Index of /mlmi04/MLMI-Talk-010/scannedSlides

Name	Last modified	Size	Description
Parent Directory	14-Oct-2004 17:35	-	
p0010s001.jpg	06-Oct-2004 11:06	332k	
p0010s002.jpg	06-Oct-2004 11:06	326k	
p0010s003.jpg	06-Oct-2004 11:06	378k	
p0010s004.jpg	06-Oct-2004 11:06	348k	
p0010s005.jpg	06-Oct-2004 11:06	404k	
p0010s006.jpg	06-Oct-2004 11:06	365k	
p0010s007.jpg	06-Oct-2004 11:06	384k	
p0010s008.jpg	06-Oct-2004 11:06	374k	
p0010s009.jpg	06-Oct-2004 11:06	312k	
p0010s010.jpg	06-Oct-2004 11:06	331k	
p0010s011.jpg	06-Oct-2004 11:06	322k	
p0010s012.jpg	06-Oct-2004 11:06	317k	
p0010s013.jpg	06-Oct-2004 11:06	301k	
p0010s014.jpg	06-Oct-2004 11:06	346k	
p0010s015.jpg	06-Oct-2004 11:06	321k	
p0010s016.jpg	06-Oct-2004 11:06	349k	
p0010s017.jpg	06-Oct-2004 11:06	318k	
p0010s018.jpg	06-Oct-2004 11:06	344k	
p0010s019.jpg	06-Oct-2004 11:06	305k	

p0010s020.jpg	06-Oct-2004 11:06	295k
p0010s021.jpg	06-Oct-2004 11:06	212k
p0010s022.jpg	06-Oct-2004 11:06	237k
p0010s023.jpg	06-Oct-2004 11:06	285k
p0010s024.jpg	06-Oct-2004 11:06	348k
p0010s025.jpg	06-Oct-2004 11:06	165k
thumb/	06-Oct-2004 11:40	-

Apache/1.3.31 Server at mmm.idiap.ch Port 80

File Edit Movie Favorites Window Help

Summary

Conversational Telephone Speech (CTS)

We present an overview of our effort to port the SRI CTS system to the Meeting task using:

- Adaptation (language and acoustic models)
- Preprocessing (noise reduction & array processing)
- Postprocessing (cross-talk elimination)

http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s002.jpg

File Edit Movie Favorites Window Help

The 2004 ICSI-SRI-UW Meeting Recognition System

International Computer Science Institute
Berkeley, CA, USA

Speech Technology & Research Laboratory
SRI International, Menlo Park, CA, USA

Signal, Speech & Language Processing Laboratory
University of Washington, Seattle, WA, USA

MLMI04

http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s002.jpg20.3.2006 8:58:04

File Edit Movie Favorites Window Help

Cast of Characters

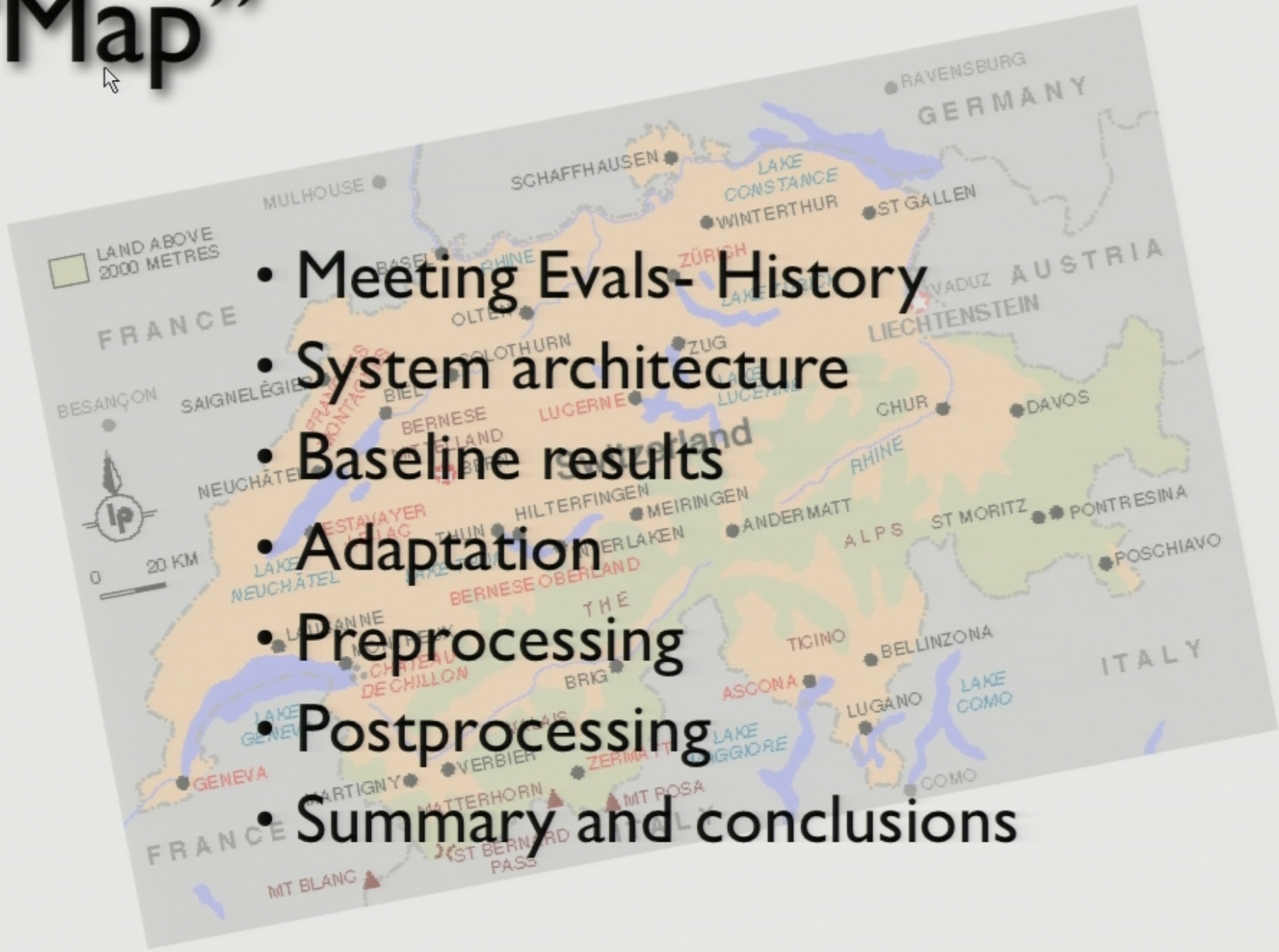
ICSI
Chuck Wooters (data czar, segmentation, postprocessing)
Nikki Mirghafori (text normalization, acoustic adaptation)
Tuomo Pirinen (array processing)
David Gelbart (noise filtering)
Barbara Peskin (advising)

