CONSOLIDATION BASED SPEECH TRANSLATION
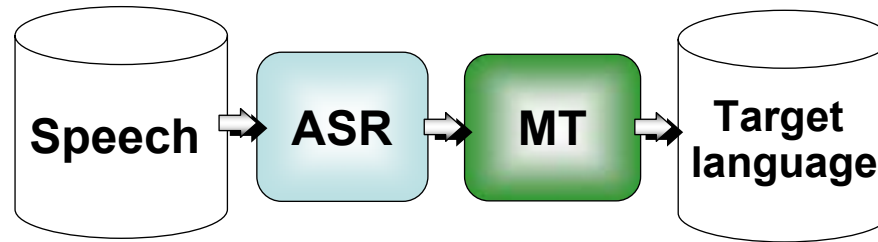
Chiori Hori†

Bing Zhao‡ Stephan Vogel‡   Alex Waibel‡

†NiCT-ATR

‡Carnegie Mellon University

The difficulties in speech translation

1. Problems in MT

   - Phrase table needs more constraints

        -> data sparseness
   - Global reordering model is missing

        MT(short) is better than MT(long)
   - OOV problem
   - Word segmentation
   - Difficulty in evaluation -> Reference + Metrics

2. Spoken language is different form written language.

- ungrammatical sentence

- colloquial experession

- Disfluency: filler, word fragment, repetitions, etc

Q: Does the same amount of speech transcription give the same quality of MT for written text?
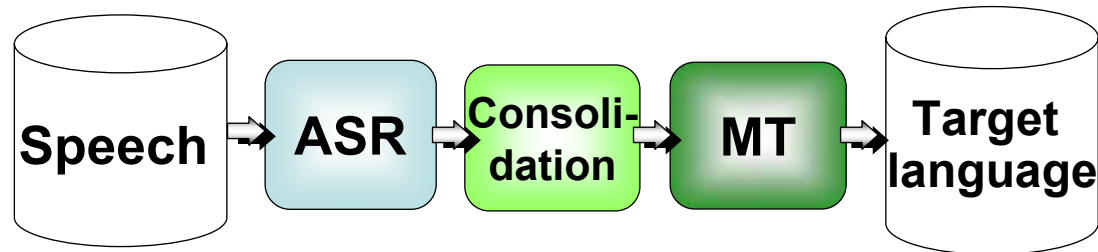
3. Tight-coupling of MT with ASR
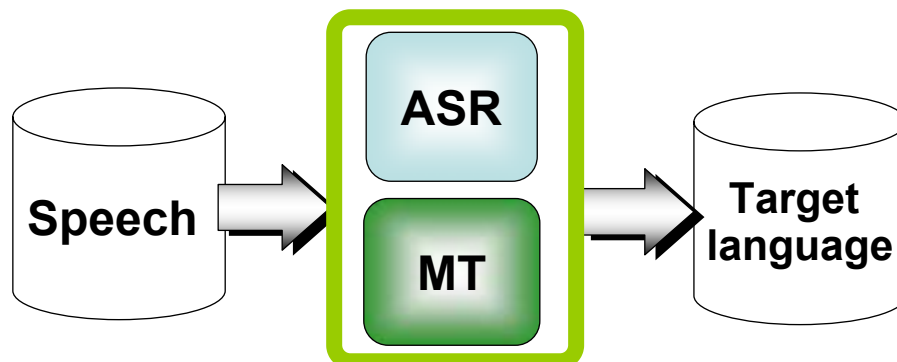
- Lattice/Confusion network translation
- Word segmentation

Q1. Does an oracle path always give the best MT?
Q2. Which aspect enhances the performance of MT in tight-coupling?
Q3. Are Linguistically better recognition results selected by MT?

4. ASR errors + Disfluency caused by spontaneity

$$\hat{T} = \arg\max_{T} P(T|O) = \arg\max_{T} \sum_{S_{CON}} \sum_{W} P(O|W, S_{CON}, T) P(W|S_{CON}, T) P(S|T) P(T)$$

$$\approx \arg\max_{T} \left[\max_{S_{CON}, W} P(O|W) P(W|S_{CON}) P(S_{CON}|T) P(T) \right]$$

Q1. Is full translation absolutely needed even when translatability is low?

Q2. Do we have to translate disfluency correctly especially for communication?

Evaluation Results for RT04
(Chinese BN to English text)

			#ST	#WD	PP
TIDES '02 BLEU=27.2 , NIST=8.7	Source		878	24337	138
	4 references		3512	105143	148
RT04 BLEU=8.2 , NIST=4.1	SRC	TRS	297	11547	536
		ASR [21.2%(c) 46.8%(w)]	297	9724	848
	1 reference		297	12105	300

Evaluation Results for RT04
(Chinese BN to English text)

Legend:
- ASR output
- CON_Confidence+sp+ip
- CON_Confidence+LM+sp+ip
- Gold standard for consolidation
- Manual Transcription

Categories: BLEU4 (One reference), BLEU4 (Multiple references), MPAccy