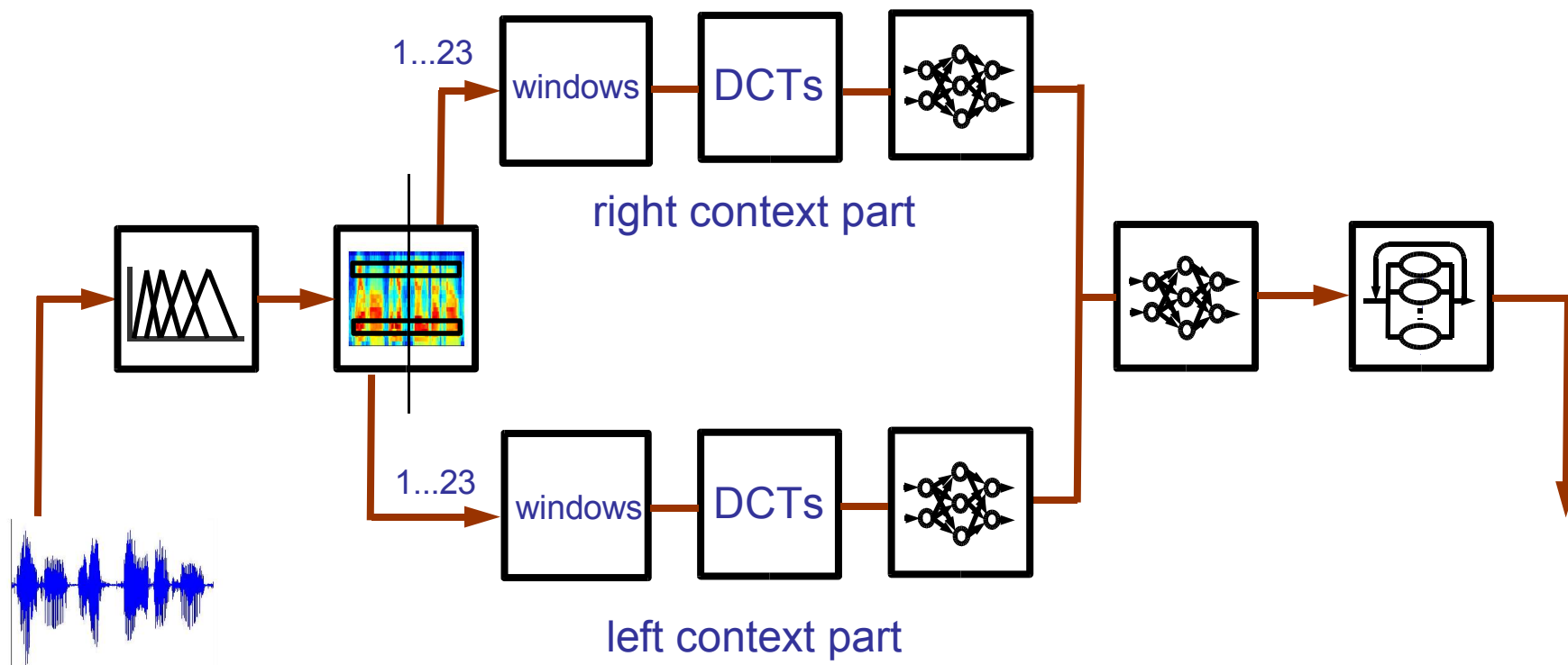- audio/video recording on one notebook.

Video processing
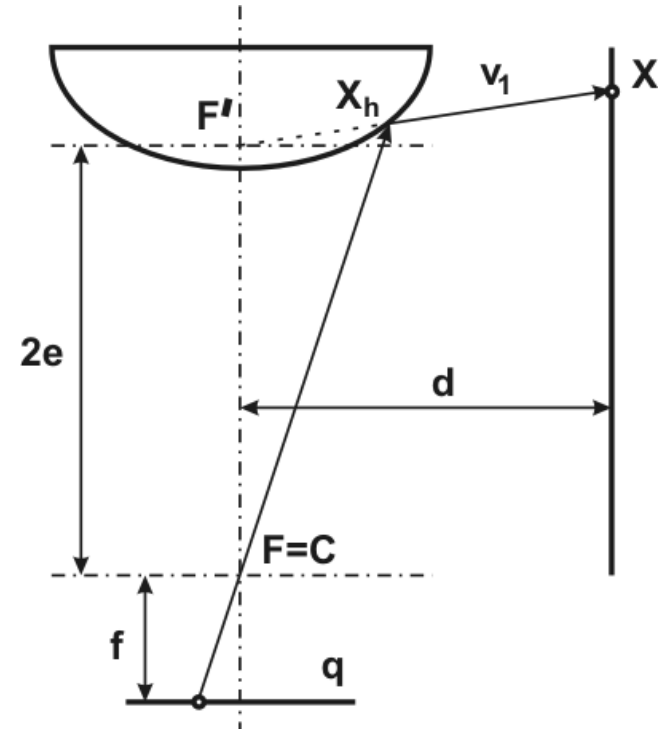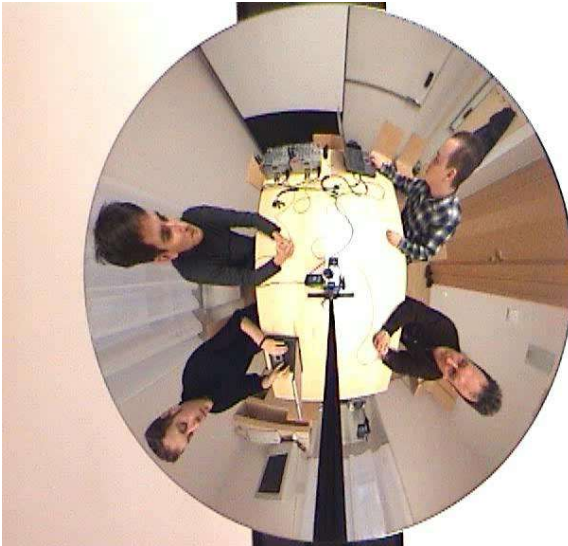
Audio processing

- Cross-correlation and energy based speaker-turn detector
- Keyword spotting

Background model

Keywords model

Acoustic-modeling for KWS

1...23 windows — DCTs —

right context part

1...23 windows — DCTs —

left context part

Implementation

Notebook 1:

- Audio and video grabbing (Direct-X), video transforms and selection of views.
- Sends data to audio server (sockets).

