
The NICT ASR System for IWSLT 2012

H. Yamamoto, Y. Wu, C. Huang, X. Lu, P. R.
Dixon, S. Matsuda, C. Hori and H. Kashioka

2012.12.06



Task of the ASR track

- Automatic transcription of TED Talks
 - TED Talks: A collection of lectures on www.ted.com



- Spoken in English
- Spontaneous style
- Various topics
 - Technology
 - Entertainment
 - Design
- Non-speech
 - laugh, applause
 - music

Our motivation

IWSLT 2011	
System	WER
MIT	15.3
KIT	17.1
LIUM	17.4
FBK	18.2
NICT	27.3

[Federico+2011]

IWSLT 2012	
Catch up in straightforward way	
and tackle new problems	
12 pt. behind!	
First challenge on English LVCSR	



What we did

- Acoustic modeling

- Spontaneous style speaking
- Recording environment
- Non-speech (noise, music)
- Speaker switching
- Non-native speaker

Collect audio of the target condition

Train and combine two types of AMs

Important, but look small relatively

- Language modeling

- Spontaneous style sentence
- Variety of topics

Extend RNNLM to use multiple features

Adapt N-gram to domain & topic

System Overview

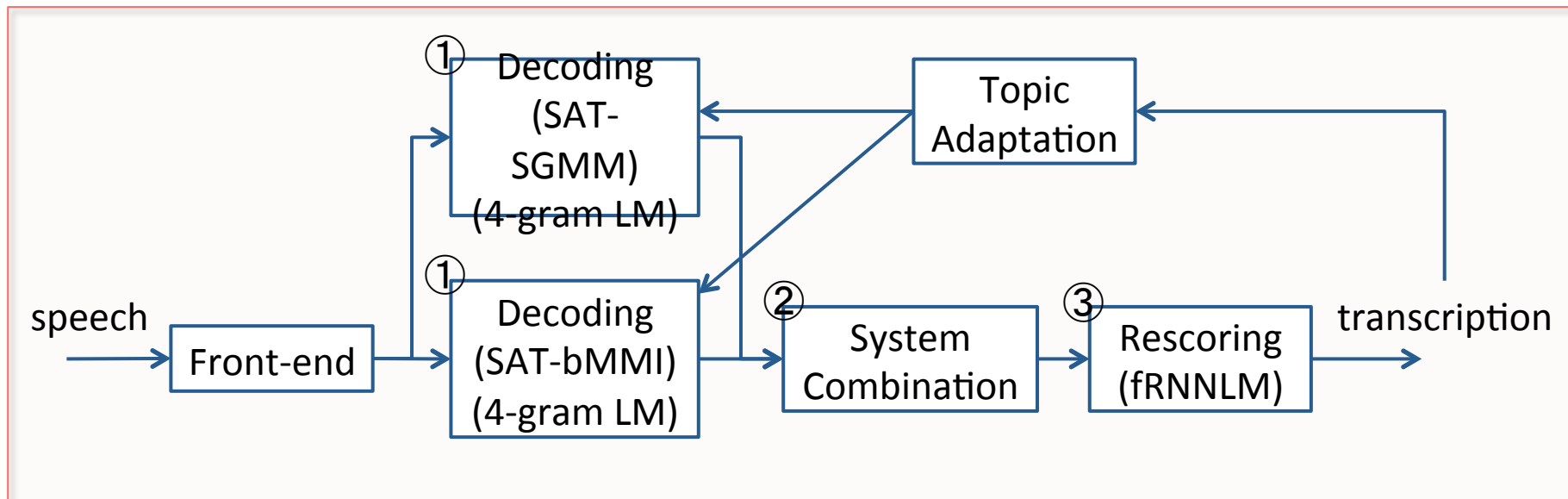
- WFST-based two-pass decoding

Two Passes

1. Domain adapted LM
2. Topic adapted LM

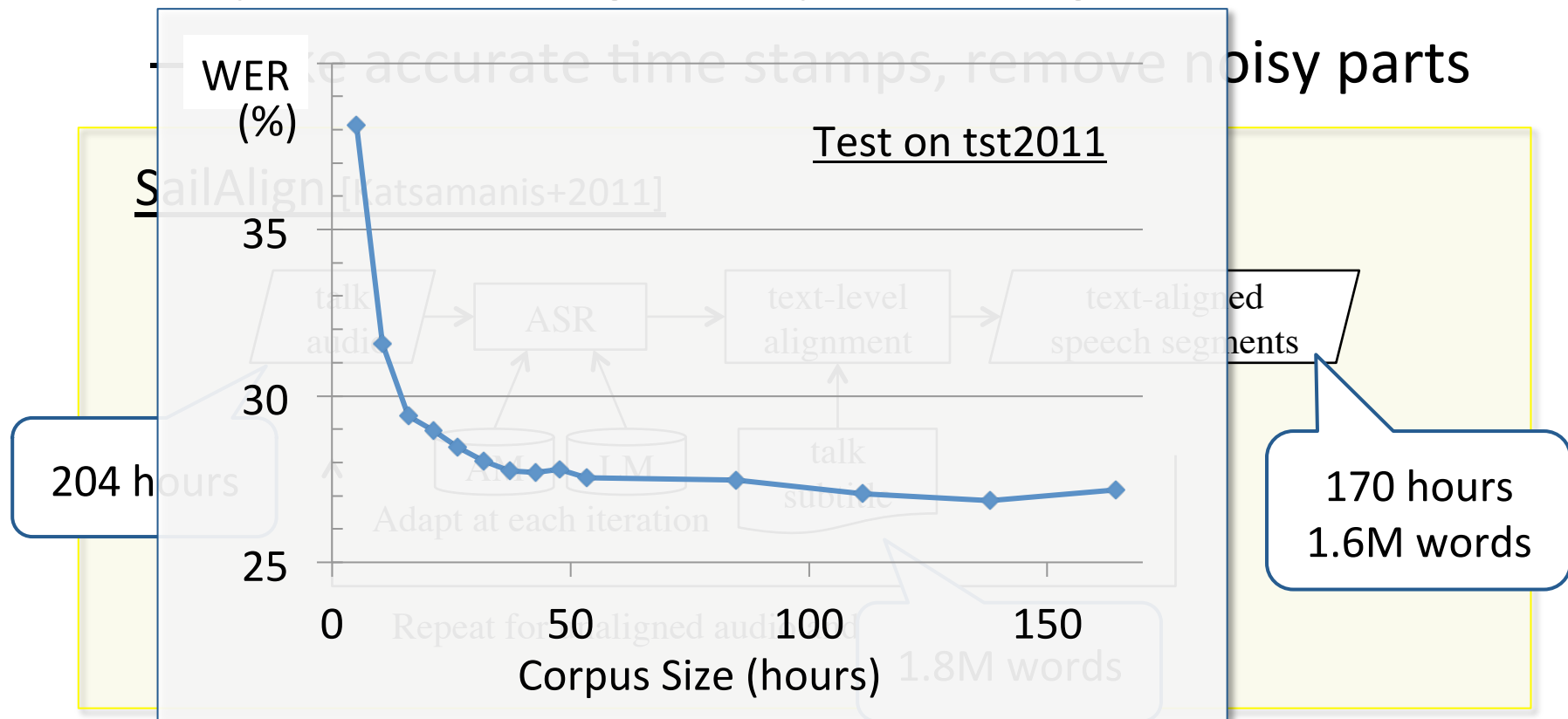
Steps in each pass

1. Decoding w/ N-gram LM
2. System Combination
3. Rescoring w/ fRNN LM



AM: Training corpus

- TED Talks AM ← crawl movies and subtitles
- Prepare text-aligned speech segments



