Direct Observation of Pruning Errors: A Search Analysis Tool

Volker Steinbiss, Martin Sundermeyer, Hermann Ney
{steinbiss,sundermeyer,ney}@cs.rwth-aachen.de

1st Errare Workshop, Ermenonville
November 22nd, 2013

Human Language Technology and Pattern Recognition
Chair of Computer Science 6
Computer Science Department
RWTH Aachen University, Germany

RWTH AACHEN UNIVERSITY

Outline

Search in Automatic Speech Recognition

- **Find most likely word sequence w_1^N given acoustic signal x_1^T**

$$\arg\max_{w_1^N} \{p(w_1^N|x_1^T)\} = \arg\max_{w_1^N} \{p(w_1^N) \cdot p(x_1^T|w_1^N)\}$$

- **Many different search algorithms possible [Aubert 2002]**
- **Here: history-conditioned lexical tree search [Ney & Ortmanns 1999]**
- **Complexity of full search**

$$O(T \cdot H \cdot G)$$

where
T**: length of acoustic signal**
H**: number of Hidden Markov Model states**
G**: number of n-grams in language model**

Pruning and Pruning Errors

- Search space too large for efficient search
- Solution: apply pruning
- Pruning = heuristic for reducing the search effort
- Pruning parameters chosen empirically,
  common practice: word error rate vs. real-time factor curves
- Problems of pruning
  - Optimum word sequence $\hat{w}_1^N$ may not be found
  - Only indirect detection of pruning errors
- Goal: detailed, non-heuristic analysis

Concepts of a Pruning Error

- **Different notions of a pruning error:**
  1. **Pruning a hypothesis that would have resulted in the globally optimal path**
     $$\exists t : \hat{s}_t \in P_t$$
     with
     $\hat{s}_1^T$: **optimum state sequence**
     P_t: **set of hypotheses pruned at time step** t
  2. **Pruning a hypothesis resulting in an additional recognition error**
     $$\exists t : \hat{s}_t \in P_t \wedge \left[\forall s_1^T \not\ni \hat{s}_t : WER(\hat{s}_1^T) < WER(s_1^T) \right]$$
     where
     $WER(s_1^T)$: **word error rate obtained for word sequence induced by** s_1^T

Detecting Pruning Errors

- **Problems**
  - ▷ **Type (1) requires exhaustive search**
  - ▷ **Type (2) immature, ‘helpful’ pruning errors**
- **Detection of pruning errors**
  - ▷ **Type (1): compute forced alignment, compare with recognition Better score implies pruning error (sufficient, but not necessary)**
  - ▷ **Type (2): run series of recognition experiments, compute error rates**

New Definition of a Pruning Error

- **Definition of a pruning error:**

  3. Pruning a hypothesis that would have resulted in the globally optimal path through the *spoken* word sequence

$$\exists t : \bar{s}_t \in P_t$$

  with $\bar{s}_1^T$: *spoken* state sequence

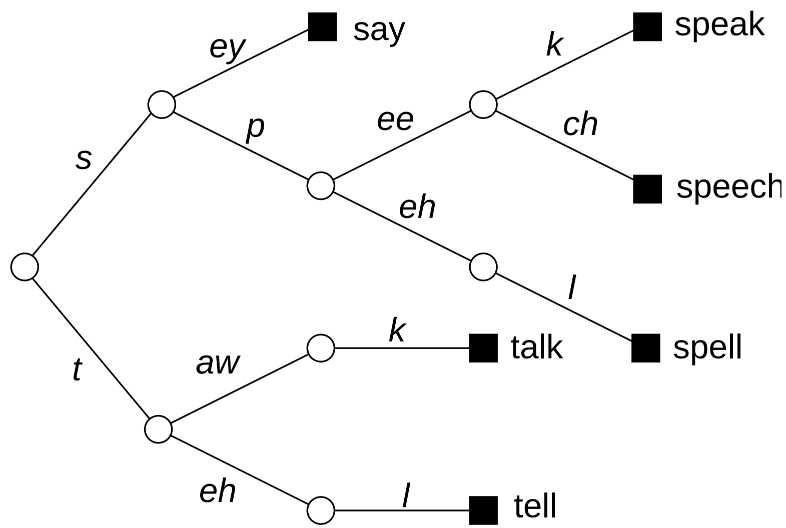
- **Assumption: spoken word sequence (= manual transcription) is known**