SRI
Andreas Stolcke (acoustic modeling, system architecture)
Martin Graciarena (POF experiments)
Jing Zheng (structured MAP adaptation)
Ramana Gadde (help with segmentation)
Wen Wang (Fisher text processing, SuperARV LM)

UW
Ivan Bulyko (language models)
Scott Otterson (segmentation features)
Mari Ostendorf (advising)

File Edit Movie Favorites Window Help

"Map"



- Meeting Evals- History
- System architecture
- Baseline results
- Adaptation
- Preprocessing
- Postprocessing
- Summary and conclusions

http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s006.jpg

The "Meeting" Task

Conducted by the U.S. National Institute of Standards and Technology (NIST).
First NIST eval on Meeting Data held in 2002 (RT-02).
Second evaluation held this in the Spring of this year (RT-04).

	Recording Conditions	
	Close mics	Distant mics
RT-02	Headset Lapel	Single Distant Mic
RT-04	Individual Headset Mics (**IHM**) (no lapel)	Multiple Distant Mics (**MDM**)
		Single Distant Mic (**SDM**)

http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s006.jpg20.3.2006 8:58:43

http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s007.jpg

File Edit Movie Favorites Window Help

The "Meeting" Task

Conducted by the U.S. National Institute of Standards and Technology (NIST).
First NIST eval on Meeting Data held in 2002 (RT-02).
Second evaluation held this in the Spring of this year (RT-04).

	Recording Conditions		
	Close mics	Distant mics	
RT-02	Headset Lapel	Single Distant Mic	primary
RT-04	Individual Headset Mics (**IHM**) (no lapel)	Multiple Distant Mics (**MDM**)	primary
		Single Distant Mic (**SDM**)	

required contrast

http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s007.jpg20.3.2006 8:58:54

http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s008.jpg

Dev and Eval Data

Development data same as RT-02 Eval with updated transcripts.
Evaluation data consists of 2 meetings from each site. 1-2 hours each.
-Recognize 11 minute segments only.

		CMU	ICSI	LDC	NIST
RT-04 Eval Data	Personal	Head	Head	Head	Head
	Tabletop	1	6	4	7
Dev Data (a.k.a. RT-02 eval)	Personal	Lapel	Head	Lapel	Head
	Tabletop	1	6	10	7

- CMU has only 1 table-top mic, others have 4-10.
- RT-02 eval (with updated transcripts) set served as devtest.

http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s008.jpg20.3.2006 8:59:03

