

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>



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

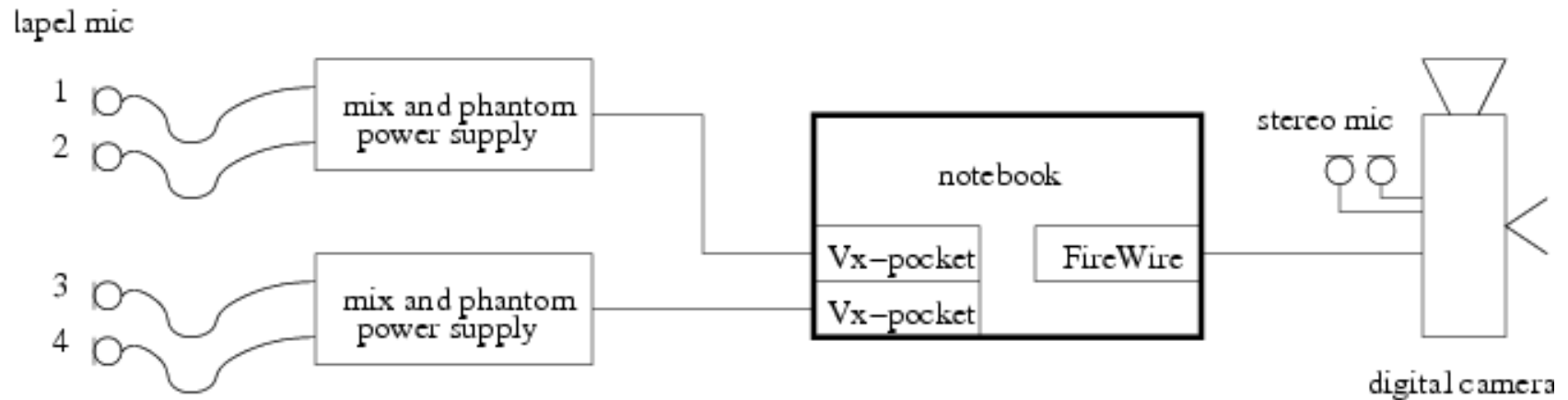
What is BUT ?

- 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 ...

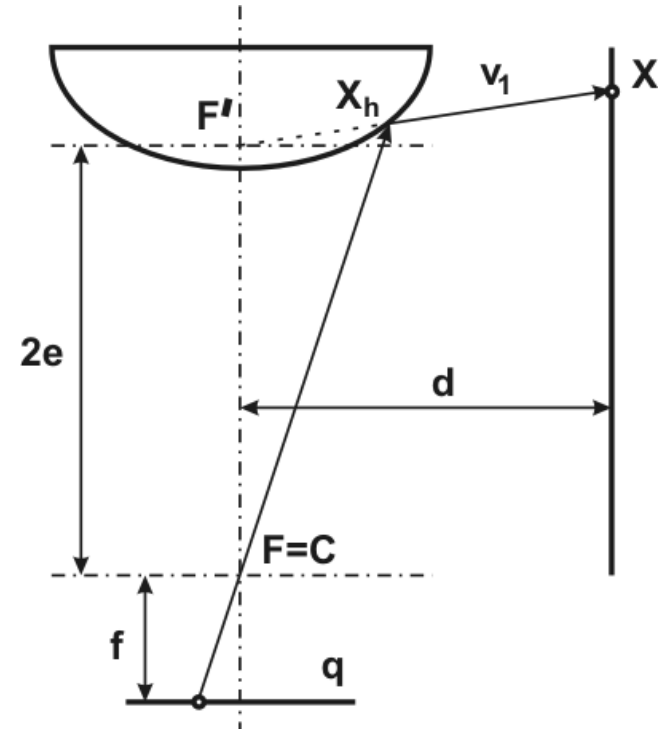
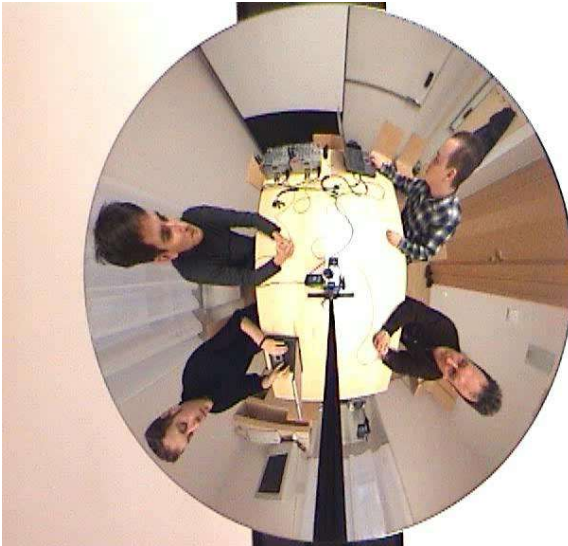
1. Mobile meeting room with on-line speaker segmentation and keyword-spotting
2. Off-line keyword-spotting on the output of LVCSR recognition lattice.