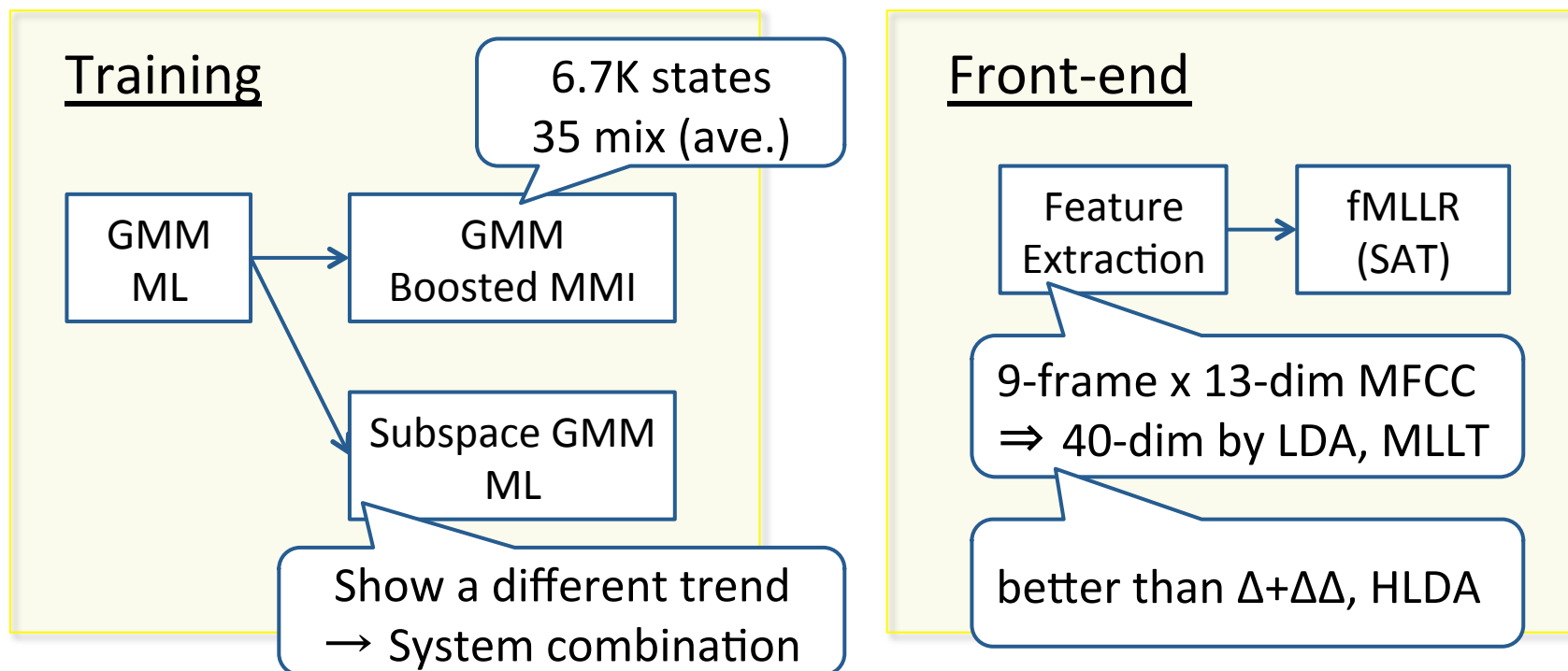
AM: Modeling

- Spontaneous speech AM

Kaldi [Povey+2011]

