# REAL-TIME SPOKEN LANGUAGE IDENTIFICATION AND RECOGNITION FOR SPEECH-TO-SPEECH TRANSLATION

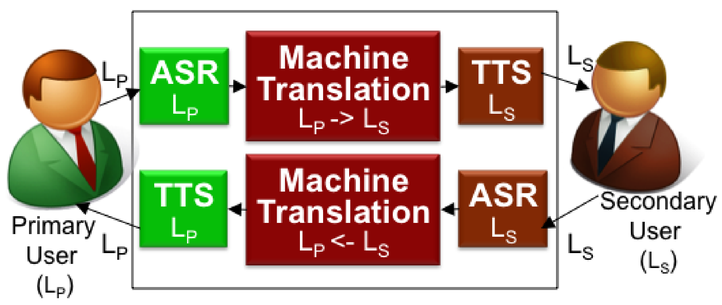
*Daniel Chung Yong Lim[1,2], Ian Lane[1] and Alex Waibel[1]*

*[1]Language Technologies Institute, Carnegie Mellon University, USA* *[2]DSO National Laboratories, Singapore*

## Language Identification for S2S Translation

- In current S2S translation systems input language must be selected before use
  - Single input microphone → input language must be switched for each turn in dialog
  - Multiple secondary languages ($L_S$ ) → must explicitly select $L_S$ before dialog begins

→ **Simplify user interface by detecting input language during translation**

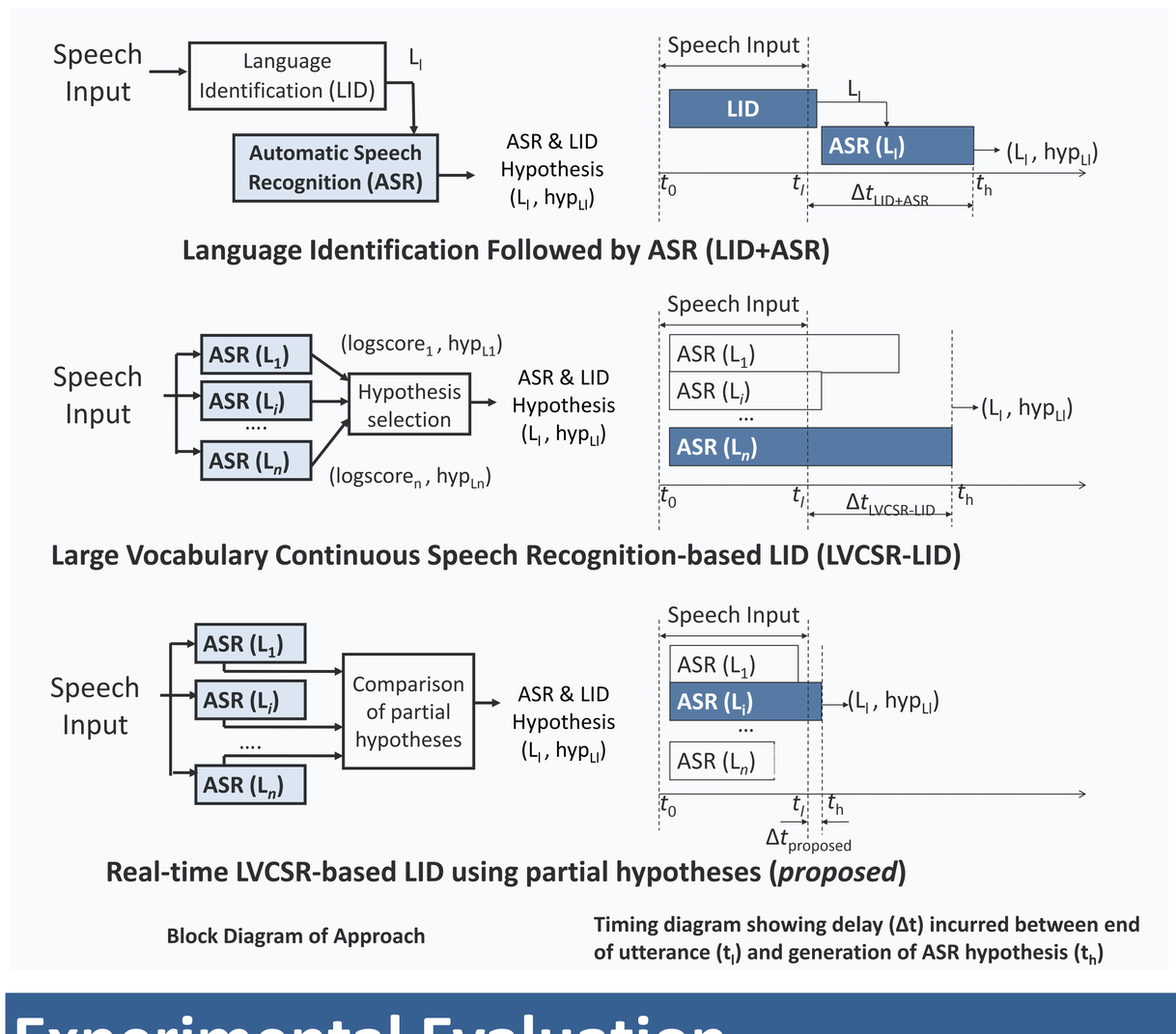




| Approach | Application |
|---|---|
| Automatically select $L_S$ based on secondary user's speech input | Portable S2S systems for locations where multiple languages used by local population ***e.g. in Afghanistan where both Dari and Pashtu are common (DARPA TransTAC)*** |
| | Locations where there are many foreign visitors **e.g. airport: single multilingual system which operates based on users speech input** |
| Automatically select translation direction using LID | Simplify interface for S2S translation systems **e.g. single microphone devices (mobile phones, desktop systems with single input mic)** |

