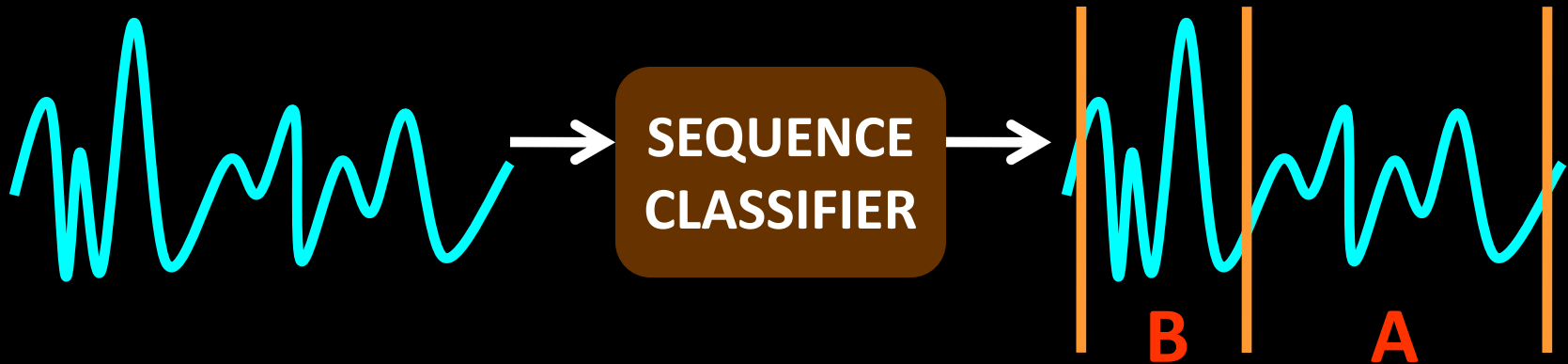# Diversity Benefits in Multi-System Sequence Classification:
Case Study of Automatic Speech Recognition (ASR)

Kartik Audhkhasi

Advisor: Prof. Shrikanth S. Narayanan

Signal Analysis and Interpretation Lab (SAIL)

What is Sequence Classification?

SEQUENCE CLASSIFIER

B   A

"Raw" Signal ——→ **"Useful" label sequence**

Speech   *Words*

Video, Motion capture   *Action events*

Biomedical sensors   *Medically-abnormal events*

Weather patterns   *Heat-waves*

Economic patterns   *Bearish/bullish trend*

...   ...

What Do "Textbook" ASR Systems Contain?

SPEECH SIGNAL → FEATURE EXTRACTOR (FE) / ACOUSTIC MODEL (AM) / LANGUAGE MODEL (LM) → GOOD BYE
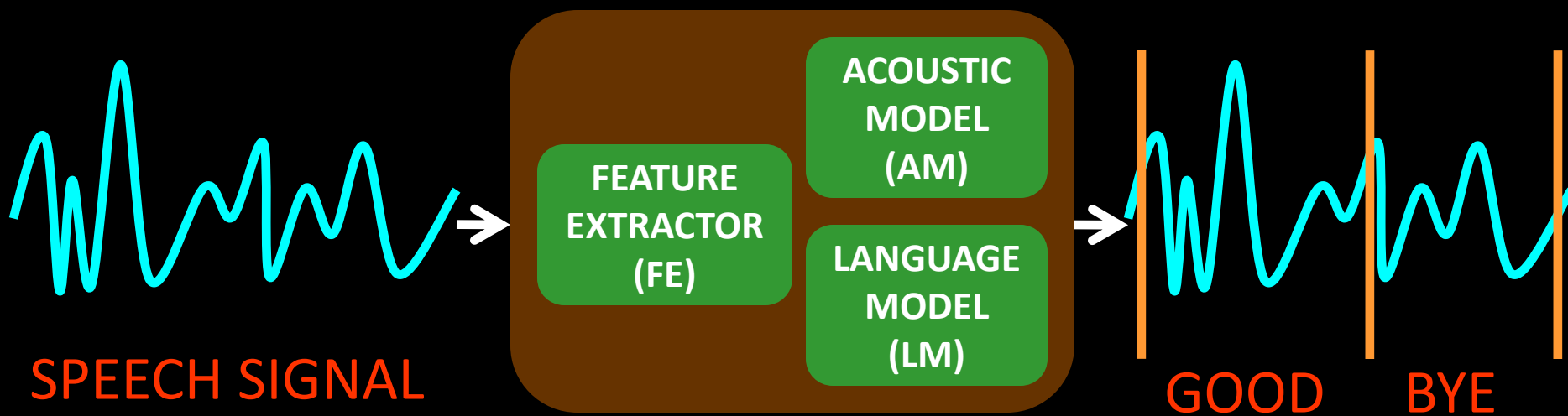
- FE – human speech signal processing.
- AM – statistics of speech features.
- LM – statistics of text.
  There are multiple models in literature for FE/AM/LM.