Proposed Approach

1. Compute spoken state sequence $\bar{s}_1^T$ by forced alignment of spoken word sequence
2. For each time frame t: Keep track of
   - alignment HMM state (spoken state)
   - preceding words (spoken history)
3. Recognition pass: check whether spoken state with spoken history is active or has been pruned
4. Evaluate results

Excursus: RWTH Decoder

- **History-Conditioned Tree Search**
  - **Search Space is organized as prefix tree**



  - **Prefix tree copy for each language model context**
  - **Language Model Lookahead**

Excursus: RWTH Pruning Methods

- **Pruning Methods**
  - **Acoustic Pruning**

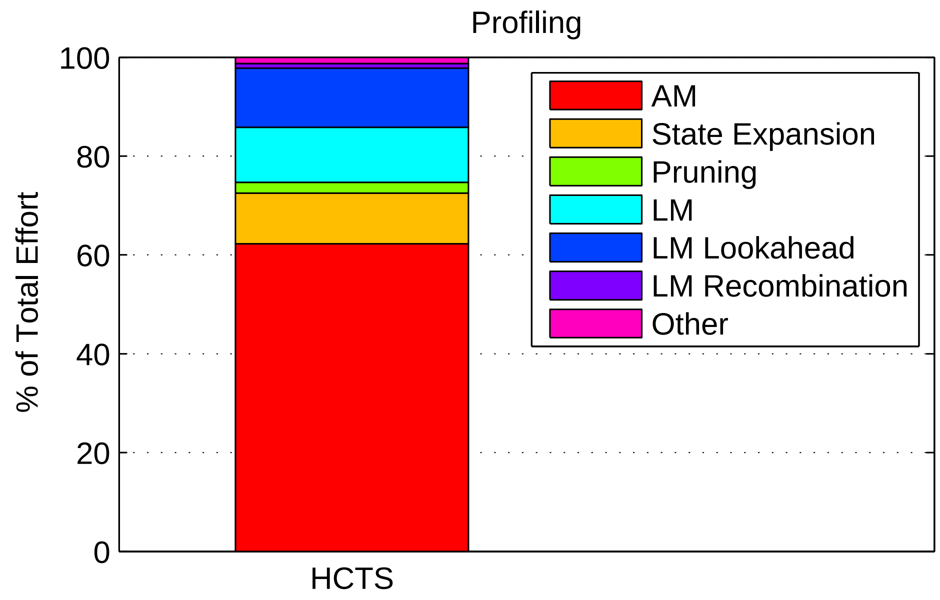
$$Q_h(t, s) < f_{\mathrm{AC}} \cdot Q_{\mathrm{AC}}(t)$$

where
$Q_{\mathrm{AC}}(t)$: **largest probability at time** t
$Q_h(t, s)$: **probability for history** h **and state** s
f_{AC}: **pruning parameter**

