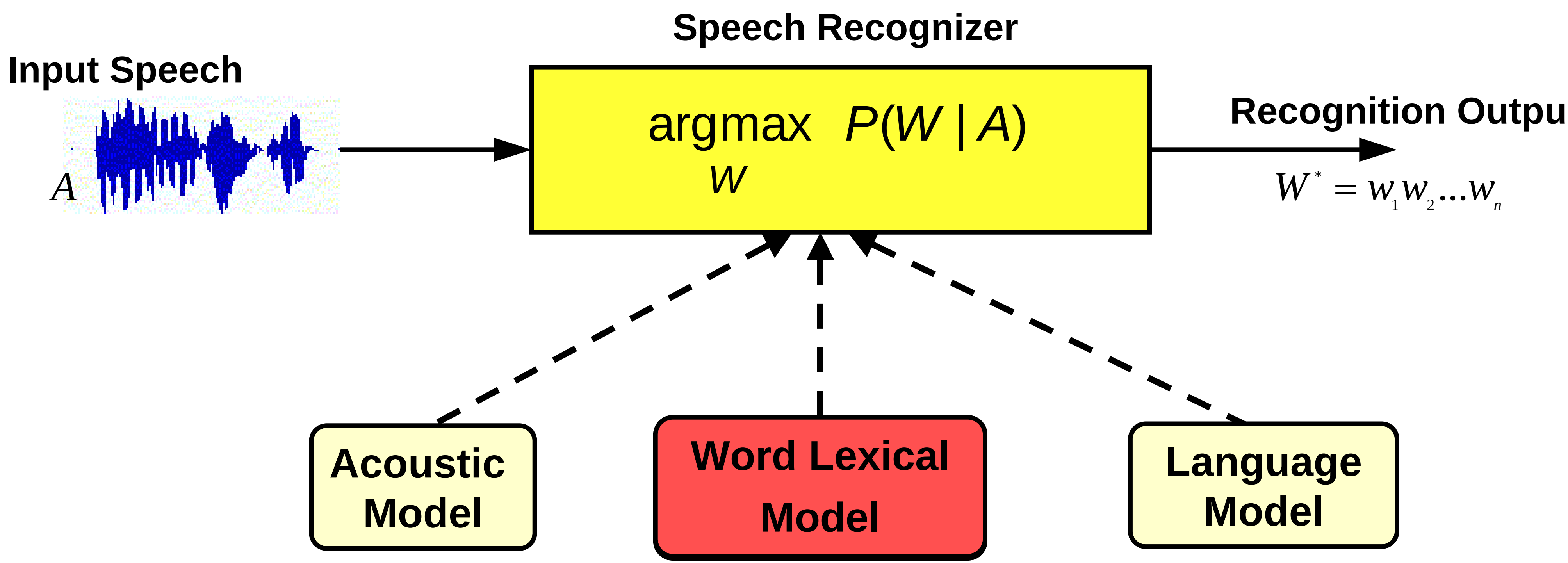
Lecture # 19
Session 2003

Modelling New Words

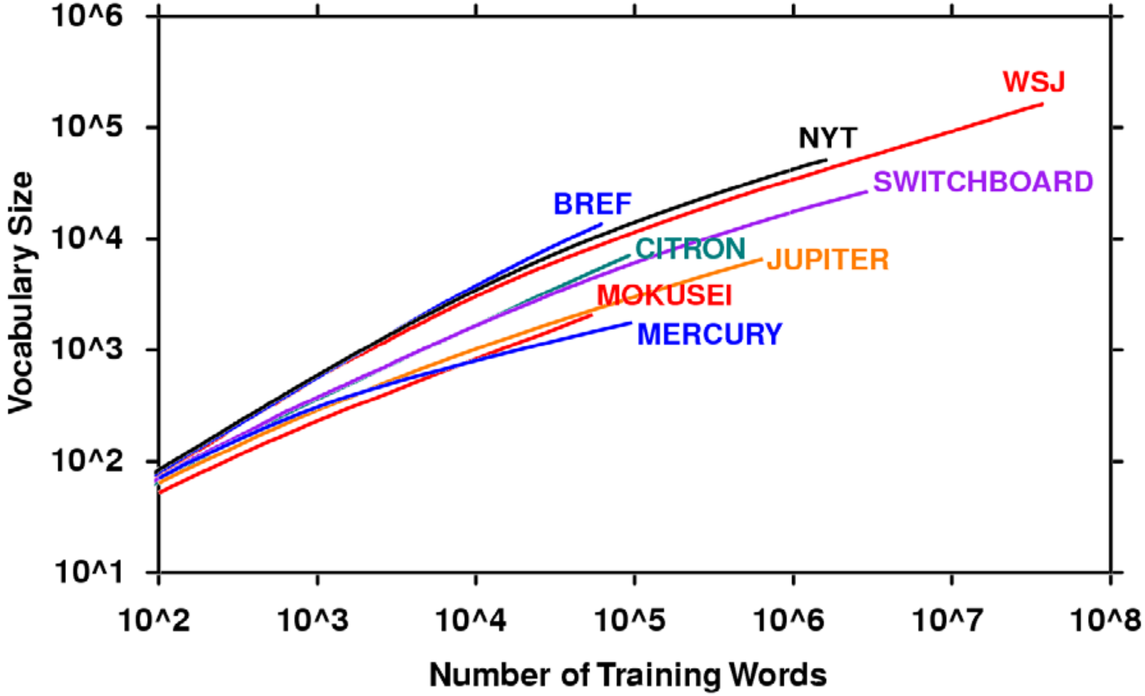
- **Introduction**
- **Modelling out-of-vocabulary (OOV) words**
  - **Probabilistic formulation**
  - **Domain-independent methods**
  - **Learning OOV subword units**
  - **Multi-class OOV models**

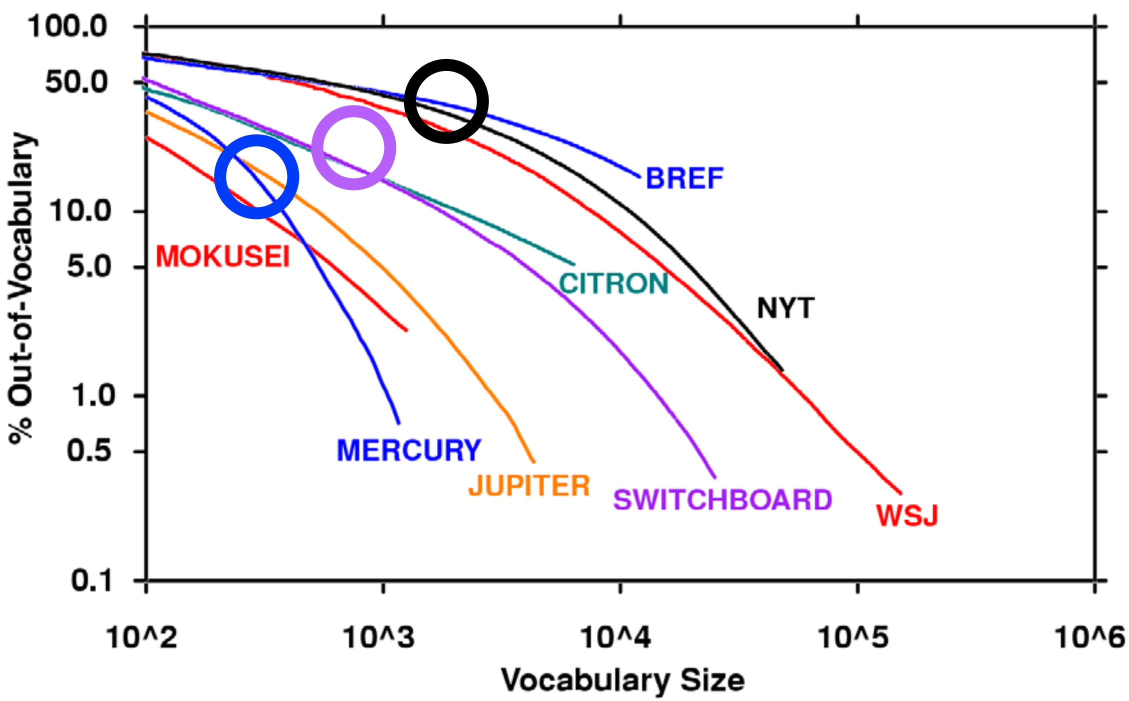
What is a new word?