**Can we accurately detect input language, perform speech recognition and select correct translation component with minimal impact to speed of end-to-end system?**

## Joint Language Identification and Recognition

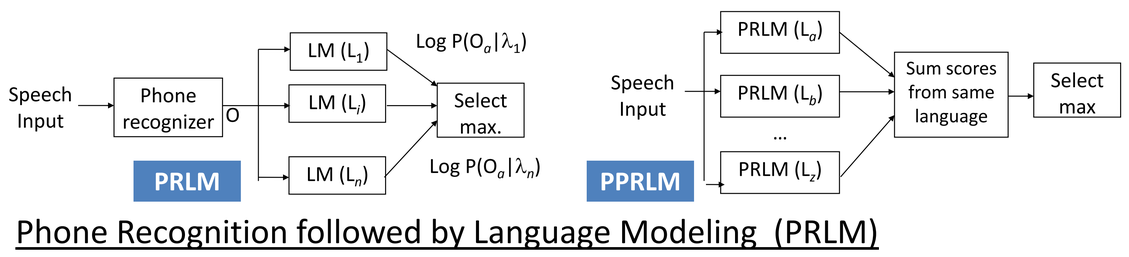
Require language identity and recognition hypothesis of input speech for downstream machine translation




**Block Diagram of Approach**

**Timing diagram showing delay (Δt) incurred between end of utterance ($t_I$) and generation of ASR hypothesis ($t_h$)**

### Approaches for Language Identification




Phone Recognition followed by Language Modeling (PRLM)

$$\hat{l} = argmax_{l \in \{L_1, \cdots, L_n\}} \log P(O|\lambda_l)$$

Parallel PRLM (PPRLM)

$$\hat{l} = argmax_{l \in \{L_1, \cdots, L_n\}} \sum \log P(O_i|\lambda_l)$$

Parallel PRLM+CRF [Lane&Lim '09]

- Apply discriminative CRF model using features extracted from PPRLM

LVCSR-based Language Identification

$$\hat{l} = argmax_{l \in \{L_1, \cdots, L_n\}} logscore_l$$

LVCSR-based Language Identification using partial hypotheses

- Compare partial hypotheses across languages (penalize languages where decoding is lagging). For language *l* the normalized partial hypothesis score is:

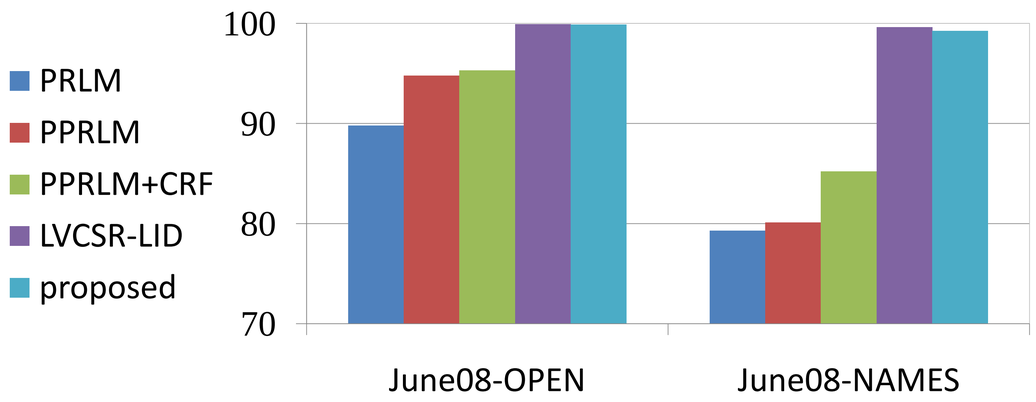
$$score_l = \frac{logscore_l}{logscore_{max}} \cdot \frac{t_{max}}{t_{max} - t_l} - w_{penalty}\left(\frac{t_{max} - t_l}{t_{max}}\right) \quad \text{where: } t_{max} = argmax_{l \in \{L_1, \cdots, L_n\}} t_l$$

$logscore_{max}$: log-scale ASR score for the system with $t_{max}$
$t_l$: amount of audio processed by language *l*'s decoder
$w_{penalty}$: time-lag penalty

Halt decoding for languages where $score_l < \theta$

## Experimental Evaluation

- Evaluate on CMU's TransTAC English-Iraqi S2S translation system
- Laptop-based single input microphone system: Discriminate between English and Iraqi




**Classification accuracies of evaluated LID approaches**

- LVCSR-based LID significantly more accurate than best phone-based approach
- LVCSR-based LID significantly slower than PRLM-based approaches
- *Proposed* approach offers similar accuracy in near realtime

| | June08-OPEN (E/I) | June08-NAMES (E/I) | Avg. |
|---|---|---|---|
| *LID+ASR* | 5038 / 7303 | 2017 / 4000 | 4590 |
| *LVCSR-LID* | 8099 /10511 | 2748 / 5033 | 6800 |
| *proposed* | 386 / 1575 | 493 / 507 | 740 |
| manual | 130 / 1317 | 290 / 346 | 520 |

**Average delay from end of speech until final recognition hypothesis (milliseconds)**

- Significant delay introduced by LID+ASR and LVCSR-based LID
- Proposed approach offers performance similar to manual language selection with minimal additional delay

**Conclusion:** proposed LVCSR-based LID approach offers selection accuracy similar to manual case with minimal additional delay introduced into end-to-end system

interACT International Center for Advanced Communication Technologies

Carnegie Mellon

lti