  - **Language Model Pruning**

$$Q_h(t, s = 0) < f_{\mathrm{LM}} \cdot Q_{\mathrm{LM}}(t)$$

- **Profiling**



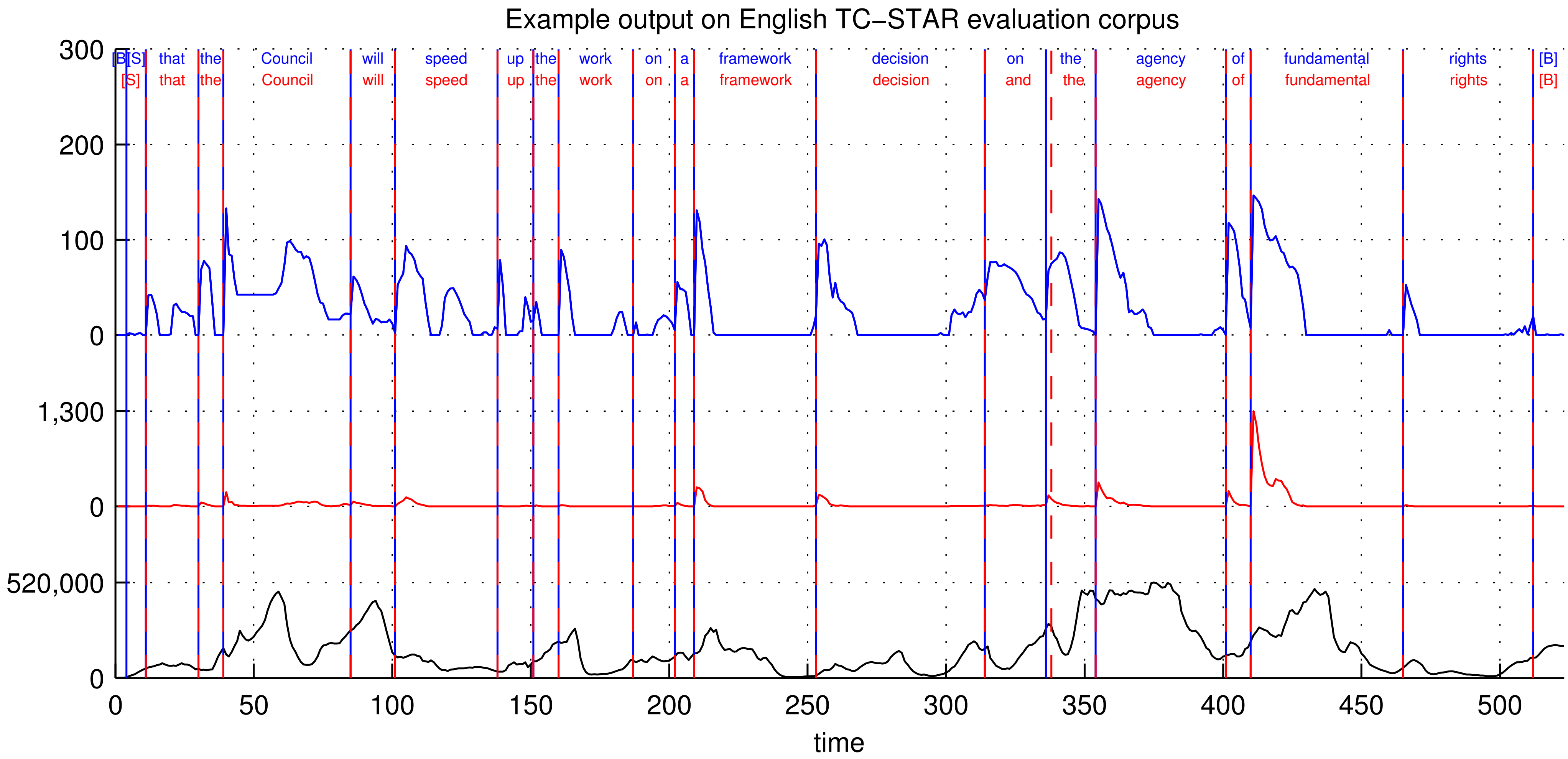
Experimental setup

- RWTH English EPPS system
- Good results in TC-STAR evaluation campaign

EPPS English	
Acoustic data	2.9 hours
Number of n-grams	7,400,000
Vocabulary size	53,300
OOV	0.0 %
WER	14.7 %