Notebook 2:

- Performs speaker segmentation and KWS
- Sends data back to NB1 for visualization

The output:

- Active speaker
- Keyword occurred

What for ?

On-line processing of meetings

- "Tele-participating at a meeting but interested only when speaker X speaks about topic Y".

- The system can alert the user when this occurs.

2. KWS in LVCSR lattices

- Output of LVCSR often not precise for a simple text search.
- LVCSR can produce more than just string of words – word lattices, but …
  - They're BIG (~100 MByte for an half-an-hour meeting).
  - They may contain many false alarms of the keyword searched
  - They must be searched efficiently.

Search of lattice, sort according to confidence, indexing

LVCSR system
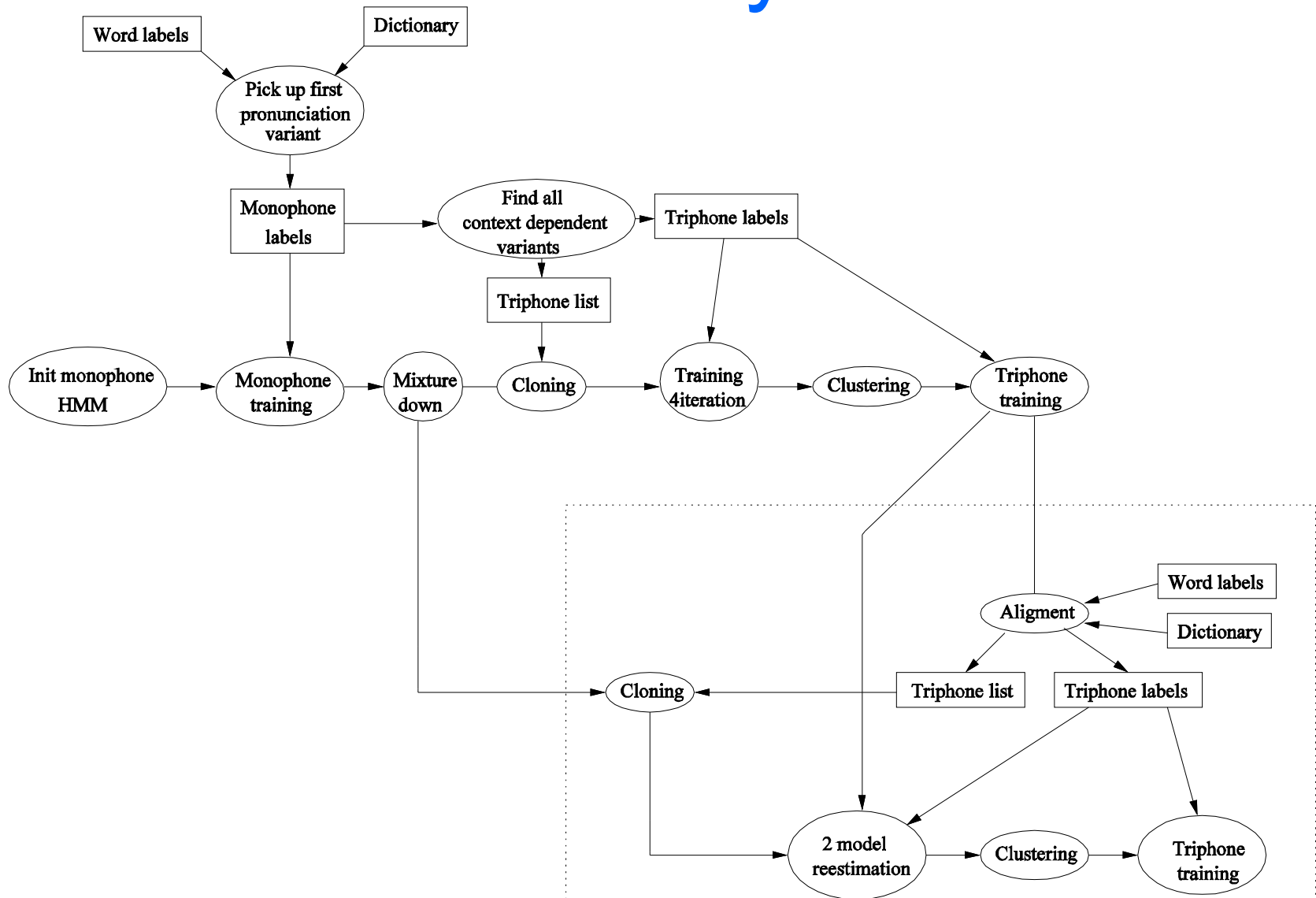
The search

- Pre-generated index-files of lattices – a fast "grep".

- Pre-computed scores of words (Baum/Welch).

- Sorting candidates according to confidence

- Going into the lattice to search for the context of found keywords (may be different from LVCSR "hard" output).

Output …

What for ?

- Better orientation in stored databases of meetings "Get fast what you need".

- Part of a complex info-retrieval system (speaker info, slides, video, …)

- Working on rare and unseen words – search in phoneme lattices + acoustic KWS.