What do State-Of-The-Art ASR Systems Contain?
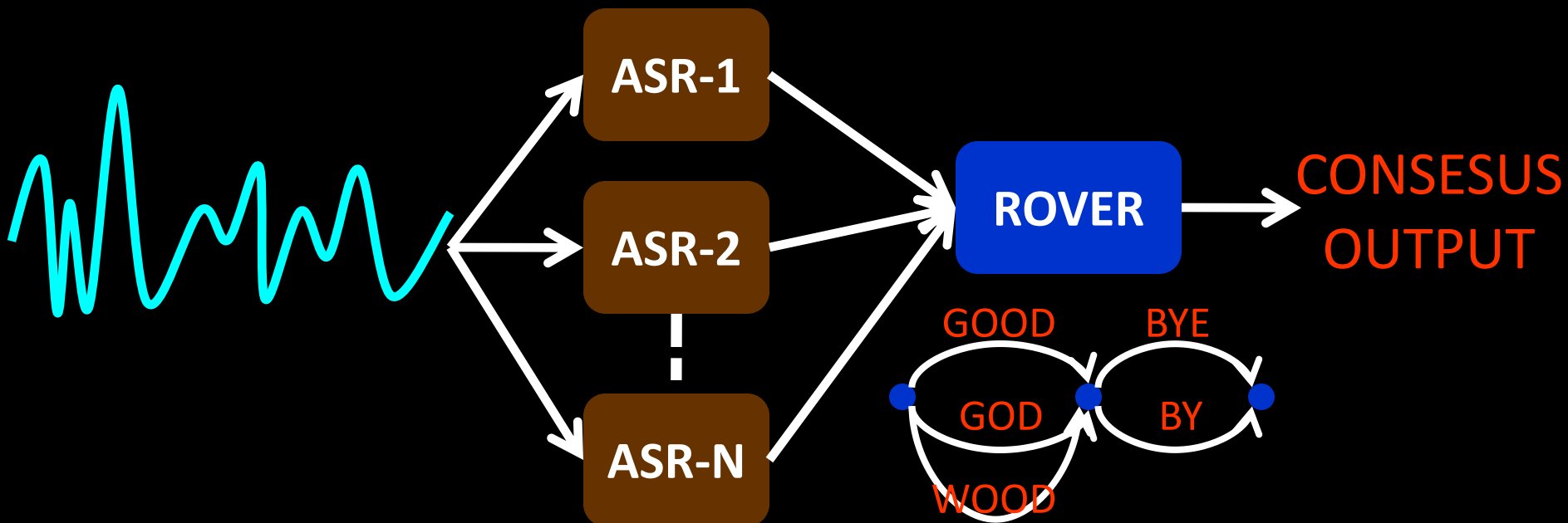
- Fusion of multiple diverse ASRs is performed in all large-scale ASR projects by major ASR research groups.

Design Strategy: Train diverse ASRs separately and hope fusion helps!

My Research Contributions

1) Defined ASR system diversity and proved its link with fusion performance as an upper-bound.

LEVENSHTEIN STRING

EDIT DISTANCE

$$E(\text{ref. sent.}, \text{fusion output}) \leq \frac{1}{N} \sum_{i=1}^{N} E(\text{ref. sent.}, \text{hyp. sent. } i)$$

$$-\frac{1}{N} \sum_{i=1}^{N} E(\text{fusion output}, \text{hyp. sent. } i) + \text{constants}$$

DIVERSITY FOR N ASRs

AVG. PERFORMANCE FOR N ASRs

[1] *(Under revision)* **K. Audhkhasi**, A. Zavou, P. Georgiou, S. Narayanan, "Theoretical analysis of diversity in an ensemble of automatic speech recognition systems", IEEE Trans. On Audio, Speech, and Language Processing.
[2] **K. Audhkhasi**, A. Zavou, P. Georgiou, S. Narayanan, " Empirical link between hypothesis diversity and fusion performance in an ensemble of automatic speech recognition systems", Proc. Interspeech, 2013.