1. Mobile meeting room



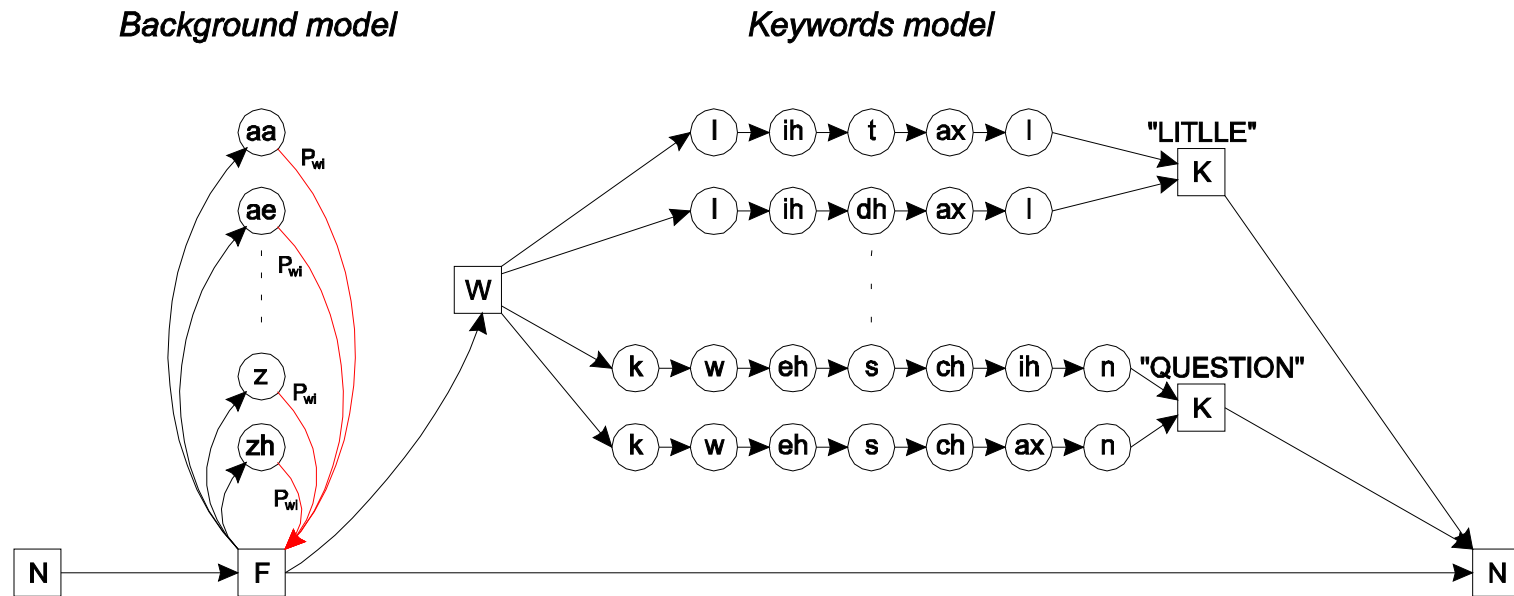
- recordings of up to 4 persons
- 4 lapel mikes
- one camera with hyperbolic mirror allowing for 360 degree view.
- audio/video recording on one notebook.