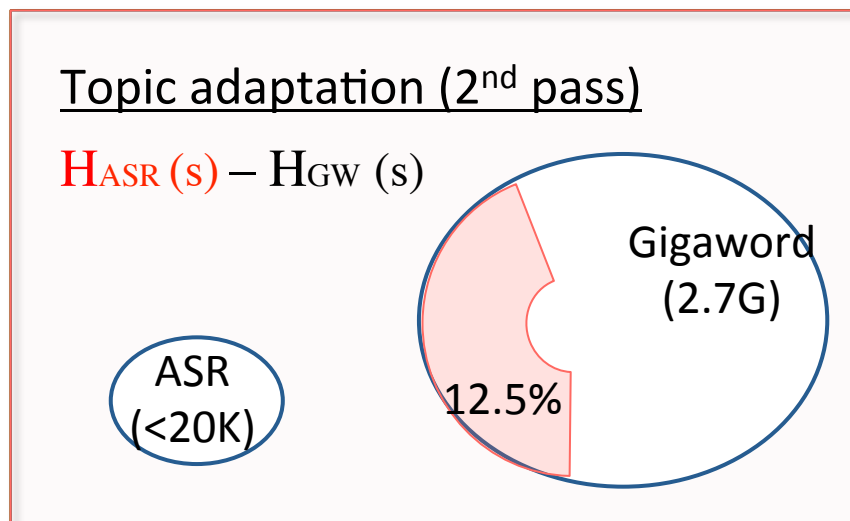
- Cross-word triphone HMM

- Two types of output pdf., GMM and SGMM

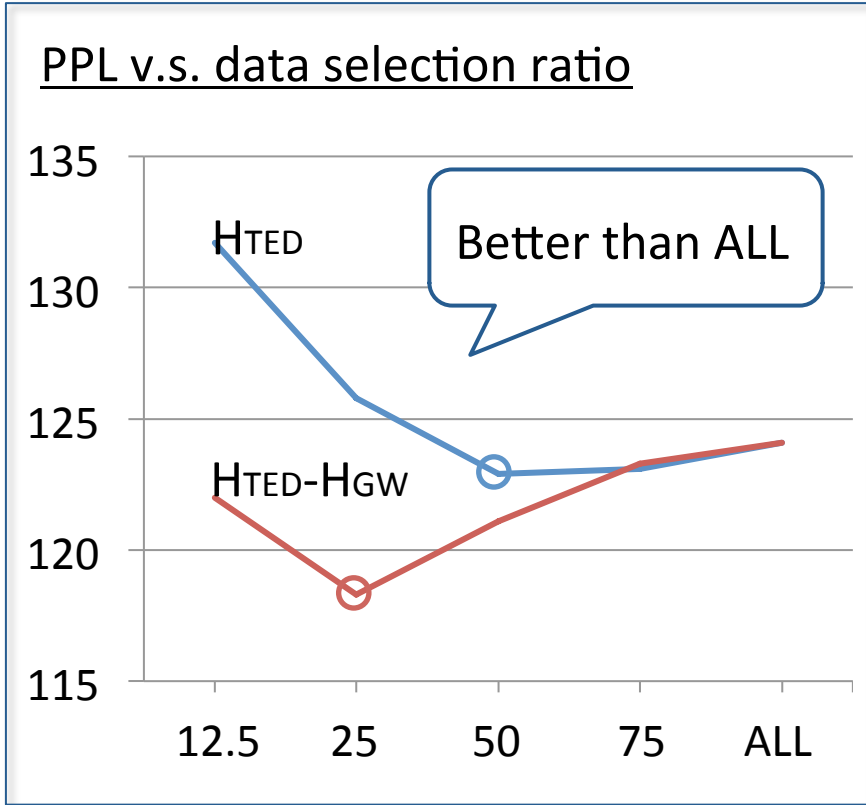


LM: N-gram

- Data selection for LM adaptation
 - Corpora: TED(in-domain), Gigaword(out-of-domain)
 - Cross-entropy difference metrics [Moore+2010]