- **Almost all speech recognizers search a finite lexicon**
  - **A word not contained in the lexicon is called out-of-vocabulary**
  - **Out-of-vocabulary (OOV) words are inevitable, and problematic!**

New Words are Inevitable!

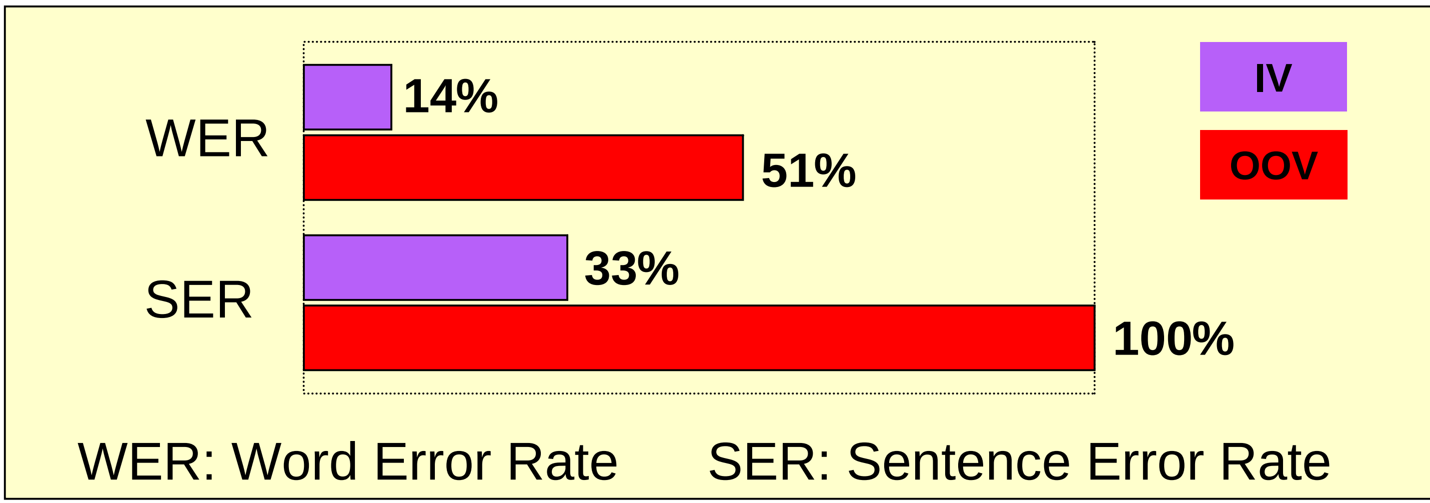




- **Analysis of multiple speech and text corpora**
  - **Vocabulary size vs. amount of training data**
  - **Out-of-vocabulary rate vs. vocabulary size**
- **Vocabulary growth appears unbounded**
  - **New words are constantly appearing**
  - **Growth appears to be language independent**
- **Out-of-vocabulary rate a function of data type**
  - **Human-machine speech**
  - **Human-human speech**
  - **Newspaper text**

New Words Cause Errors!

- **Out-of-vocabulary (OOV) words have higher word and sentence error rates compared to in-vocabulary (IV) words**



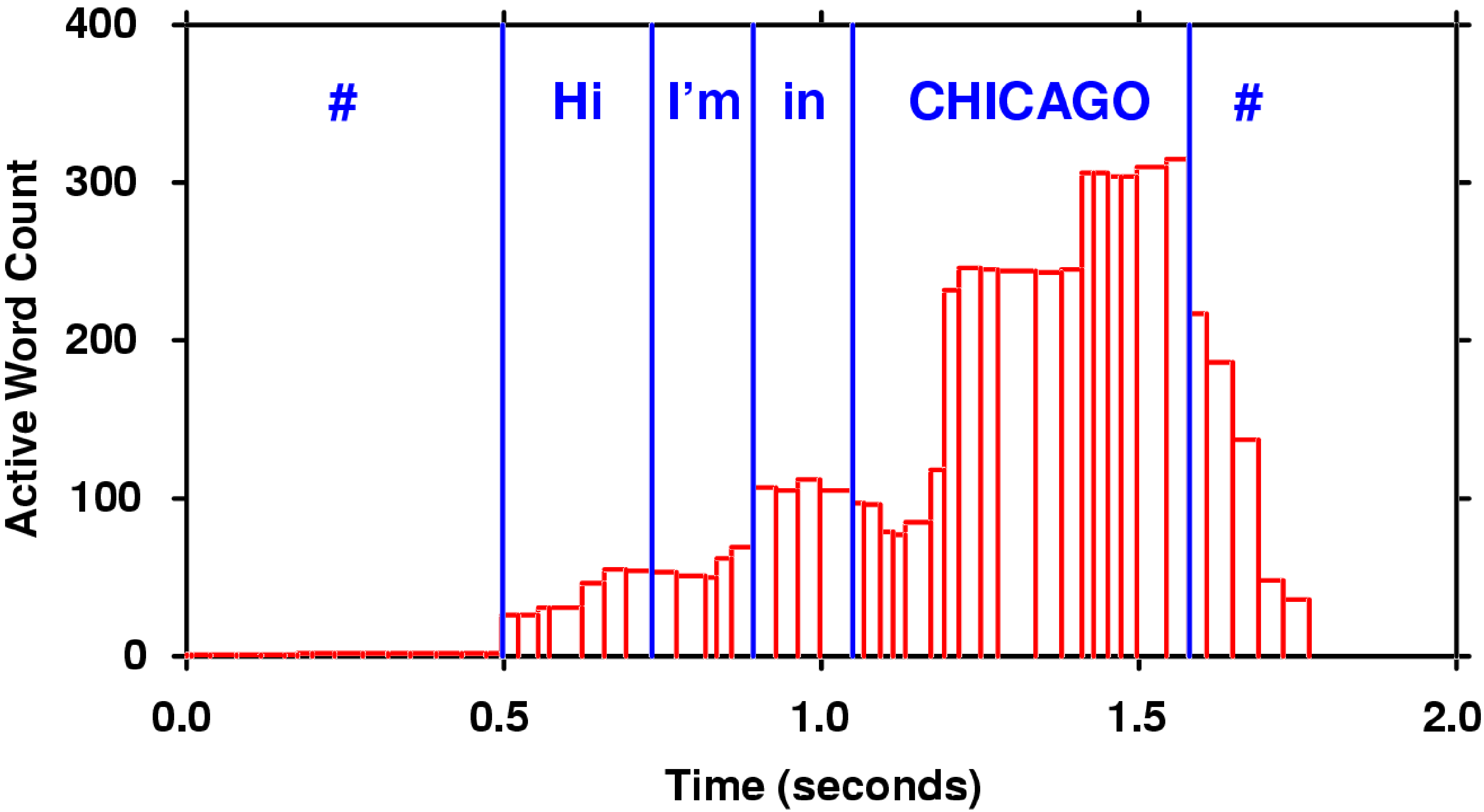
- **OOV words often cause multiple errors, e.g., "Symphony"**

  Ref: "Members of Charleston *Symphony* Orchestra are being treated…"

  Hyp: "Members of Charleston *simple your stroke* are being treated…"

New Words Stress Recognizers!

- **Search computation increases near presence of new words**



New Words are Important!