http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s009.jpg

File Edit Movie Favorites Window Help

Meeting Training Data

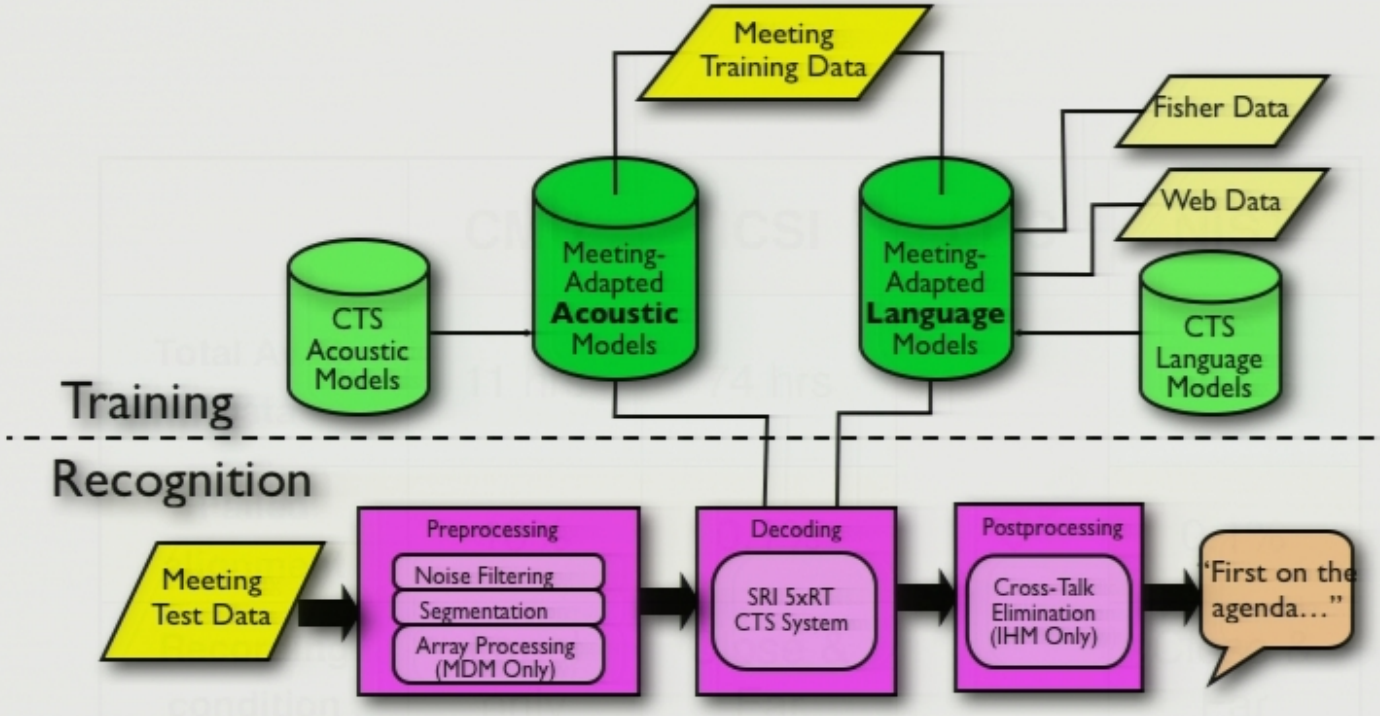
Data available for building the RT-04 Meeting recognition system.

	CMU	ICSI	LDC	NIST
Total Avail. Data	11 hrs	74 hrs	No Data	14 hrs
Failed alignment	10%	0.1%		0.1%
Recording condition	Lapel only	Close & Far		Close & Far

http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s009.jpg20.3.2006 8:59:10