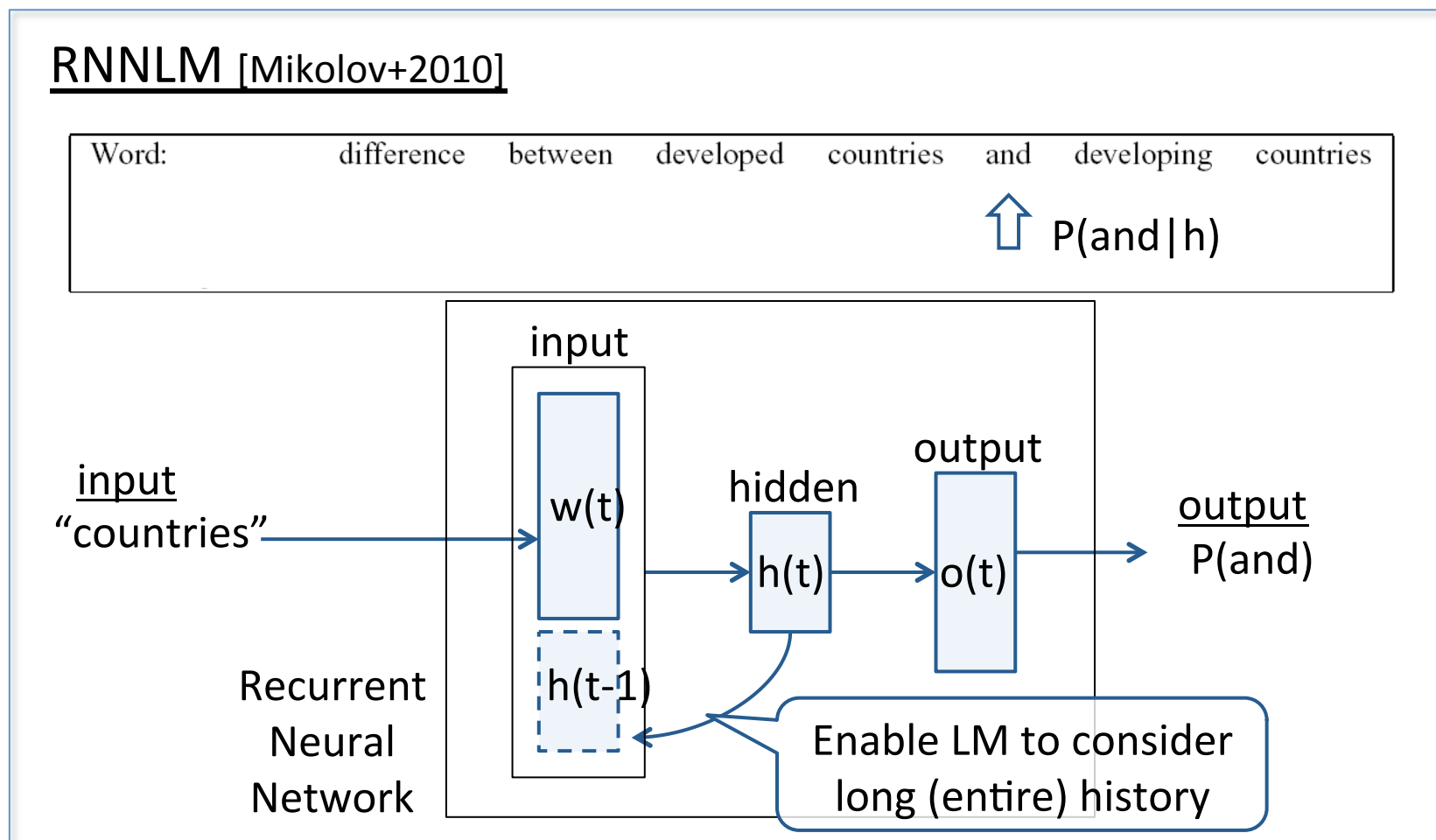


mKN smoothed
4-gram LM



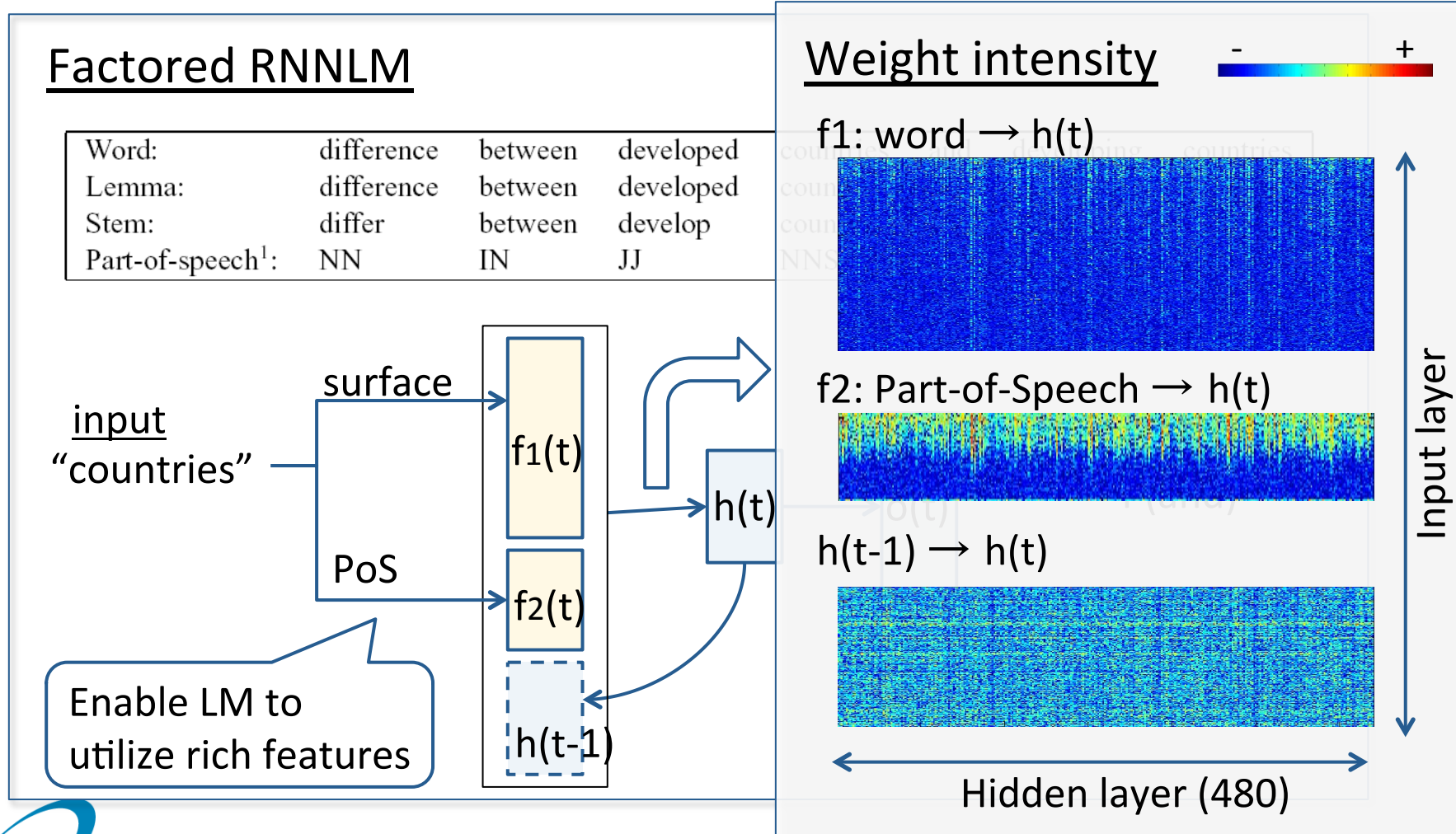
LM: Factored RNNLM [Wu+2012]

- Incorporate multiple features into RNNLM



LM: Factored RNNLM [Wu+2012]

- Incorporate multiple features into RNNLM



Experimental Results

- WER(%) of our transcriptions.