- **New words are often important content words**

Weather **Broadcast News**

- **Content words are more likely to be re-used (i.e., persistent)**

New Word Challenges

- Four challenges with new words:
  1) Detecting the presence of the word
  2) Determining its location within the utterance
  3) Recognizing the underlying phonetic sequence
  4) Identifying the spelling of the word

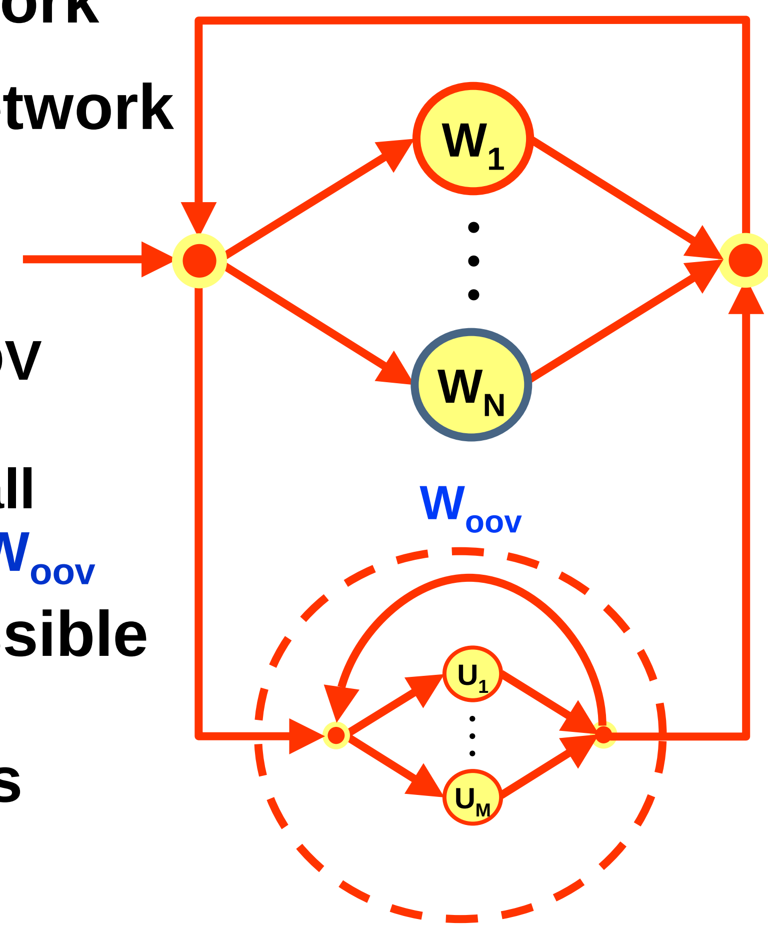
- Applications for new word models:
  – Improving recognition, detecting recognition errors
  – Handling partial words
  – Enhancing dialog strategies
  – Dynamically incorporating new words into vocabulary

Approaches to OOV Modelling

- **Increase vocabulary size!**
- **Use confidence scoring to detect OOV words**
- **Use subword units in the first stage of a two-stage system**
- **Incorporate an unknown word model into a speech recognizer**
  - **An extension of a filler, or garbage, model for non-words**

Incorporating an OOV Model into ASR
(Bazzi, 2002)

- **Hybrid search space: a union of IV and OOV search spaces**
  1) **Start with standard lexical network**
  2) **Construct separate subword network**
  3) **Add subword network to word network as a new word, W_{oov}**
     - **Cost, C_{oov}, is added to control OOV detection rate**
     - **During language model training, all OOV words are mapped to label W_{oov}**
- **A variety of subword units are possible (e.g., phones, syllables, …)**
- **A variety of topological constraints**
  - **Acoustic-phonetic constraints**
  - **Duration constraints**
  - **…**



The OOV Probability Model

- **The standard probability model:**

$$W^* = \arg\max_{W} \; P(A \mid W) \; P(W)$$

- **Acoustic models: same for IV and OOV words**
- **Language models: a class *n*-gram is used for OOV words**

An OOV word is hypothesized if:

$$\exists \;\; \mathbf{p} = p_1 p_2 .. p_N, \;\; such \;\; that \;\; \forall w_i \in L$$

$$P(A \mid \mathbf{p}) \;\; P(\mathbf{p} \mid OOV) P(OOV) \;\; > \;\; P(A \mid w_i) \;\; P(w_i)$$

Acoustic Model Probability

Language Model Probability

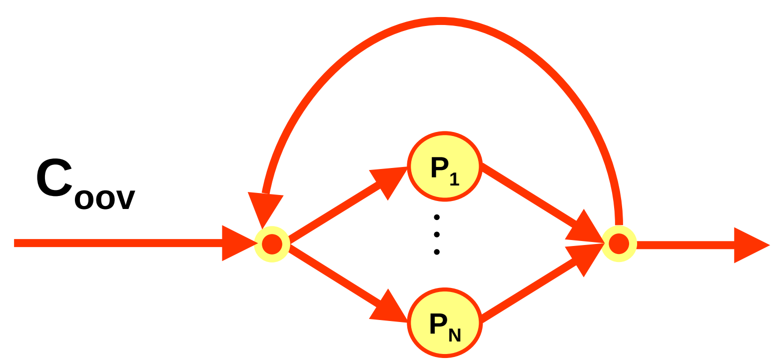
Advantages of the Integrated Approach

- **Compared to filler models**
  - **Same acoustic models for IV and OOV words**
    - * Probability estimates are comparable
  - **Subword language model**
    - * Estimated for the purpose of OOV word recognition
  - **Word-level language model predicting the OOV word**
  - **Use of large subword units**
  - **All of the above within a single framework**

- **The best of both worlds: fillers and two-stage**
  - **Early utilization of lexical knowledge (fillers)**
  - **Detailed sublexical modelling (two-stage)**

A Corpus-Based OOV Model

- **The corpus-based OOV model uses a typical phone recognition configuration**
  - **Any phone sequence of any length is allowed**
  - **During recognition, phone sequences are constrained by a phone *n*-gram**
  - **The phone *n*-gram is estimated from the same training corpus used to train the word recognizer**