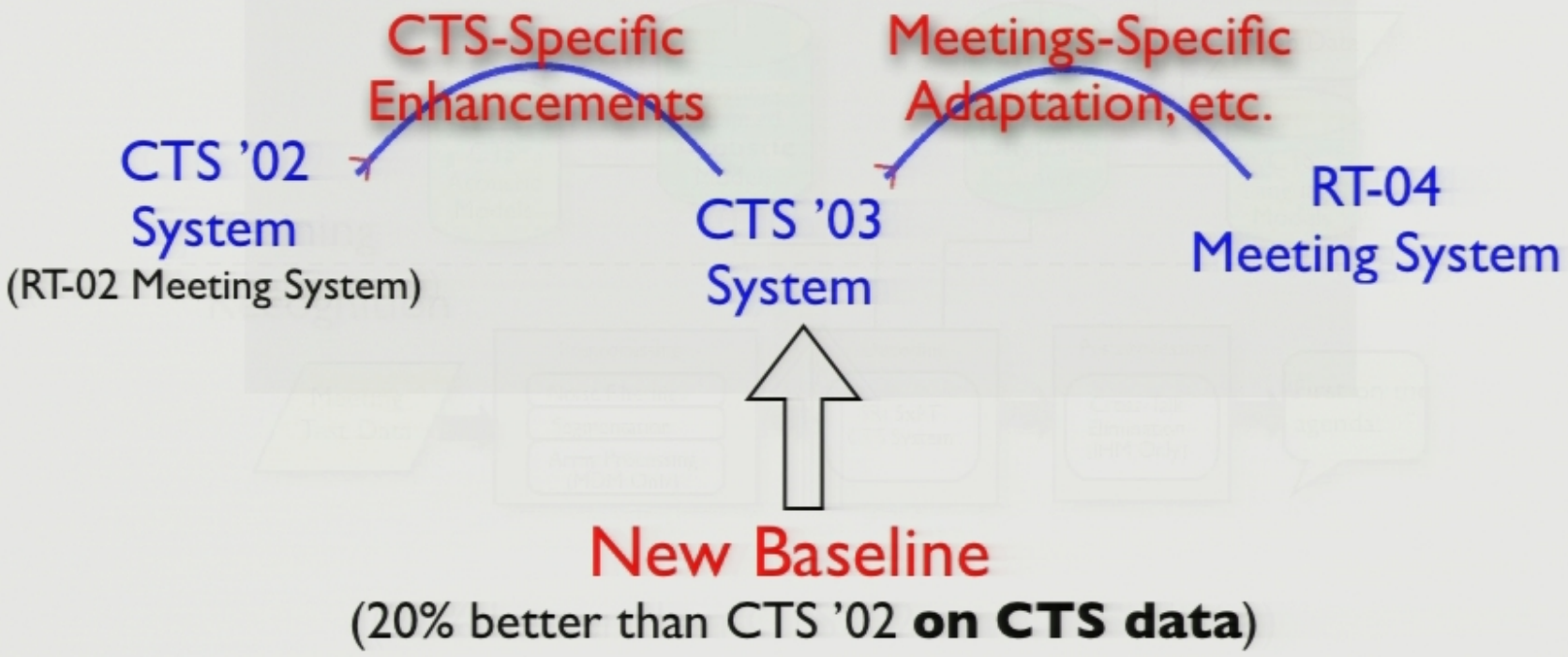
File Edit Movie Favorites Window Help

System Architecture



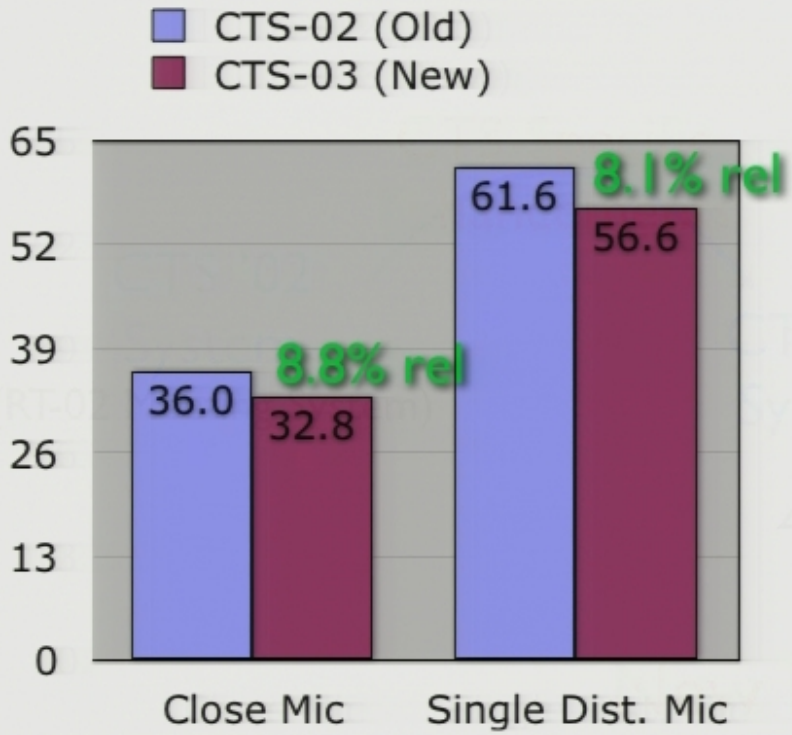
File Edit Movie Favorites Window Help

Baseline System Development



http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s012.jpg

Old Baseline vs. New Baseline on Meeting Data



Not as much gain as on CTS (20%), but a nice improvement.

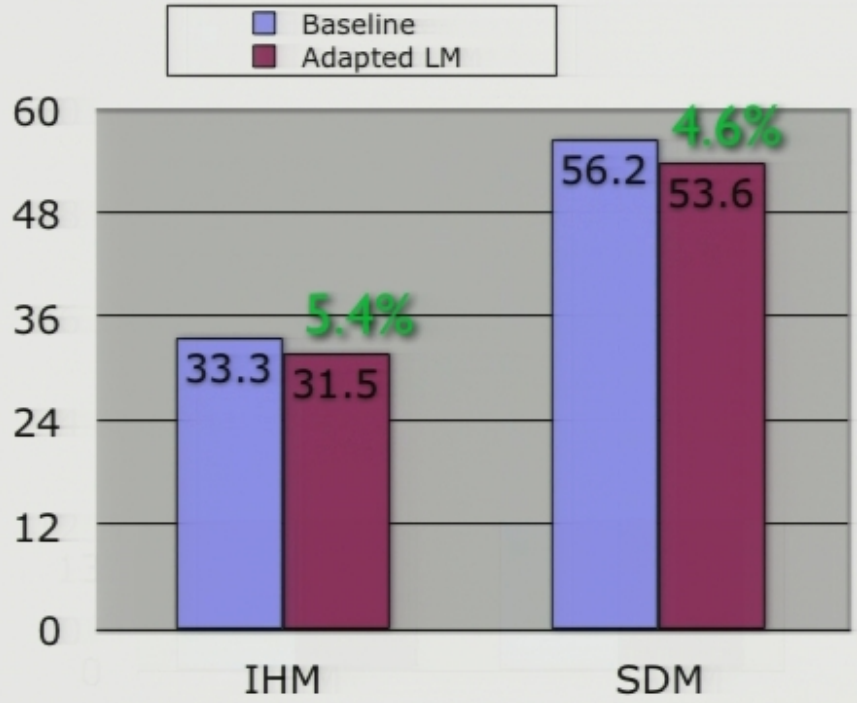
Error rate on RT-02 eval meeting data.

http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s012.jpg20.3.2006 8:59:39

File Edit Movie Favorites Window Help

Language Model Adaptation

- Started from CTS LM
- Added components for meeting domain
- Extended the Vocabulary