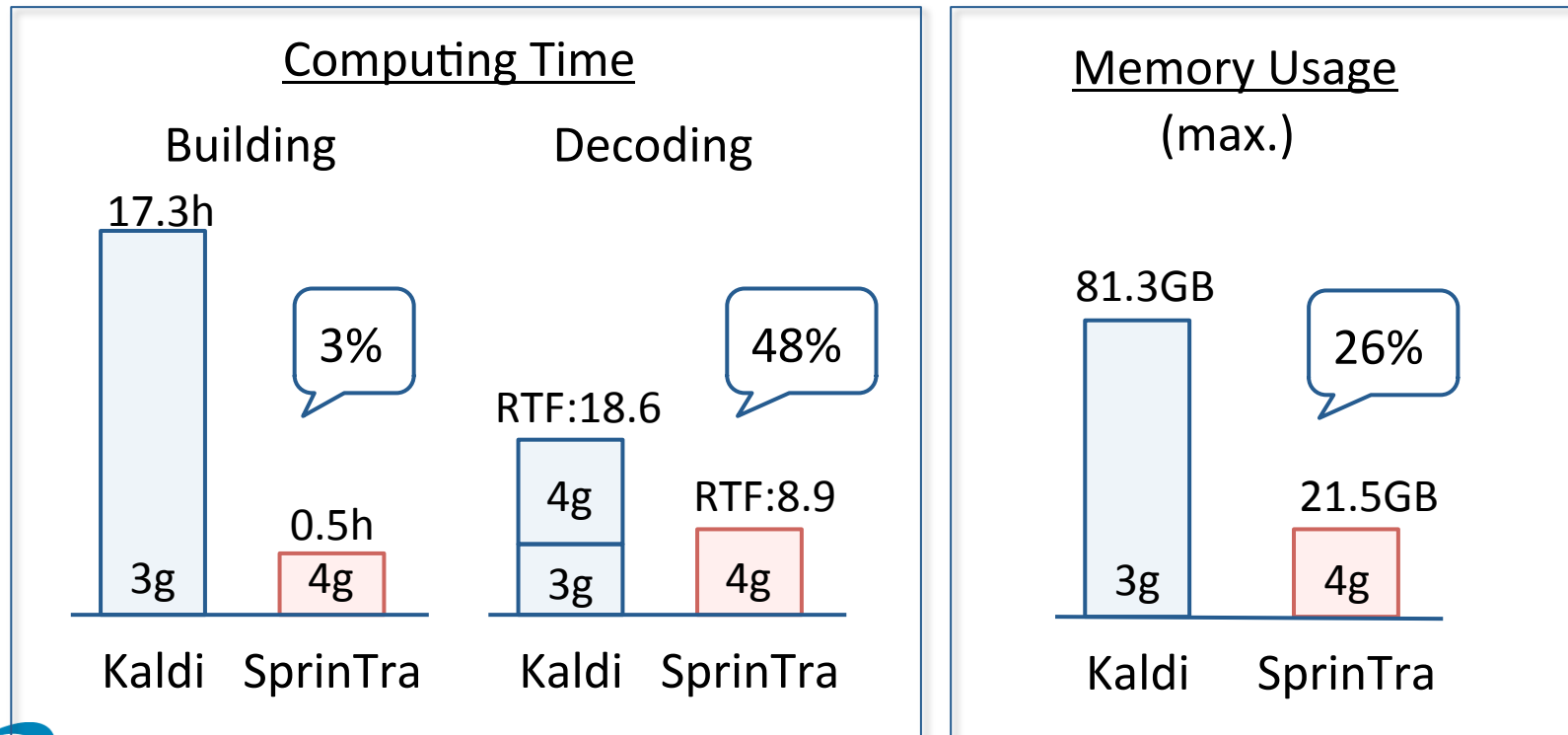
	<u>tst2011</u>		<u>tst2012</u>	
	1 st pass	2 nd pass	1 st pass	2 nd pass
Step 1: GMM	12.3	11.8	13.9	13.4
Step 1: SGMM	12.9	12.5	14.2	13.5
Step 2: Comb.	12.0	11.5	13.3	12.8
Step 3: fRNN	10.9	10.6	12.1	12.0