Video processing

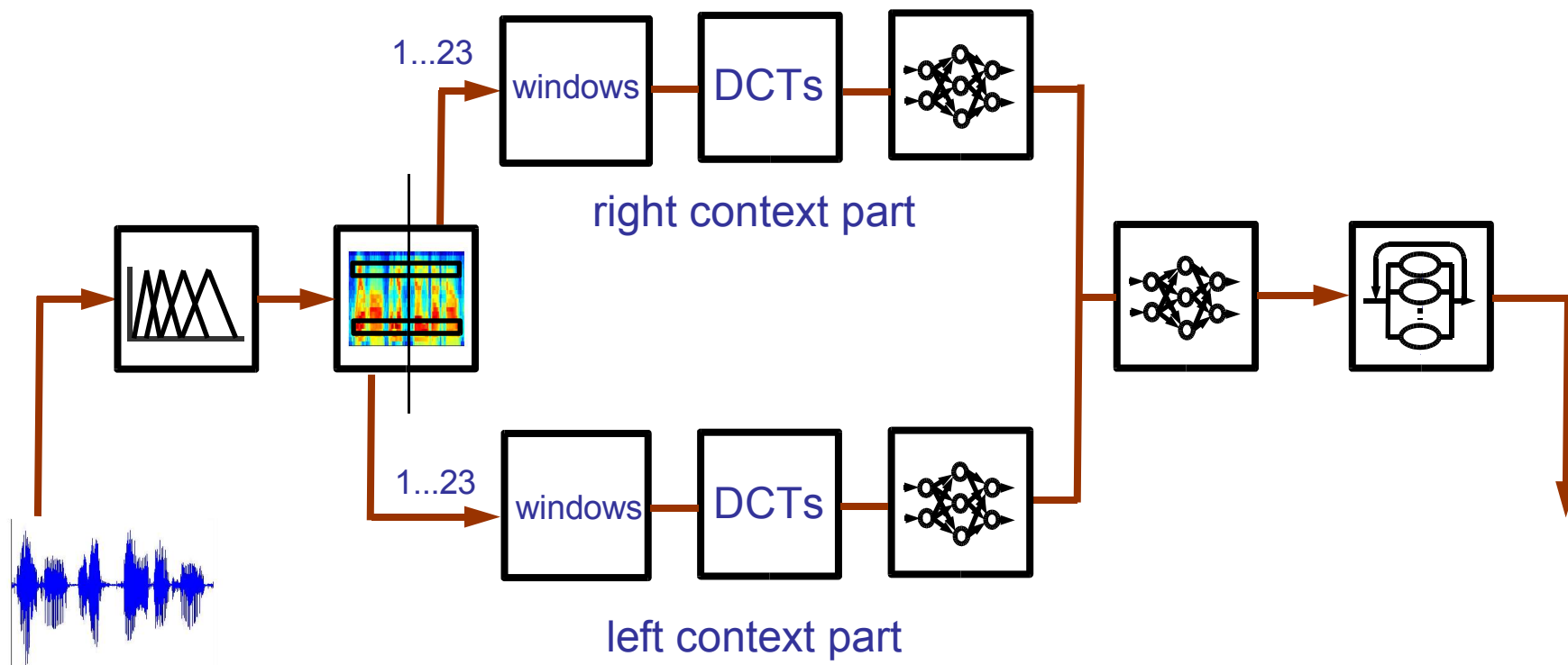


Audio processing

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



Acoustic-modeling for KWS



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.