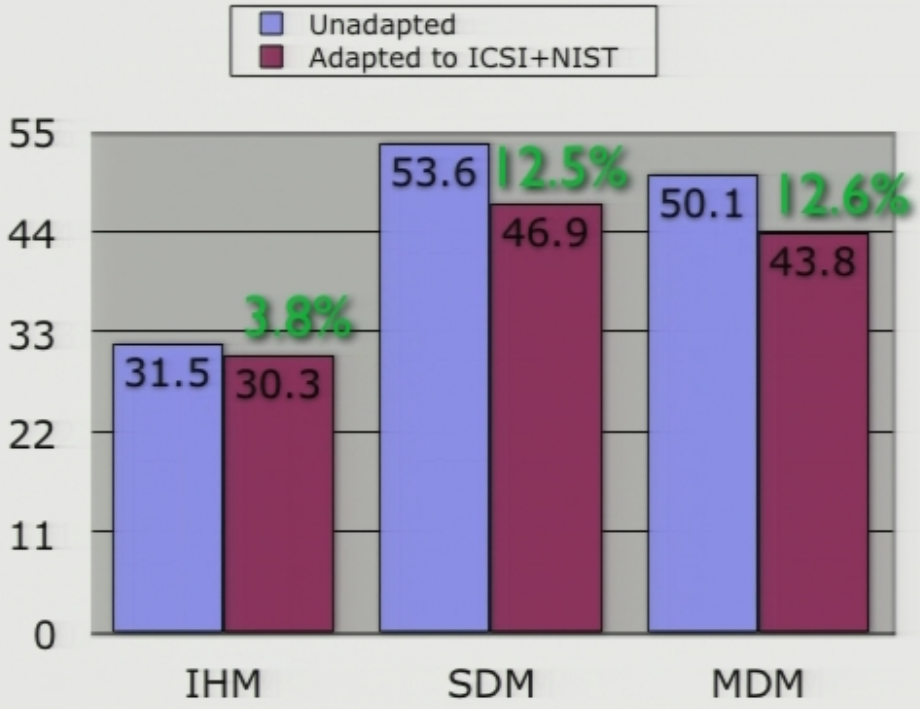
Roughly 5% improvement on both IHM and SDM.

Error rate on RT-04 dev data

http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s014.jpg

Acoustic Model Adaptation

- MAP-adapt acoustic models for each meeting source (using data from one or several sources)



Best to adapt to *all* distant mic channels

Captures acoustic variability due to both speaker and mic location (noise levels, reverberation)

Adapting to only central mic gives about 45-75% of possible gain.

Error rate on RT-04 dev data

http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s014.jpg20.3.2006 9:00:02

File Edit Movie Favorites Window Help

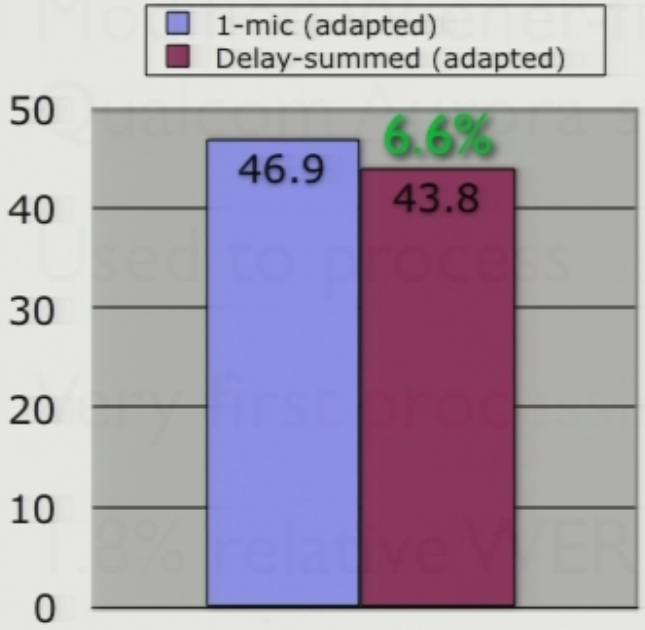
Noise Filtering

- Modified Wiener-filtering developed for ICSI-OGI-Qualcom Aurora system
- Used to process all distant mic channels
- Very first processing step, on full meeting length
- 1.8% relative WER reduction with baseline recognizer

http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s016.jpg

Array Processing

- Idea: add all distant mic signals after shifting (shift to compensate for delays)
- Speech is enhanced
- Noise is phase delayed and partially cancelled in summing



Baseline used "central" or "best" single mic.

Acoustic adaptation: still better (5% relative) to adapt to all distant channels, NOT to delay-summed signal.

