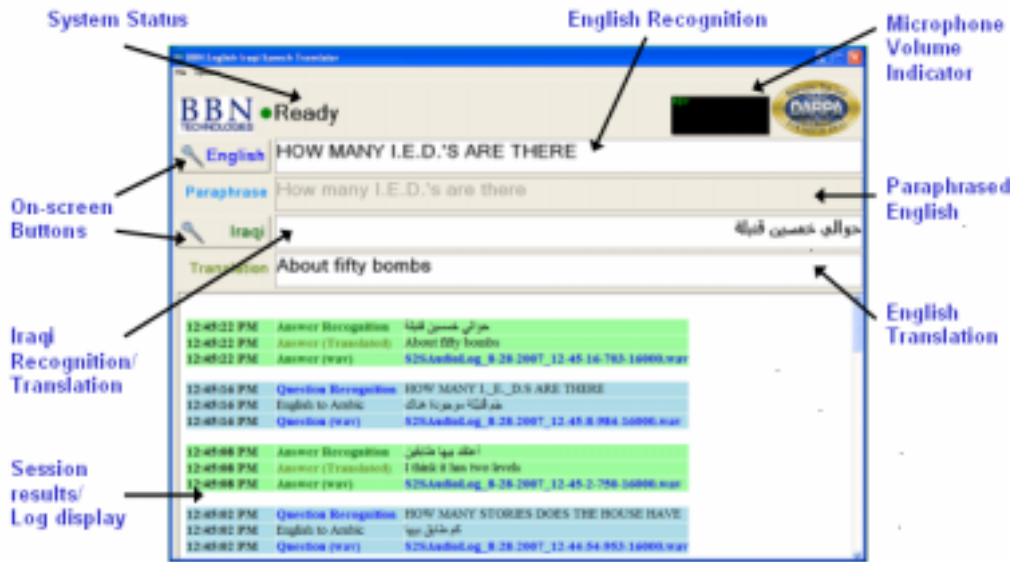# Challenges and Future Directions for Speech-to-Speech Translation

Rohit Prasad, Prem Natarajan, David Stallard, Fred Choi, Shirin Saleem, Chia-lin Kao, Krishna Subramanian, Kriste Krstovski

BBN Technologies, Cambridge, MA, USA

IEEE ASRU 2007 Panel "Problems and Future Directions of Speech Translation Technology"

BBN TECHNOLOGIES

BBN TransTalk™: Transcending the Language Barrier

- **Enable free form two-way spoken communication across a language barrier**

- **BBN TransTalk integrates the following components:**
  - **BBN Byblos speaker-independent, n-gram automatic speech recognition (ASR) engine**
  - **BBN's phrase-based statistical machine translation (MT) coupled with a "Question Canonicalizer"**
  - **Cepstral's Text-to-Speech engine**

- **Focus on English/Iraqi for force protection domain**
  - **Other languages: Farsi, Levantine, and Indonesian**

- **Being advanced under the DARPA TRANSTAC program**
  - **Participants: BBN, CMU, IBM, SRI, USC**

BBN TECHNOLOGIES

BBN TransTalk for Multiple Platforms

Laptop System

PDA System

Wearable Headset System

IEEE ASRU 2007 Panel "Problems and Future Directions of Speech Translation Technology"

BBN TECHNOLOGIES

Challenges for Speech-to-Speech (S2S) Translation

- **Algorithmic: Robust to changes in speakers, environment, and topics of the conversation**

- **Dialog Management: Prevent dialog from going off track when the system or the user makes errors**

- **Evaluation Metric: Devise a metric for measuring improvements and optimizing *end-to-end* performance**

- **Hardware: Portable, low power, rugged, affordable, etc.**

- **User Interface: Simple but effective eyes-free and hands-free application control**

Key Question: *What are severe errors made by our current S2S system?*

BBN
TECHNOLOGIES

Text-to-Text (T2T) Error Analysis for Iraqi to English (I2E)

- **Developed a novel methodology for assigning severity to an error category based on relative reduction in Likert Score it causes[1]**

Category	Count	Severity	%Total Damage
Wrong Concept	51	1.72	21.8
Word Sense	57	1.42	20.2
Missing Concept	43	1.65	17.6
Pronoun Error	63	0.85	13.4
Missing Non-Concept	34	0.82	6.9
Word Order	30	0.76	5.7
Polarity	8	1.99	4.0
Disfluency	7	1.43	2.5
Inserted Non-Concept	16	0.56	2.4