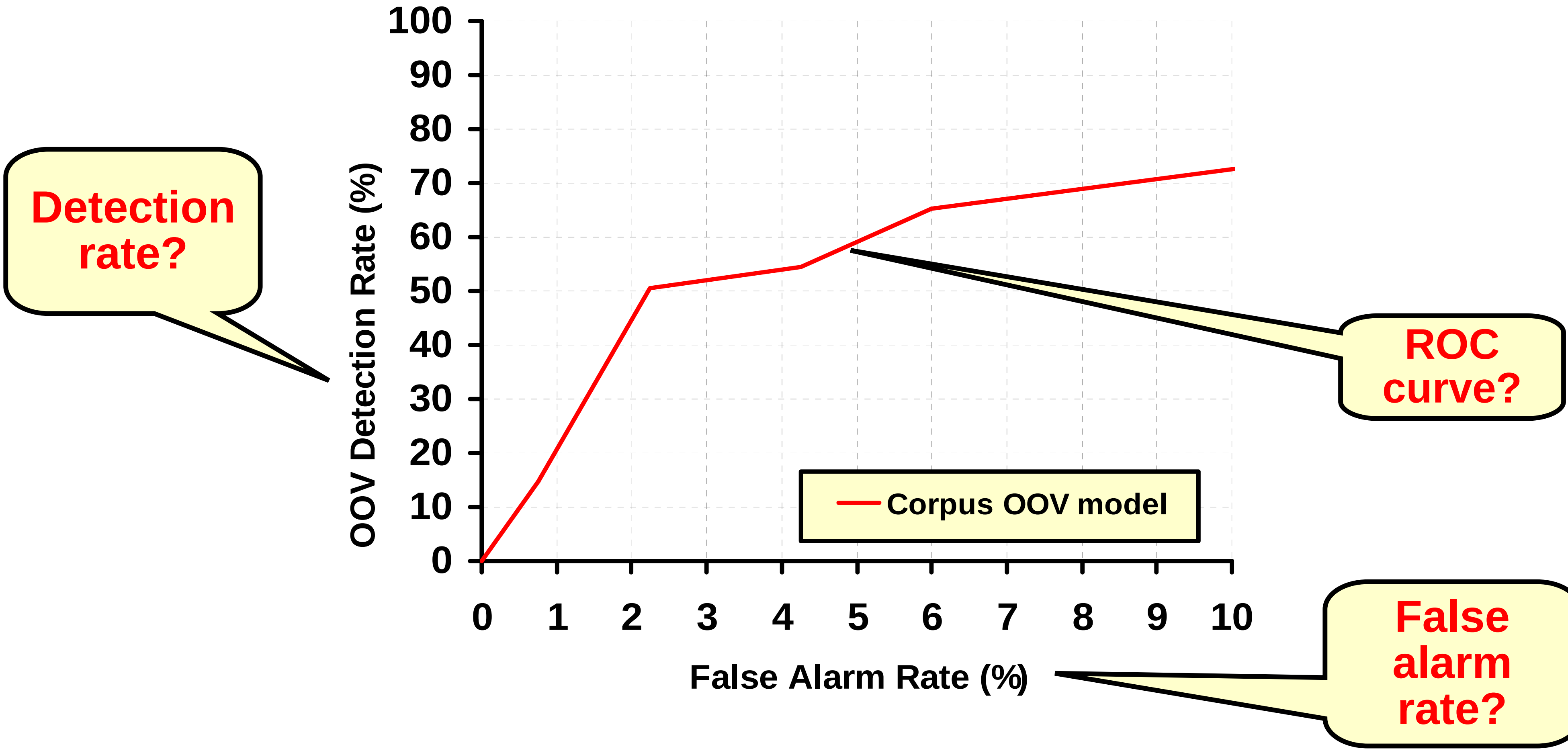


Experimental Setup

- **Experiments use recognizer from the JUPITER weather information system**
  - **SUMMIT segment-based recognizer**
  - **Context-dependent diphone models**
  - **88,755 utterances of training data**
  - **2,009 words in recognizer vocabulary**
  - **OOV rate: 2.2% (15.5% utterance-level)**
  - **OOV model uses a phone bigram**

- **Experiments use 2,029 test utterances from calls to JUPITER**
  - **1,715 utterances with only IV words**
  - **314 utterances contain OOV words**

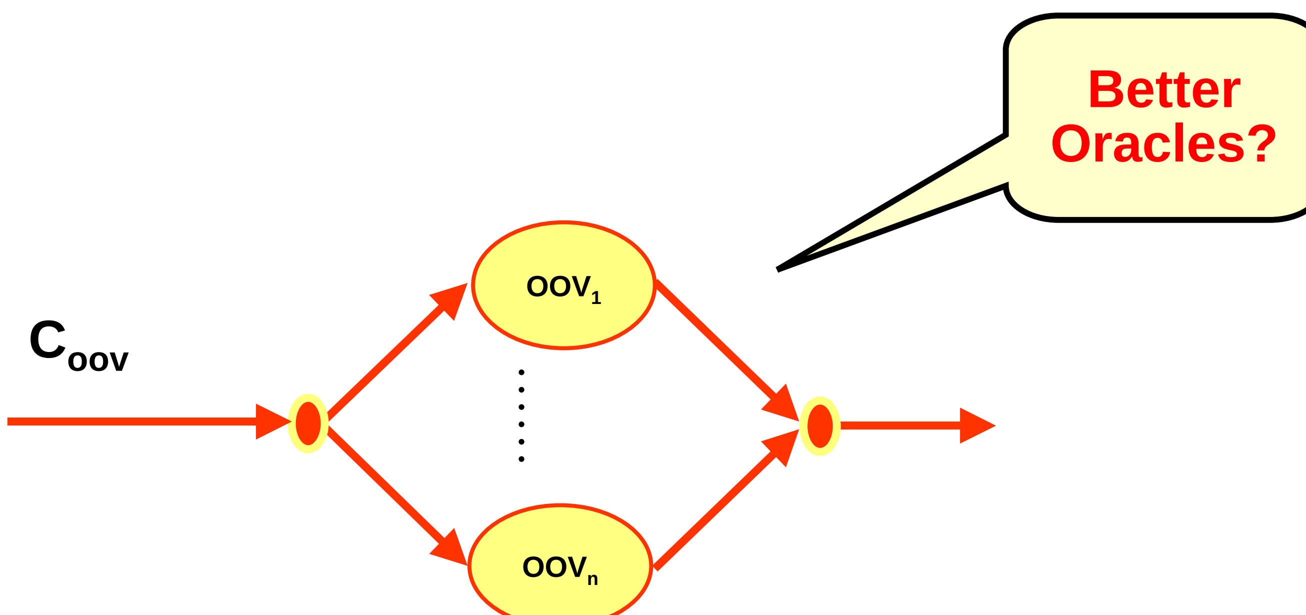
Corpus Model OOV Detection Results



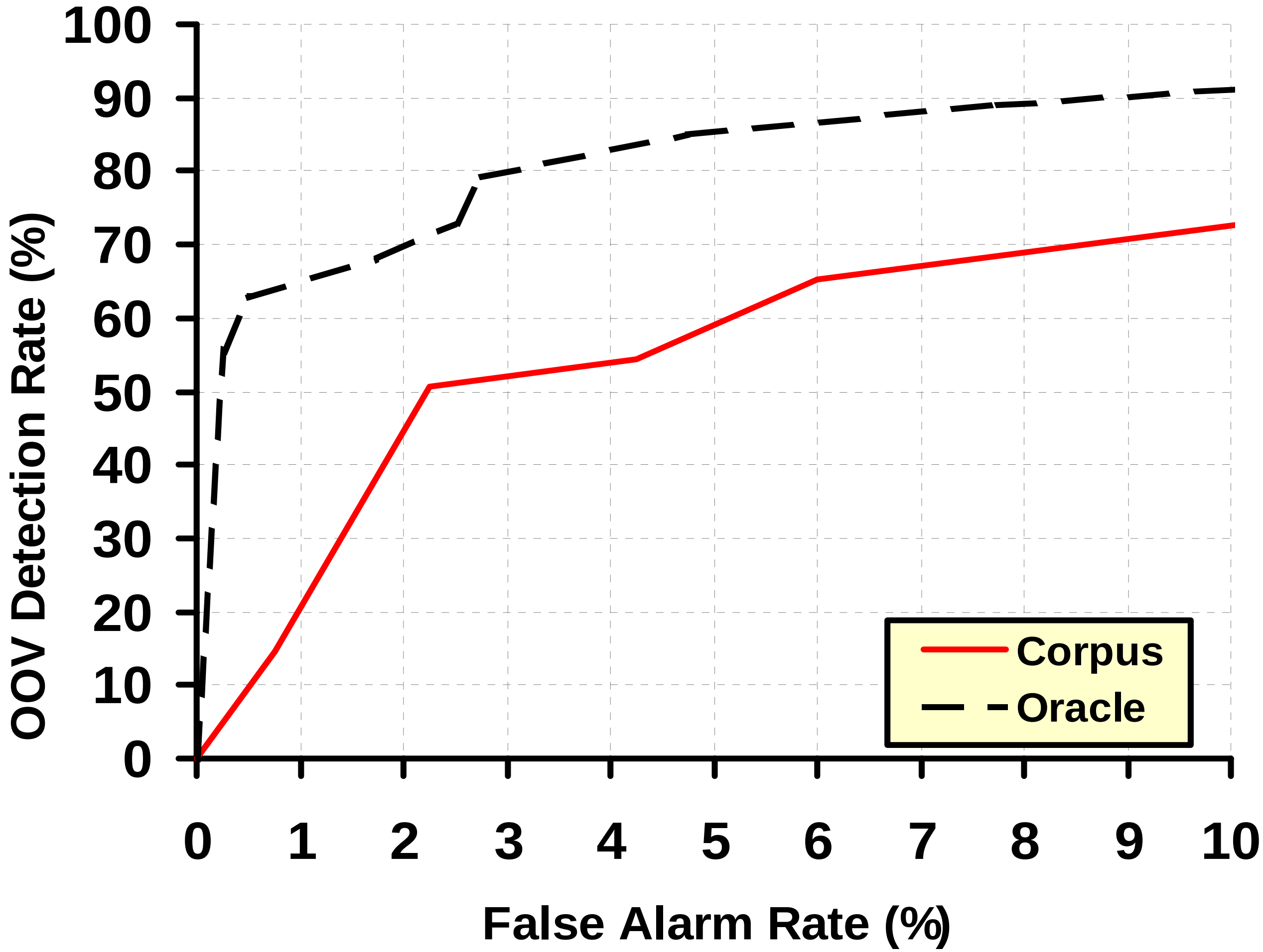
- **Half of the OOV words detected with 2% false alarm**
- **At 70% detection rate, false alarm is 8.5%**