Each step contributed to reduce error

Topic adaptation also worked

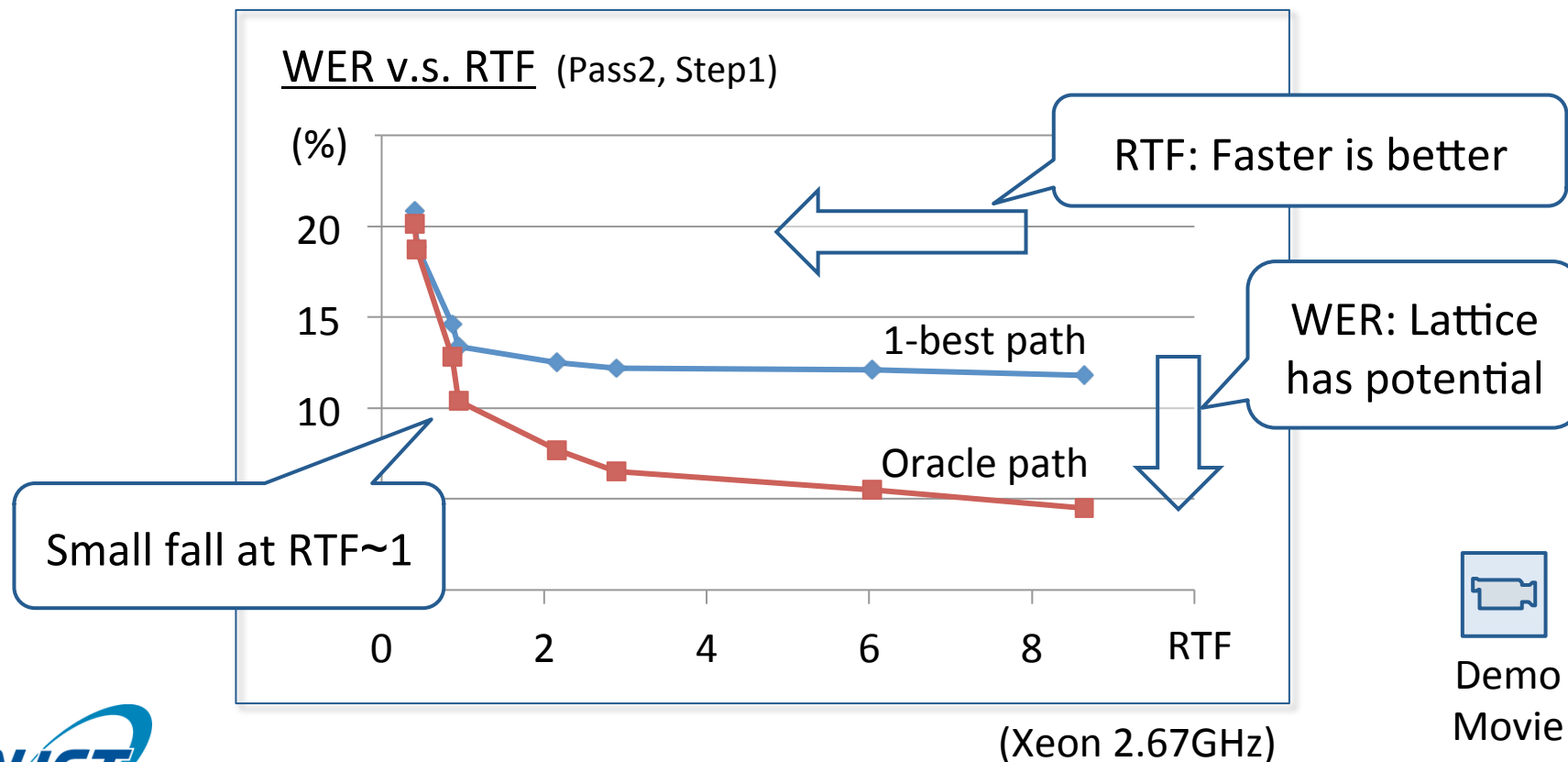
After the submission, fRNNLM adaptation achieved 11.9%.

- SprinTra decoder [Dixon+2012]
 - On-the-fly WFST composition scheme.
 - Low computing resource, no degradation in WER.



(Xeon 2.67GHz)

- Decoding speed is also a key issue of ASR
 - To response quickly for online ASR (e.g. Closed captioning)
 - To process tons of data for offline ASR (e.g. Audio indexing)



Thanks for your kind attention!