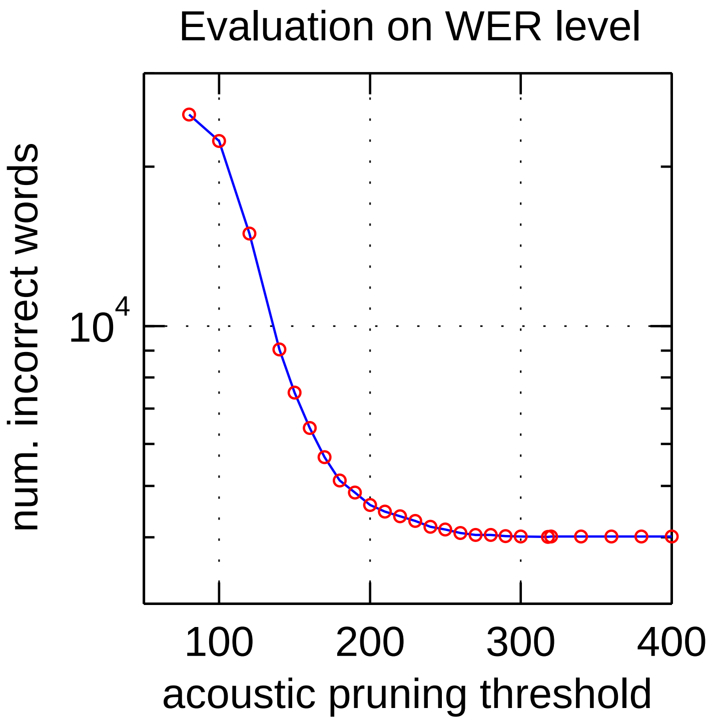
- Single pass recognition with VTLN adaptation
- OOV rate of original system is 0.1 %

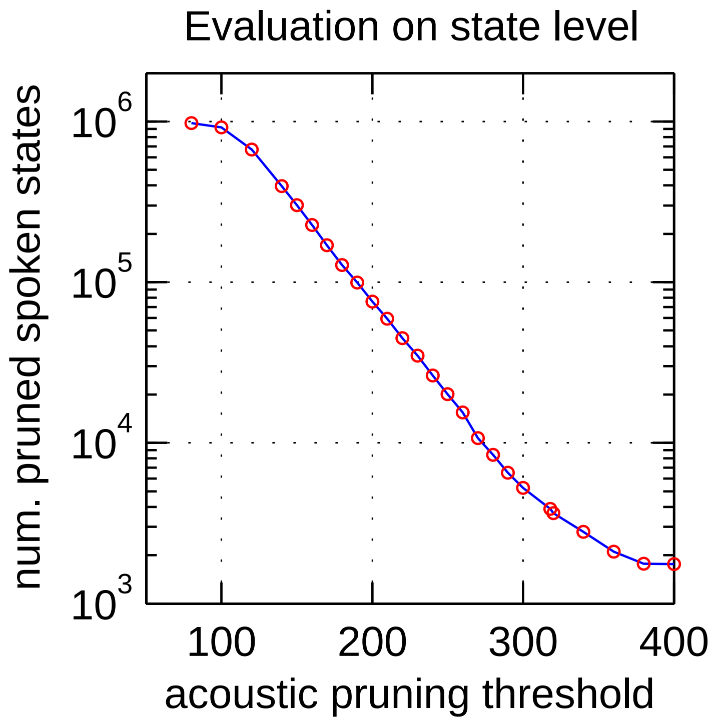
Example Output



- **Top to bottom: transcription, recognized word sequence, score of spoken state, number of hypotheses better than spoken state, number of hypotheses before pruning**