The Oracle OOV Model

- **Goal:** quantify the best possible performance with the proposed framework
- **Approach:** build an OOV model that allows for only the phone sequences of OOV words in the test set
- **Oracle configuration is *not* equivalent to adding the OOV words to the vocabulary**



Oracle Model OOV Detection Results

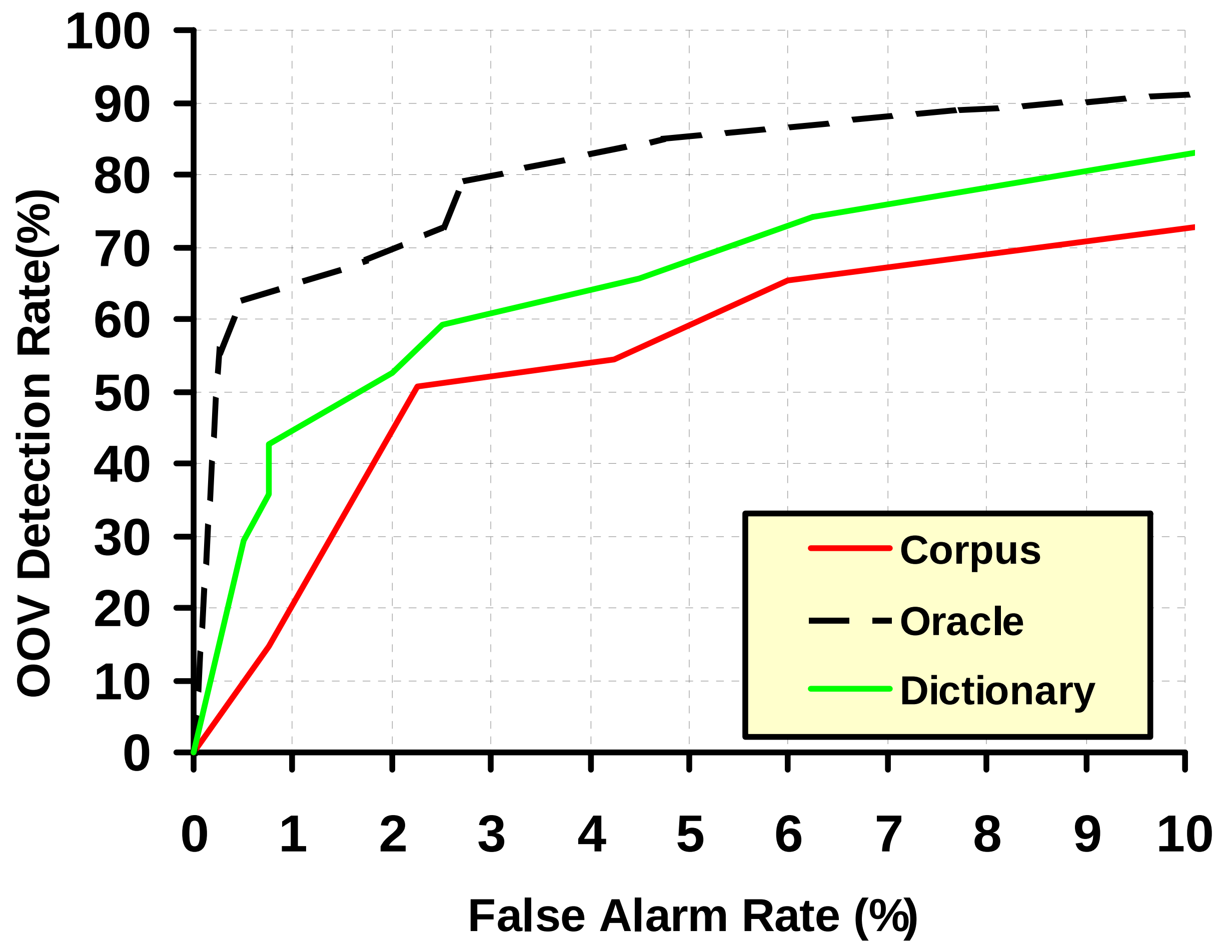


Significant room for improvement!

A Domain-Independent OOV Model

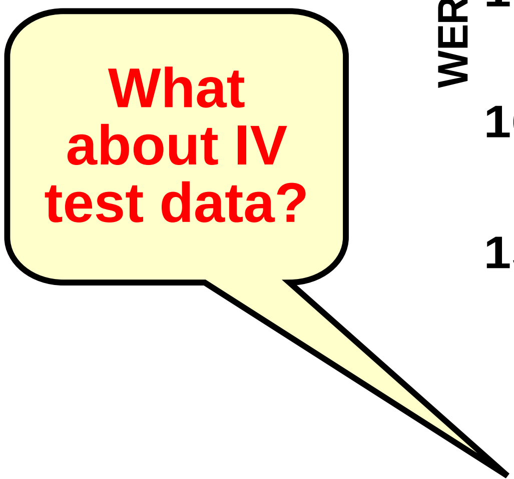
- **Drawbacks of the corpus model**
  - **Favors more frequent words since it is trained on phonetic transcriptions of complete utterances**
  - **Devotes a portion of the *n*-gram probability mass to cross-word sequences**
  - **Domain-dependent OOV model might not generalize**
- **A dictionary OOV model is built from a generic word dictionary instead of a corpus of utterances**
  - **Eliminates domain dependence and bias to frequent words**
- **Experiments use LDC PRONLEX Dictionary**
  - **90,694 words with a total of 99,202 pronunciations**