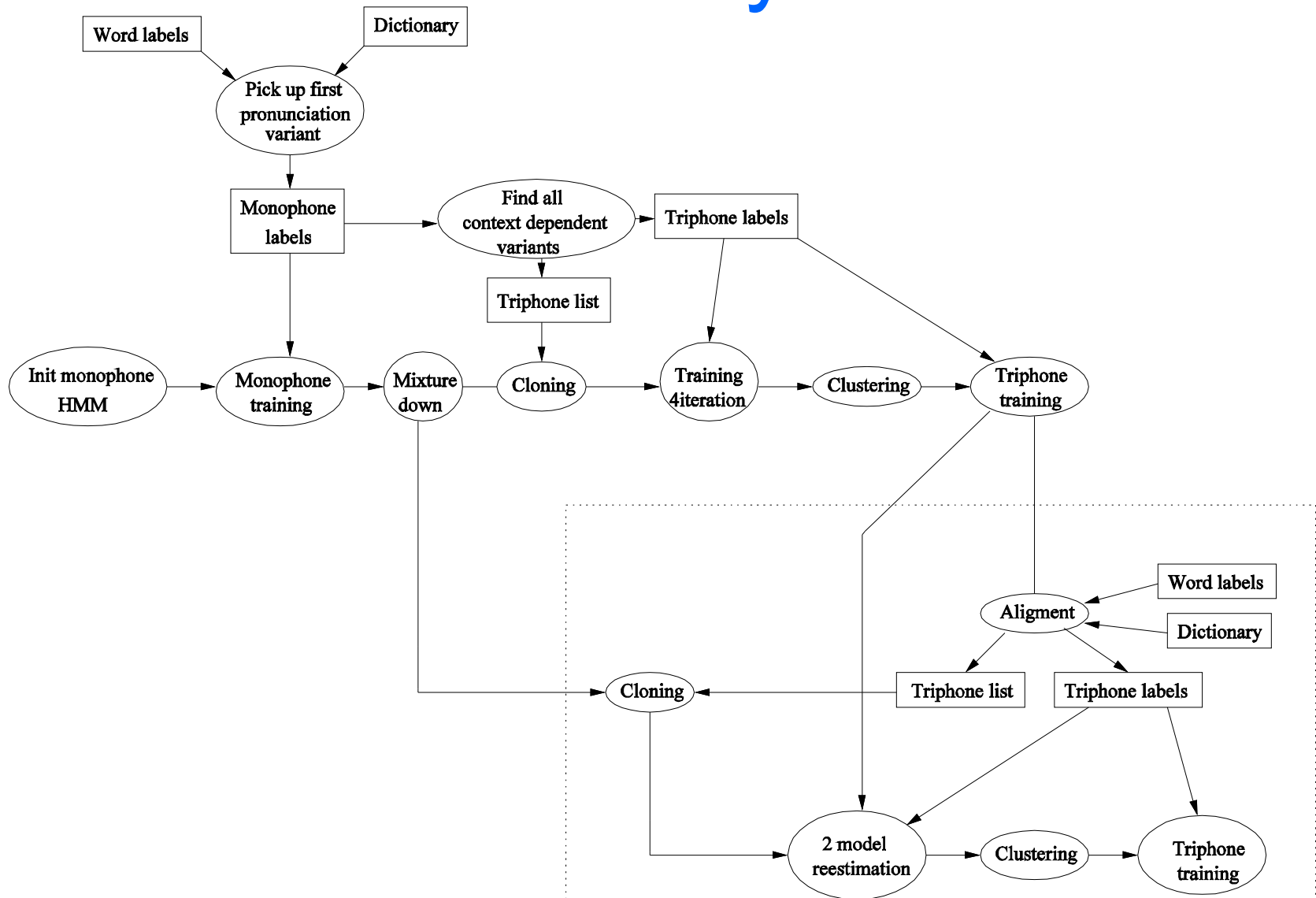
2. KWS in LVCSR lattices

- 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 ...

Ferret : Scripted-Meeting-TST-03 - Microsoft Internet Explorer - [Pracujete v režimu offline]

Soubor Úpravy Zobrazit Oblíbené Nástroje Nápověda

Zpět Hledat Oblíbené Média

Adresa E:\ami\pilot2\index.html

Přejít Odkazy

AMI Pilot meeting No. 2

00:00:00:00:0

Add Source...

ASR Find

KWS

[0:00] SO [0:00]
 [0:00] SO IT [0:01]
 [0:00] THAT WAS ALMOST [0:10]
 [0:00] NO NO THAT'S WOULD ALMOST I MEAN
 WE DON'T ACTUALLY HAVE ANYTHING
 ALMOST [0:04]
 [0:03] OKAY [0:03]
 [0:08] OH WE GOING [0:10]
 [0:10] SO IF YOU KNOW NOWING WHO KNOWS
 US [0:12]
 [0:19] WELL THAT'S [0:21]
 [0:24] YEAH [0:27]
 [0:29] UH [0:31]
 [0:35] UH [0:36]
 [0:42] UH [0:42]
 [0:46] THE KITE THIS OTHER WE'VE JUST GO
 BACK QUICK AGENDA SO THIS MEETING FOR
 THAT I GUESS THE GOAL OF THIS IS TO I
 AGREE ALL THE FUNCTIONAL THE LITTLE UM
 DISCUSS THE CORE FUNCTIONAL DO THAT
 WOULD HAVE BEEN DIRECT ROUTE [1:01]

[8:19] (-9.7561) UM I ALSO I LOOKED AT I.
 UM BY CONFORMS TO SEE WHAT PEOPLE
 COMPLAINDING AND HELP IN THE ENTER
 MIKE AND AND MORE BUT THE CAR
 INSTEAD OF THE [8:27]
 [10:09] (-57.5143) INTO THAT SOLD
 PROBABLY SO THAT'S ALSO THINGS THAT
 PEOPLE REALLY WANT AND WHAT
 COMPLAINDING NOISE YOU KNOW THEY'LL
 FORMALISE DIDN'T THERE A WHILE
 [10:18]

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.