Run after segmentation

Works best AFTER noise filtering

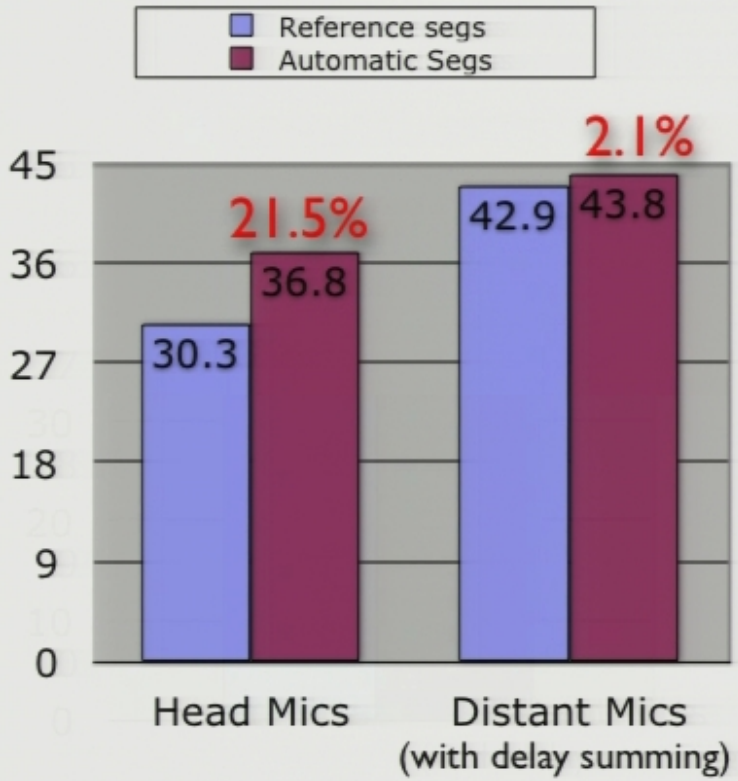
Error rate on RT-04 dev data

http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s016.jpg20.3.2006 9:00:28

File Edit Movie Favorites Window Help

Segmentation

- Segmenting meeting into isolated utterances.
- How much are we hurt by the automatic segmentors?



Biggest problem: cross-talk in IHM

Tried to address with postprocessing

Also developing cross-channel features (UW)

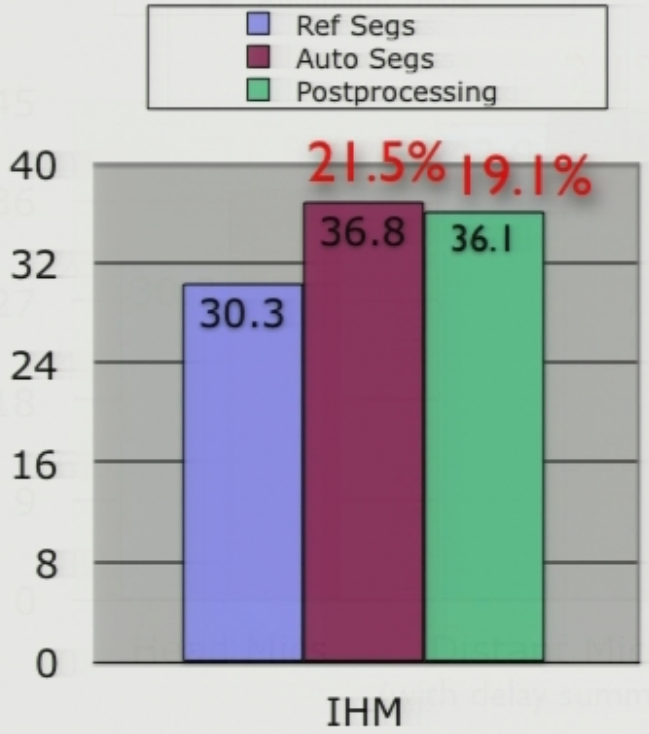
Error rate on RT-04 dev data

http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s018.jpg

File Edit Movie Favorites Window Help

Cross-talk Suppression

- Standard technique: delete words with low posteriors
- More sophisticated: look for words overlapping more than 50%, and delete those with low posteriors



Still lots of room for improvement (features that look across channels? or confidence measures?)

Post eval diagnostics: helpful for lapel, not headmounted mics

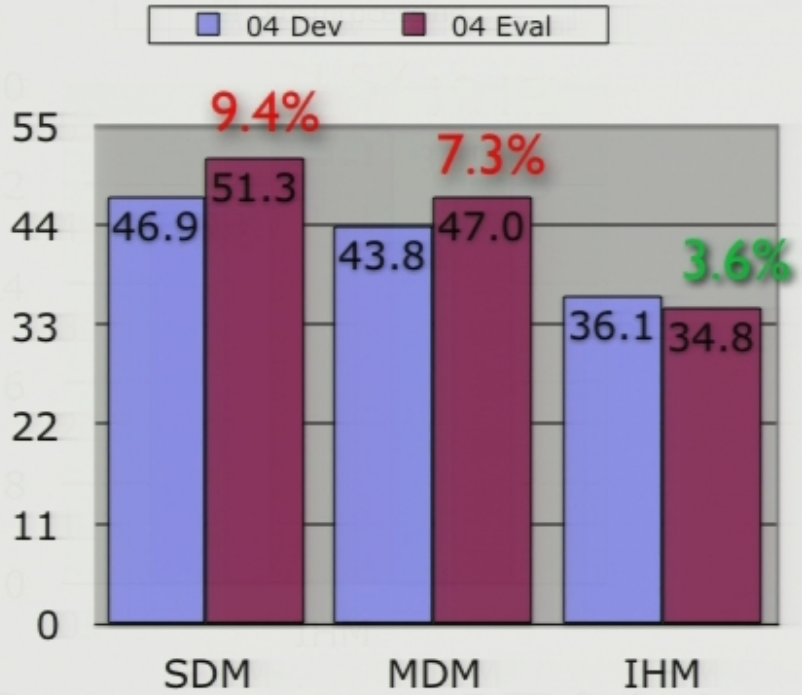
Error rate on RT-04 dev data

http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s018.jpg20.3.2006 9:00:47

http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s019.jpg

RT-04 Dev vs Eval

- How does the development data compare to the evaluation data?