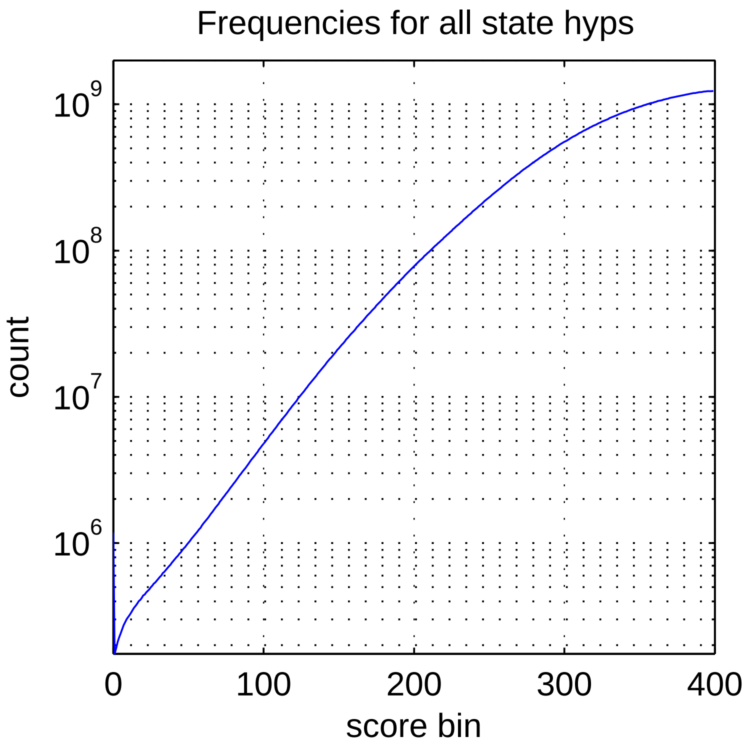
Search Error Analysis

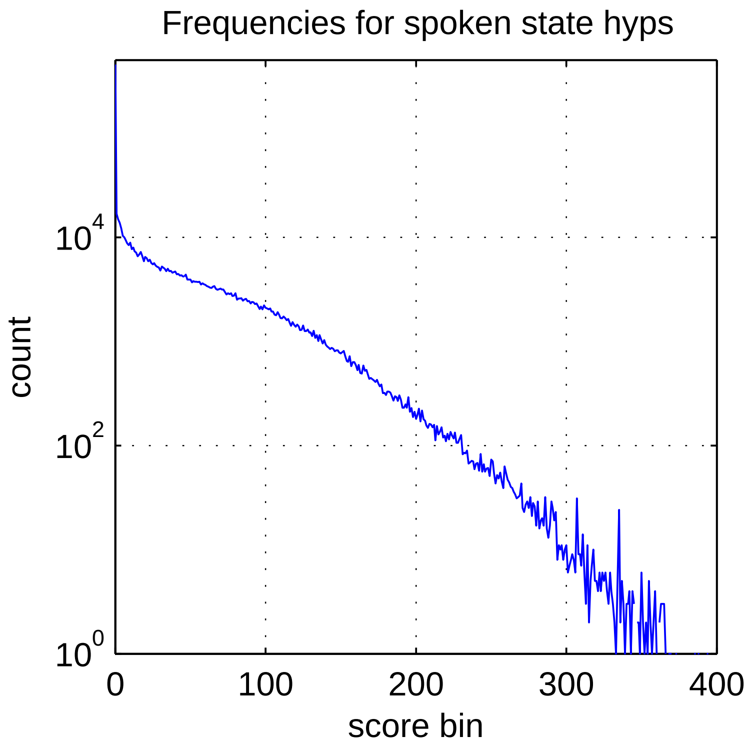




- **Word error rate level analysis:**
  - **Sensitive for strong pruning**
  - **Insensitive for weak pruning**
  - **Slightly noisy**
- **Spoken state level analysis:**
  - **Constant sensitivity**
  - **No noise**