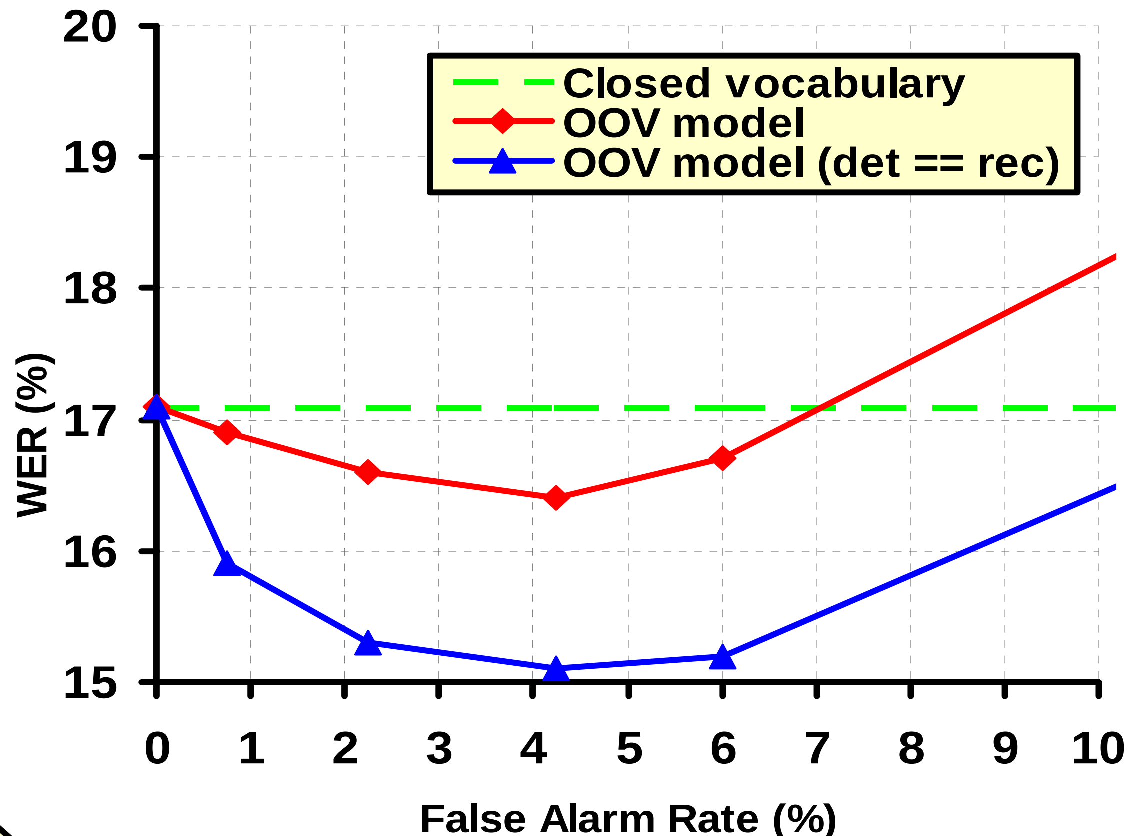
Dictionary Model OOV Detection Results



At 70% detection rate, false alarm rate is reduced from 8.5% to 5.3%

Impact on Word Error Rate

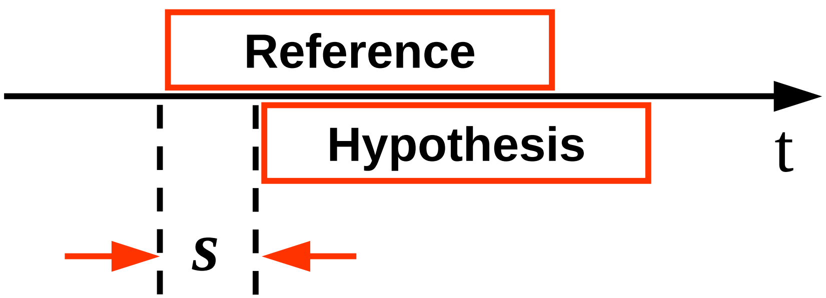


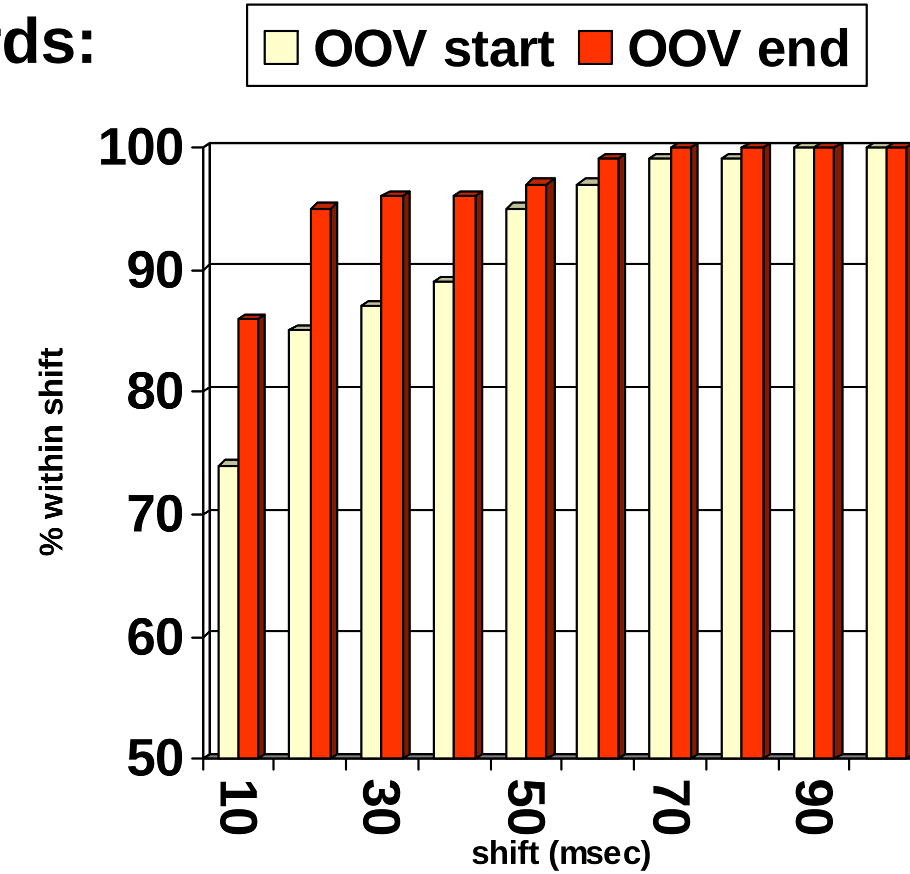


- WER on entire test set is reduced from 17.1% to 16.4%
- WER can be reduced from 17.1% to 15.1% with an identification mechanism

Other Performance Measures

- **Accuracy in locating OOV words:**





- **OOV phonetic error rate (PER):**

PER	Substitutions	Insertions	Deletions
37.8%	18.9%	6.0%	12.9%

Is that any good?

Learning OOV Sub-Word Units

- **Goal:** incorporate additional structural constraints to reduce false hypothesis of OOV words
- **Idea:** restrict the OOV network recognition to specific multi-phone units

How do we obtain the set of multi-phone units?

- **A data-driven approach:** measure phone co-occurrence statistics (e.g., mutual information) within a large dictionary to incrementally propose new multi-phone units

Learning Multi-Phone Units

- **An iterative bottom-up algorithm**
  - **Starts with individual phones**
  - **Iteratively merges unit pairs to form longer units**
- **Criterion for merging unit pairs is based on the weighted mutual information (MI_w) of a pair:**

$$MI_w(u_1,u_2) = p(u_1,u_2)\log\frac{p(u_1,u_2)}{p(u_1)p(u_2)}$$

- **At each iteration, the *n* pairs with highest MI_w are merged**
- **The number of multi-phone units derived depends on the number of iterations**
- **One byproduct is a complete parse of all words in the vocabulary in terms of the learned units**