The more difficult sources (CMU and LDC) constituted a larger portion of the 04 Eval data.

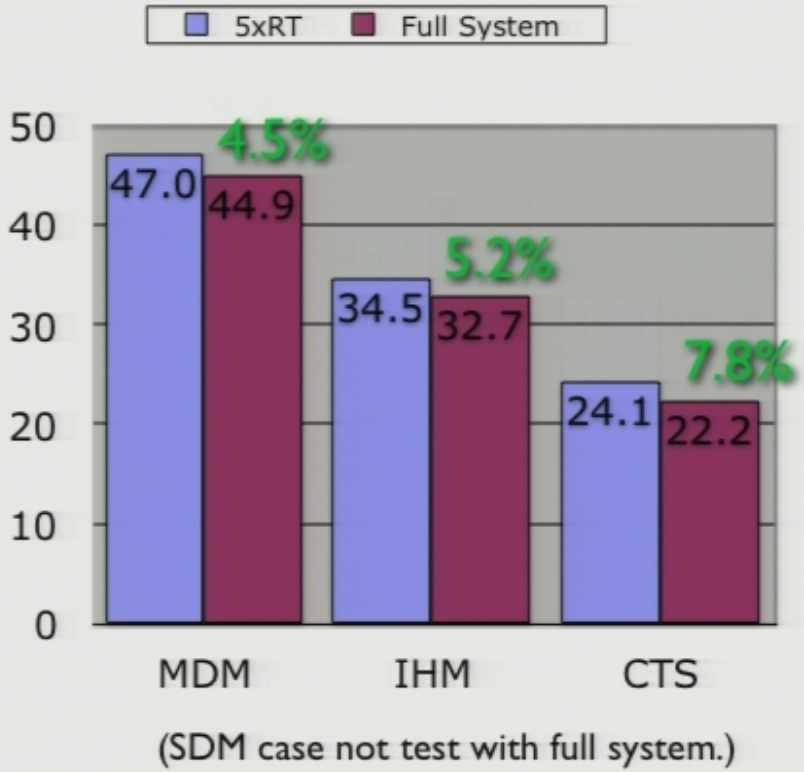
http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s019.jpg20.3.2006 9:00:58

http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s020.jpg

File Edit Movie Favorites Window Help

Full CTS System Results

- Full CTS eval system (20xRT)
- Performance on RT-04 Eval data



(SDM case not test with full system.)

Relative improvement on Meeting data is less than on CTS.

http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s020.jpg20.3.2006 9:01:07

File Edit Movie Favorites Window Help

Summary

	Baseline
IHM:	33.3
SDM:	56.2
MDM:	50.1

http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s022.jpg

File Edit Movie Favorites Window Help

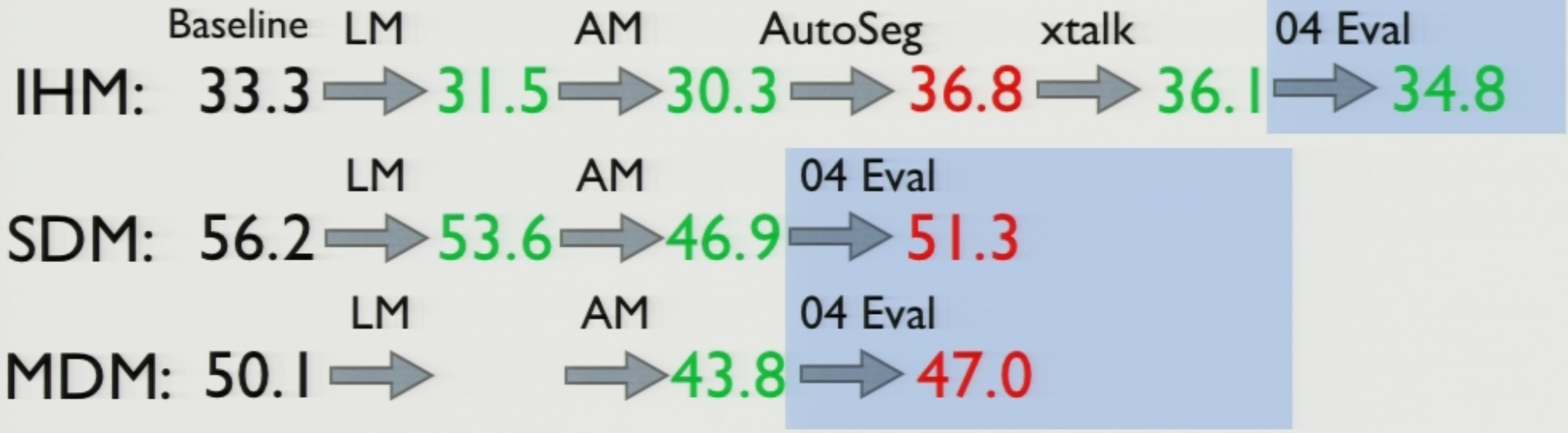
Summary

	Baseline	LM	AM	AutoSeg
IHM:	33.3	31.5	30.3	36.8
		LM	AM	
SDM:	56.2	53.6	46.9	
		LM	AM	
MDM:	50.1		43.8	

http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s022.jpg20.3.2006 9:01:23

http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s023.jpg

Summary



http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s023.jpg20.3.2006 9:01:33