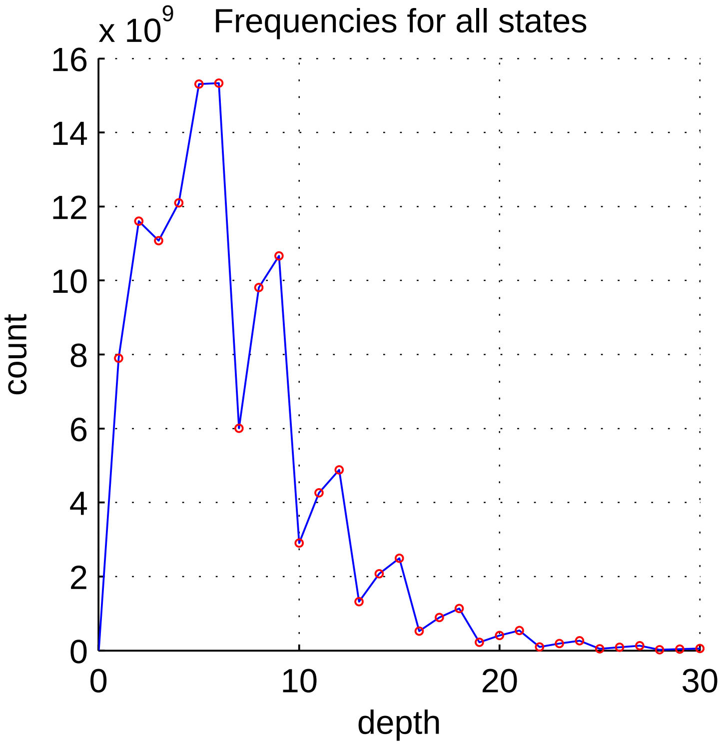
Analysis of State Space

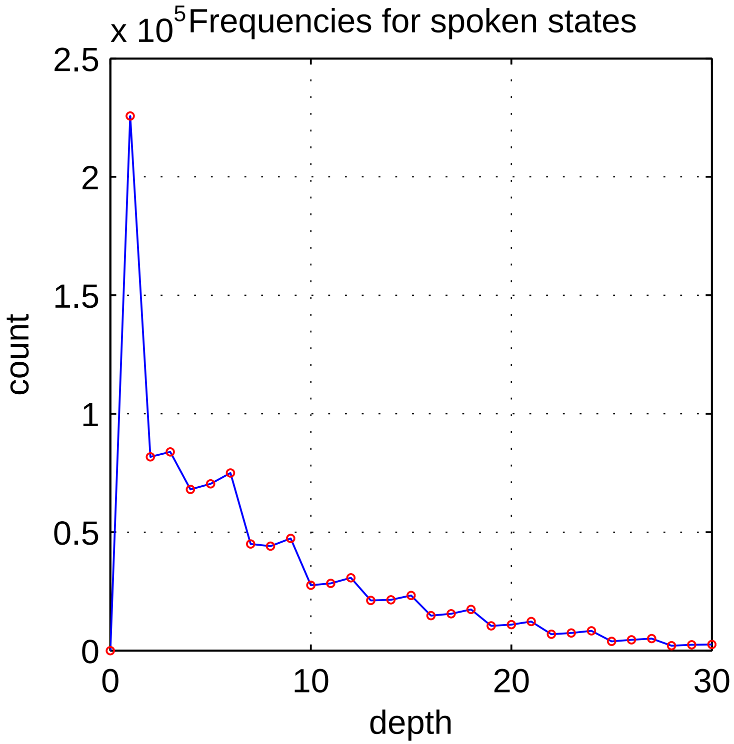




- **All states:**
  Number of hypotheses grows strongly with increasing score
- **Spoken states:**
  Fewer hypotheses with higher scores (facilitates acoustic pruning)
- **Tight pruning: 21,177 hypotheses, only 637 better than spoken one!**