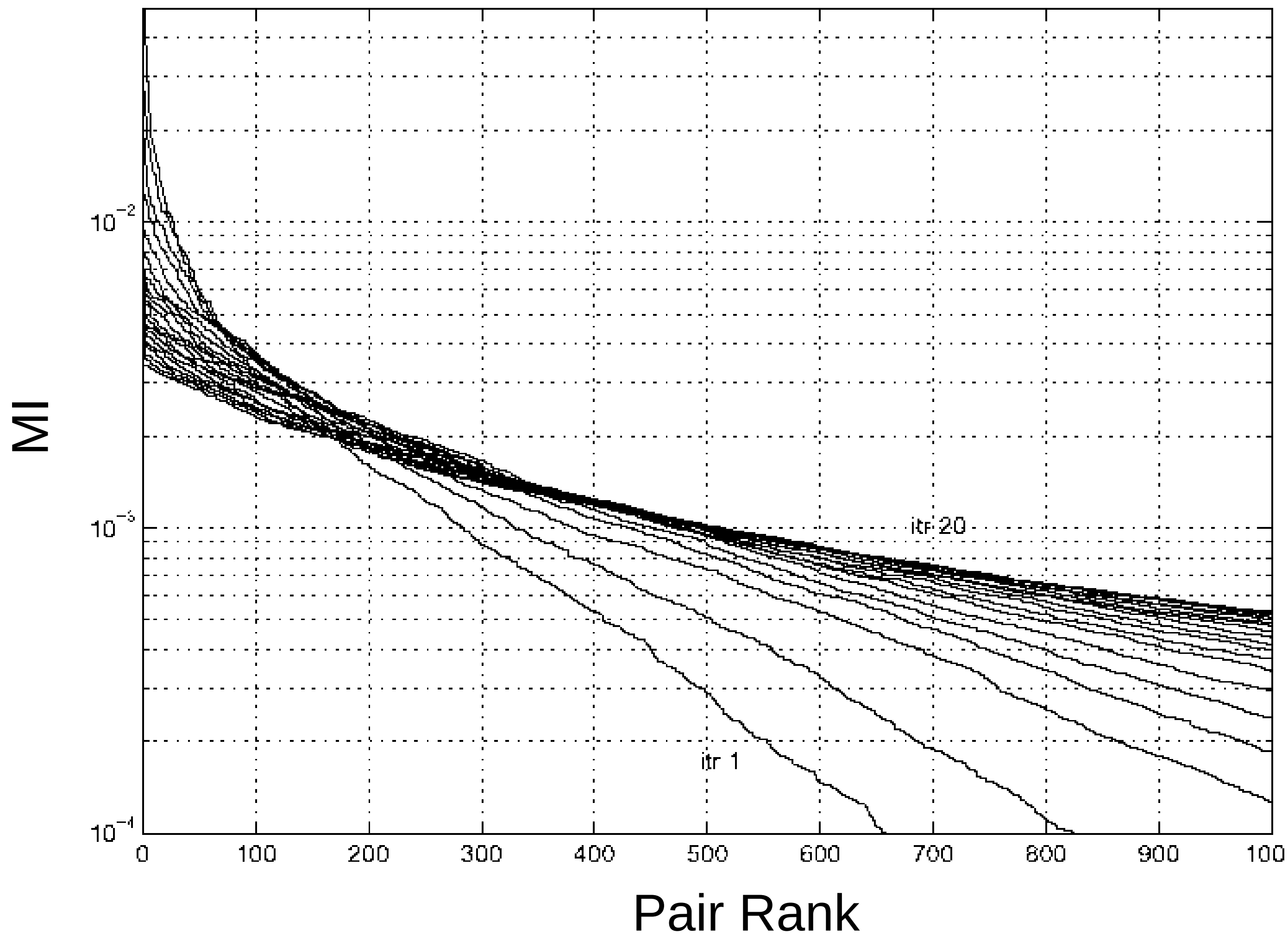
MMI Results

- **Initial set of units is the phone set (62 phones)**
- **Final unit inventory size is 1,977 units (after 200 iterations, and 10 merges per iteration)**
- **OOV model perplexity decreases from 14.0 for the initial phone set to 7.1 for the derived multi-phone set**
- **67% of derived units are legal English syllables**
- **Average length of a derived unit is 3.2 phones**

- **Examples:**

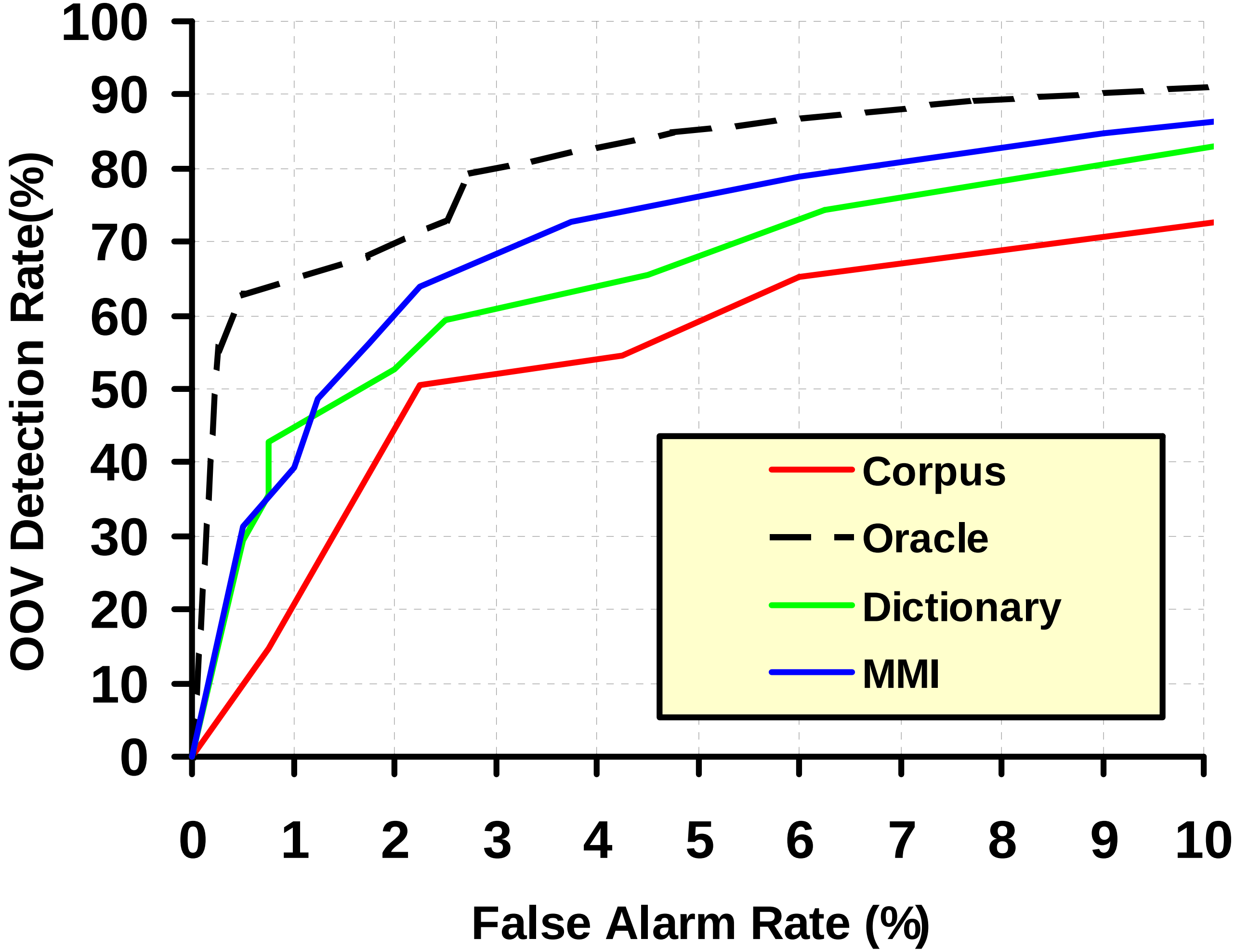
Word	Pronunciation
whisperers	(w_ih) (s) (p_ax_r) (axr_z)
yugoslavian	(y_uw) (g_ow) (s_l_aa) (v_iy) (ax_n)
shortage	(sh_ao_r) (tf_ax) (jh)

MMI Clustering Behavior



MI levels off for top ranking pairs; after several iterations (can be useful as a stopping criterion)

MMI Model OOV Detection Results



- **At 70% detection rate, false alarm rate is reduced to 3.2%**
- **Phonetic error rate is reduced from 37.8% to 31.2%**

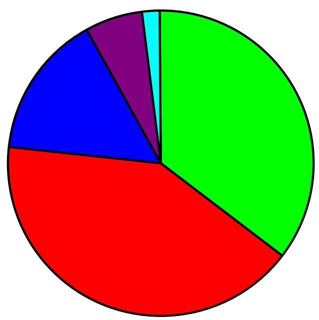
OOV Detection Figure of Merit

- Figure of merit (FOM) measures the area under the first 10% and the full 100% of the ROC curve
- The random FOM shows performance for a randomly guessing OOV model (ROC is the diagonal $y=x$)

OOV Model	100% FOM	10% FOM
Corpus	0.89	0.54
Dictionary	0.93	0.64
MMI	0.95	0.70
Oracle	0.97	0.80
Random	0.50	0.10

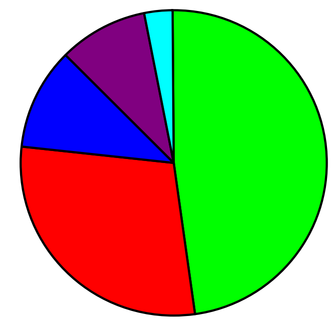
A Multi-Class OOV Model

- **Motivation:** **finer modelling of unknown word classes**
  - **At the phonetic level: similar phonotactic structure**
  - **At the language model level: similar linguistic usage patterns**