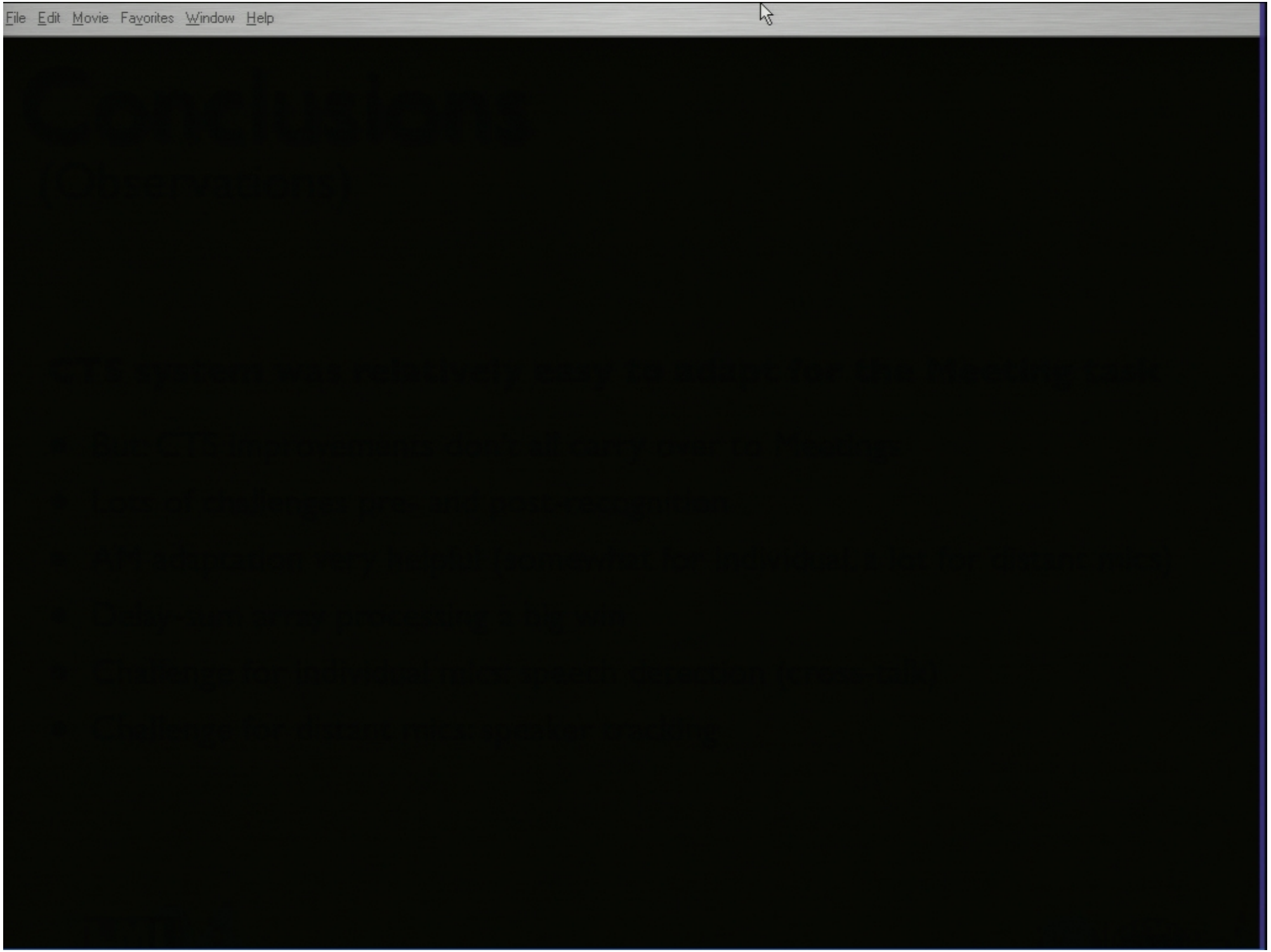
http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s024.jpg

Conclusions
(Observations)

CTS system was relatively easy to adapt for the Meeting task

- But: CTS improvements don't all carry over to Meetings
- Lots of challenges pre- and post-recognition
- AM adaptation very helpful (somewhat for individual, a lot for distant mics)
- Delay-sum array processing a big win
- Challenge for individual mics: speech detection (cross-talk)
- Challenge for distant mics: speaker tracking

http://mmm.idiap.ch/mlmi04/MLMI-Talk-010/scannedSlides/p0010s024.jpg20.3.2006 9:01:43


File Edit Movie Favorites Window Help