(147) and i actually SET at one conference a couple years ago JIMMY my DAMNED data because you people can't be trusted to

(13) one of the senior doctors at my hospital charlie SAFFRON and his colleague warner slack HAD been saying for decades that the most underutilized resource in

(131) just incredible

(132) who knows what WILL be able to do when we learn to make more use of

(133) AND the punch line is that HE YOU'RE IN a half later i was there when this magnificent young woman my daughter got married

(134) and when she came down THE steps QUINTETS just her NEED for that moment i was so glad that she DID have to say THAT WERE mother i wish dad could have been here

(135) and this is what we're doing when we make OUR FARE better

(136) now i want to talk about a couple of other patients who are doing everything in their power to improve health care

(137) this is RICH IN A HOLIDAY a painter in washington DC whose husband died of kidney cancer a year after my disease

(138) she's painting here a mural of his horrible final weeks in the hospital

(139) one of the things that she discovered was that her husband's medical record in this paper folder was just disorganized

(140) she thought you know if i have a NEW TRICIA IN FACT label on the side of a cereal box why can't there be something that simple telling every new YORK SOME comes on duty every doctor the basics about my husband's condition

(141) so she CAME TO this medical facts mural with THE TRICKS AND label THE something like that

(142) she then last year painted this diagram

(143) she studied HEALTHCARE like me

(144) she came to realize that there were a lot of people who'd written patient advocate POTS BUT you just don't hear about IT medical conferences

(145) patients are such an underutilized resource

(146) well as it SAT in my introduction i've GOTTEN somewhat KNOWN for saying that patients should have access to their data

SprinTra(RTF8)

(1) it's an amazing thing that we're here to talk about the year of patients rising
(2) you heard stories earlier today about patients who are taking control of their cases patients who are saying you know what i know what the odds are but i'm going to look for more information

(3) i'm going to define what the terms of my success OR

(4) AND to be sharing with you how four years ago i almost died I found out i was in fact already almost dead

(5) and what i then found out about what's called the INPATIENT movement ALL explain what that term means

(6) i HAVE been blogging under the name patient DAVID when i discovered this i just renamed myself EAT patient dave

(7) regarding the word patient when i first started a few years ago getting involved in health care and attending meetings as just a casual observer i noticed that people would talk about patients as if it was somebody who's not in the room here somebody out there

(8) some of our talks today we still act like that

(9) but i'm here to tell you THE patient is not a third person WORK

(10) you yourself will find yourself in a hospital bed or your mother your child THEIR heads NOT EVEN people who say yes i know exactly what you mean

(11) so WHEN THE when you hear what I'VE NEVER TALKED about here today first of all i want to say that i am here on behalf of all the patients that i have ever met all the ones i haven't met

(12) this is about letting patients play a more active role in helping health care in fixing health care

SprinTra
(RTF1)

SprinTra
(RTF8)

References

- [1] M. Federico, et al., “Overview of the IWSLT 2012 Evaluation Campaign,” in *Proc. of IWSLT, 2012*.
- [2] A. Katsamanis, et al., “SailAlign: Robust long speech-text alignment,” in *Proc. of Workshop on New Tools and Methods for Very-Large Scale Phonetics Research, 2011*.
- [3] D. Povey, et al., “The Kaldi Speech Recognition Toolkit,” in *Proc. of Workshop on Automatic Speech Recognition and Understanding, 2011*.
- [4] R. Sproat, et al., “Normalization of non-standard words,” *Computer Speech and Language, Vol. 15*, pp. 287–333, 2001.
- [5] J. R. Novak, et al., “Improving WFST-based G2P Conversion with Alignment Constraints and RNNLM Nbest Rescoring,” in *Proc. of Interspeech, 2012*.
- [6] R. Moore and W. Levis, “Intelligent selection of language model training data,” in *Proc. of ACL, 2010*.
- [7] A. Stolcke, et al., “SRILM at Sixteen: Update and Outlook,” in *Proc. of Workshop on Automatic Speech Recognition and Understanding, 2011*.
- [8] T. Mikolov, et al., “Recurrent neural network based language model,” in *Proc. of Interspeech, 2010*.
- [9] Y. Wu, et al., “Factored Language Model based on Recurrent Neural Network,” in *Proc. of COLING, 2012*.
- [10] Y. Wu, et al., “Factored Recurrent Neural Network Language Model in TED Lecture Transcription,” in *Proc. Of IWSLT, 2012*.
- [11] P. R. Dixon, et al., “A Comparison of Dynamic WFST Decoding Approaches,” in *Proc. of ICASSP, 2012*.