Analysis of Prefix Tree State Space

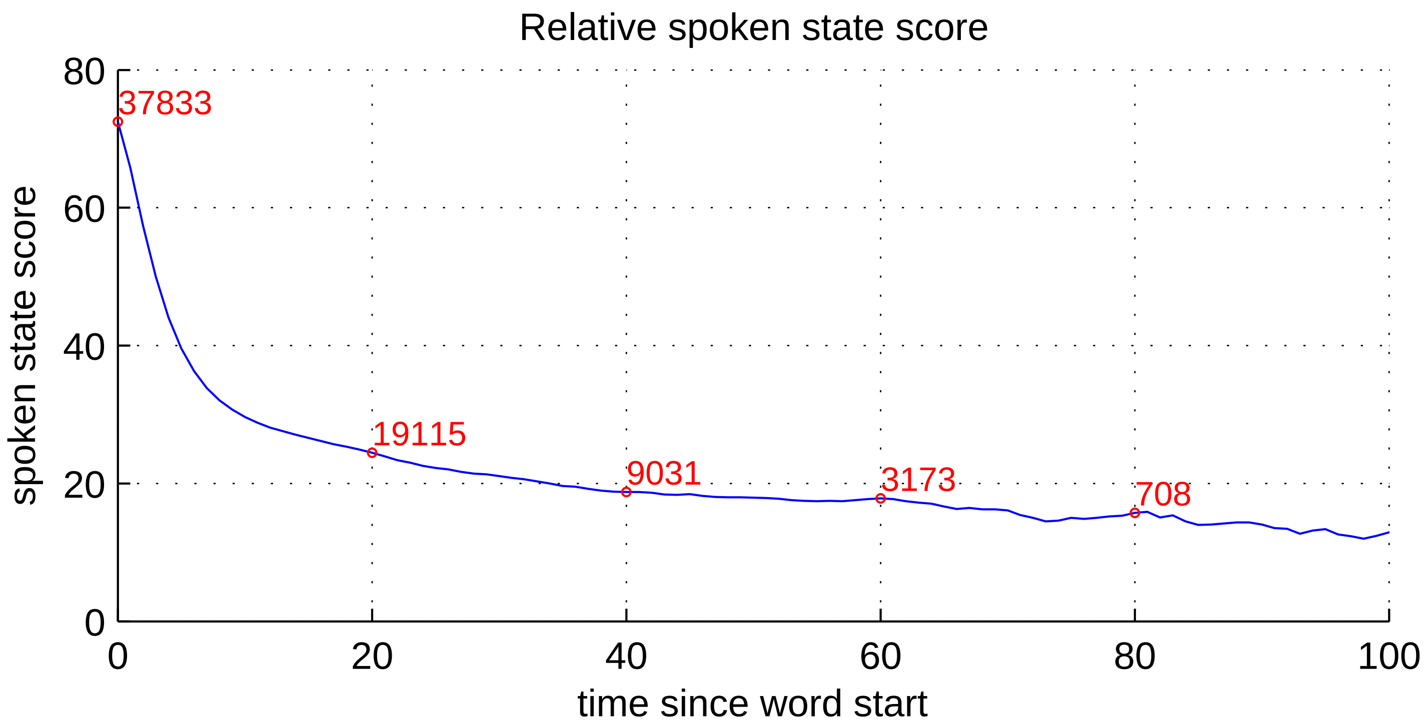




- **All states (see also [Aubert 2002]):**
  First 2 phonemes: 60 % of state hypotheses (before pruning)
  First 7 phonemes: 99 % ~
- **Spoken states:**
  State frequencies mainly follow word length distribution

Analysis of Word Hypotheses

- **Observation: no conceptual difference between scores of state/word hypotheses**
- **But: spoken state scores decrease over time**



Language model pruning can be tighter

Conclusion

- Tool for observation of effects of pruning
- Analysis at state hypothesis level
- Experiments show validity of approach
- Pruning methods not at their limits yet!?
- General concept, applicable to other decoders and search problems
- Future work
  - Location of errors
  - Calibration of pruning methods

References

[Ney & Ortmanns 1999] “Dynamic programming search for continuous speech recognition”, IEEE Signal Processing Magazine, Vol. 16, No. 5, pp. 64–83, 1999

[Pylkkönen 2005] “New pruning criteria for efficient decoding”, in Ninth European Conference on Speech Communication and Technology, 2005

[Lööf et al. 2007] “The RWTH 2007 TC-STAR evaluation system for European English and Spanish”, in Proc. of International Conference on Speech Communication and Technology, Antwerp, Belgium, 2007, pp. 2145–2148

[Aubert 2002] “An overview of decoding techniques for large vocabulary continuous speech recognition”, Computer Speech and Language, Vol. 16, No. 1, pp. 89–114, 2002

Thank you for your attention

Martin Sundermeyer

{steinbiss,sundermeyer,ney}@cs.rwth-aachen.de

http://www-i6.informatik.rwth-aachen.de/