My Research Contributions

Approximation Accuracy: RMSE = 0.57, Correlation Coefficient = 0.99

Each point = fusion of different set of real ASRs
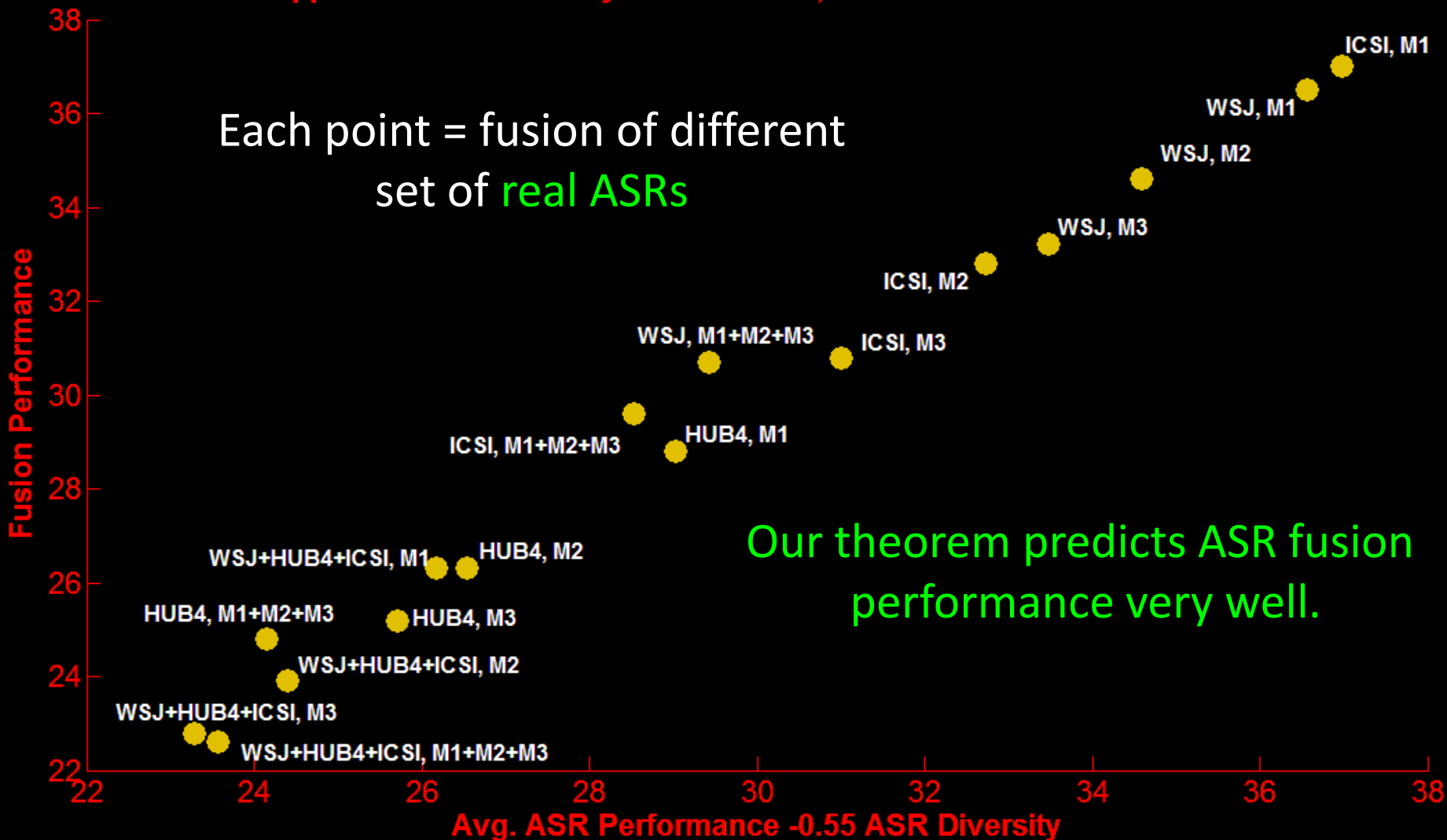
Our theorem predicts ASR fusion performance very well.

My Research Contributions

2) Recent approaches for training diverse ASR systems:

– Minimum Bayes Risk Leveraging (MBRL) [Breslin2006]

– Complementary Phone Error (CPE) Training [Diehl2012]

– Diverse Maximum Mutual Information (DMMI) Training [Tachioka2013]

are all special cases of:

$$\min\left\{\frac{1}{N}\sum_{i=1}^{N}E(\mathrm{ref.\,sent.,hyp.sent.}i)-\frac{\gamma}{N}\sum_{i=1}^{N}E(\mathrm{fusion\,output,hyp.sent.}i)\right\}$$

Minimizing upper-bound on fusion error rate leads to
Diverse Minimum Bayes Risk (DMBR) Training

My Thesis Research

"A Computational Framework for Ensembles of Diverse Experts"

Modeling
What is a realistic statistical model for a system of diverse human/machine experts?

Analysis
How does ensemble diversity affect fusion performance?

Design
How do we design ensembles of diverse experts?

Globally-Variant Locally-Constant (GVLC) Model

Generalized Ambiguity Decomposition (GAD)

Diverse MaxEnt models and ASR systems

Diversity Benefits in Multi-System Sequence Classification:
Case Study of Automatic Speech Recognition (ASR)

Kartik Audhkhasi

Advisor: Prof. Shrikanth S. Narayanan

Signal Analysis and Interpretation Lab (SAIL)