Weather

NAME
NOUN
VERB
ADJECTIVE
ADVERB



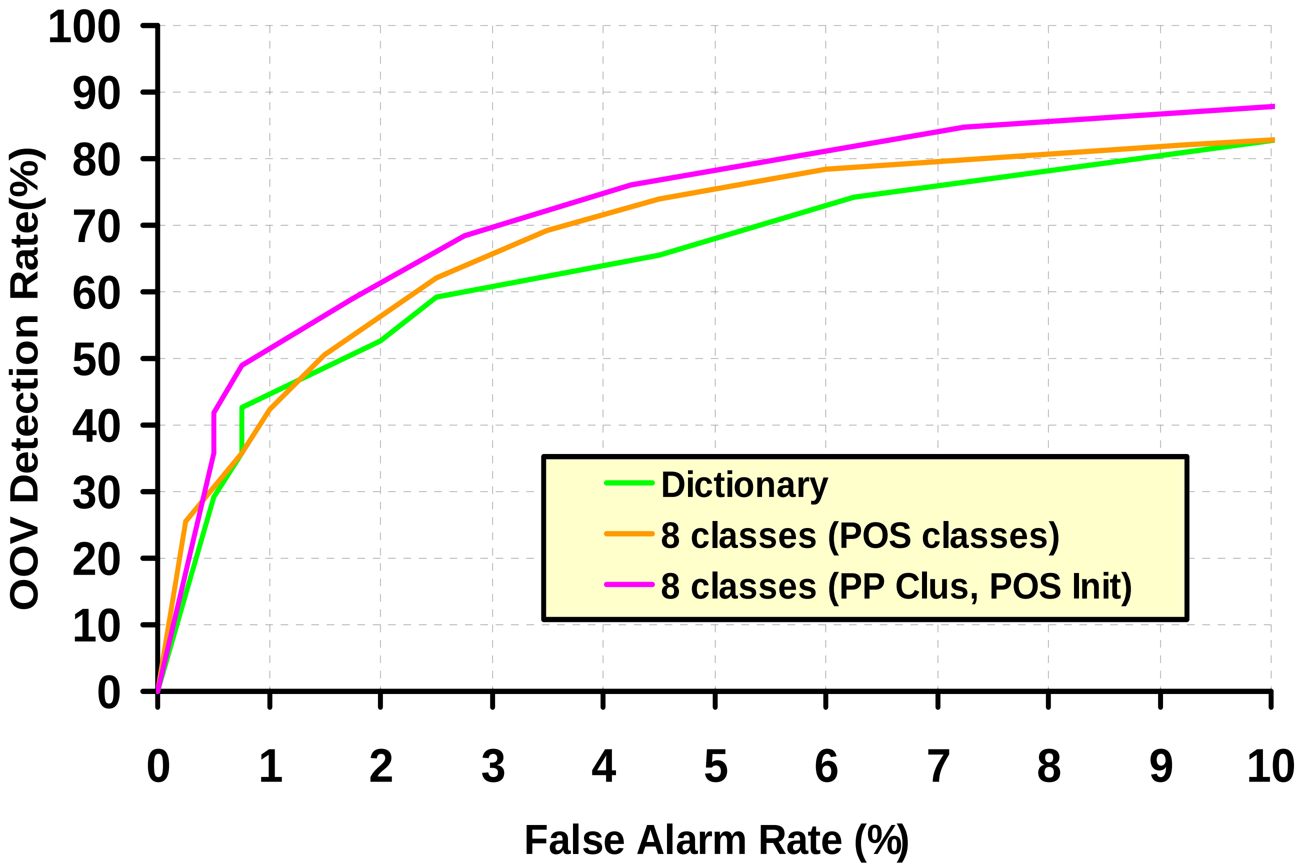
Broadcast News

- **Approach:** **extend the OOV framework to model multiple categories of unknown words**
  - **A collection of OOV networks in parallel with IV network**
  - **Word-level grammar G_N predicts multiple OOV classes**

Multi-Class Experiments

- **Class assignments in terms of part-of-speech tags**
  - **Derived from a tagged dictionary of words (LDC COMLEX)**
  - **Word-level language model trained on eight POS classes**
  - **Multiple sub-word LMs used for the different POS classes**
- **Class assignments based on perplexity clustering**
  - **Create a phone bigram language model from initial clusters**
  - **Use K-means clustering to shift words from one cluster to another**
  - **On every iteration, each word is moved to the cluster with the lowest perplexity (highest likelihood)**

Multi-Class Model OOV Detection Results



- **Multi-class method improves upon dictionary OOV model**
- **POS model achieves 81% class identification accuracy**
- **Perplexity clustering performs better than POS classes**

Multi-OOV Language Model Contribution

Condition/FOM	G_1 *n*-gram	G_8 *n*-gram
1 OOV network	0.64	0.65
8 OOV networks	0.68	0.68

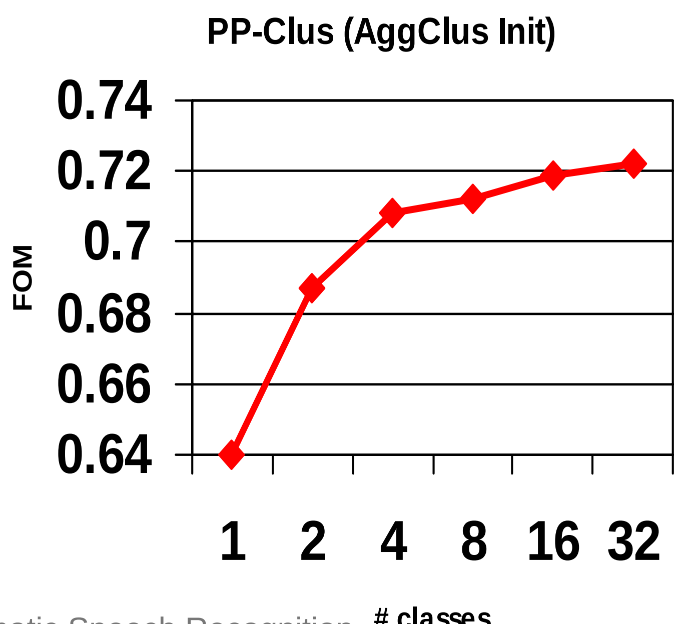
- **Most of the gain is from the multiple OOV networks**
  - **Phonotactics more important than language model constraints**
- **Behavior may be different for other domains**

Deriving Multi-Classes by Clustering

- **Clustering can be used to suggest initial multi-classes**
  - **Bottom-up clustering to initialize word class assignment**
  - **Distance metric based on the phone bigram similarity**
  - **An average similarity measure is used to merge clusters:**

$$d_{avg}(X_m, X_n) = \frac{1}{C_m C_n} \sum_{w_i \in X_m} \sum_{w_j \in X_n} d(w_i, w_j)$$

- **An arbitrary number of classes can be clustered**
- **Classes can be smoothed with perplexity clustering**



Model	Classes	10% FOM
Dictionary	1	0.64
POS Classes	8	0.68
PPClus (AggClus Init)	8	0.71
PPClus (POS Init)	8	0.72

Other Related Research Areas

- Measuring impact on OOV recognition to understanding
- Improving OOV phonetic accuracy
- Extending the approach to model out-of-domain utterances
- Developing OOV-specific confidence scores
  - To improve detection quality
- Modelling other kinds of out-of-domain sounds (e.g., noise)

References

A. Asadi, "Automatic detection and modeling of new words in a large vocabulary continuous speech recognition system," Ph.D. thesis, Northeastern University, 1991.

I. Bazzi, "Modelling out-of-vocabulary words for robust speech recognition," Ph.D. thesis, MIT, 2002.

G. Chung, "Towards multi-domain speech understanding with flexible and dynamic vocabulary," Ph.D. thesis, MIT, 2001.

L. Hetherington, "The problem of new, out-of-vocabulary words in spoken language systems," Ph.D. thesis, MIT, 1994.