- **Most damaging: Wrong Concept, Word Sense, Missing Concept**

- **Polarity has highest weight, but causes small damage**

- **Pronoun Error is the most common**

[1]*D. Stallard, et al., "Recent Improvements and Performance Analysis of ASR and MT in a Speech-to-Speech Translation System," To Appear ICASSP 2008.*
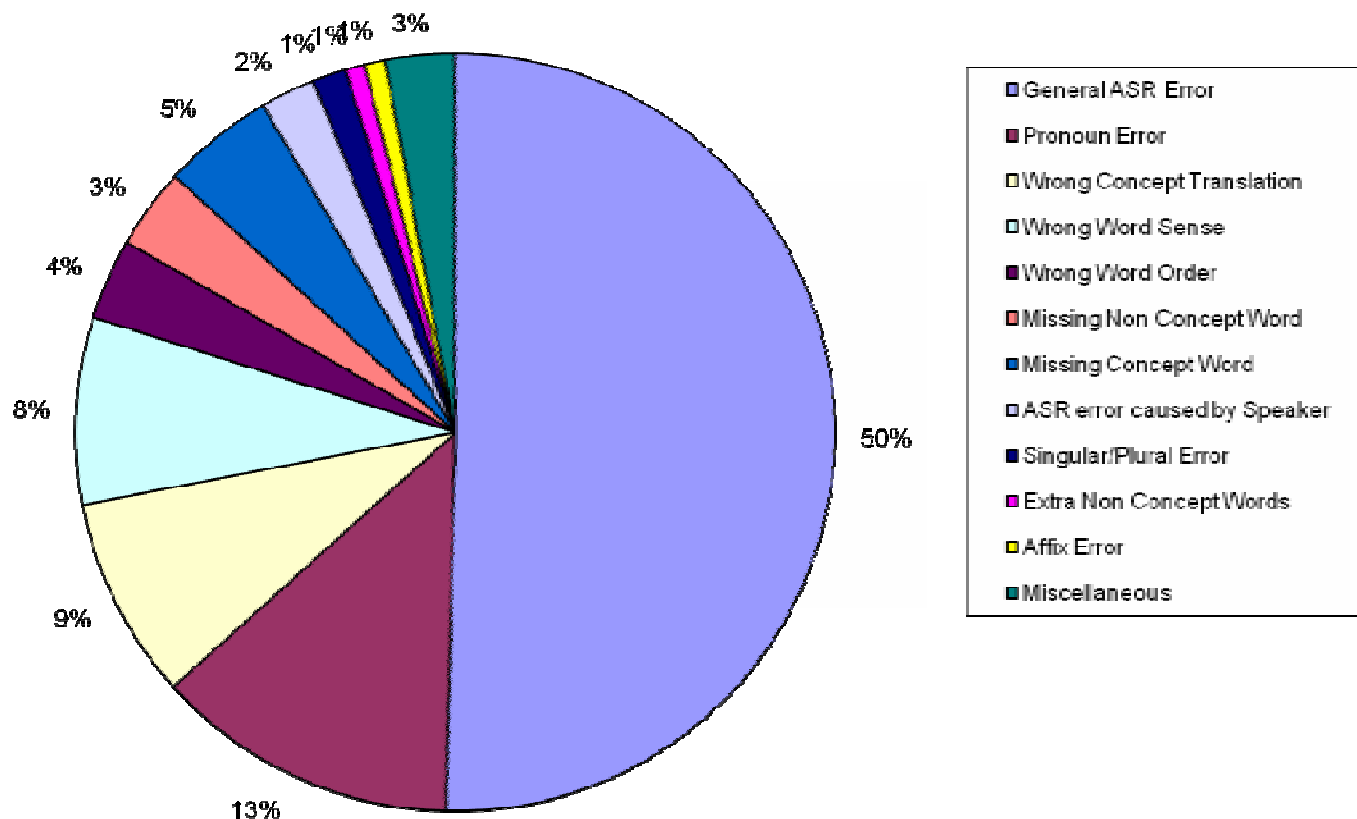
IEEE ASRU 2007 Panel "Problems and Future Directions of Speech Translation Technology"

BBN TECHNOLOGIES

T2T Analysis for English to Iraqi (E2I)

Category	Count	Severity	%Total Damage
Word Sense	39	**1.56**	**21.9**
Wrong Concept	33	**1.90**	**21.7**
Pronoun Error	59	0.94	**19.6**
Missing Concept	24	**1.82**	15.4
Missing Non-Concept	24	**1.07**	9.0
Word Order	20	0.58	4.1
Singular vs. Plural	13	0.70	3.2
Affix	8	0.99	2.8

- **The damaging errors are the intuitively "major" ones**
- **Pronoun Error is again the most common**
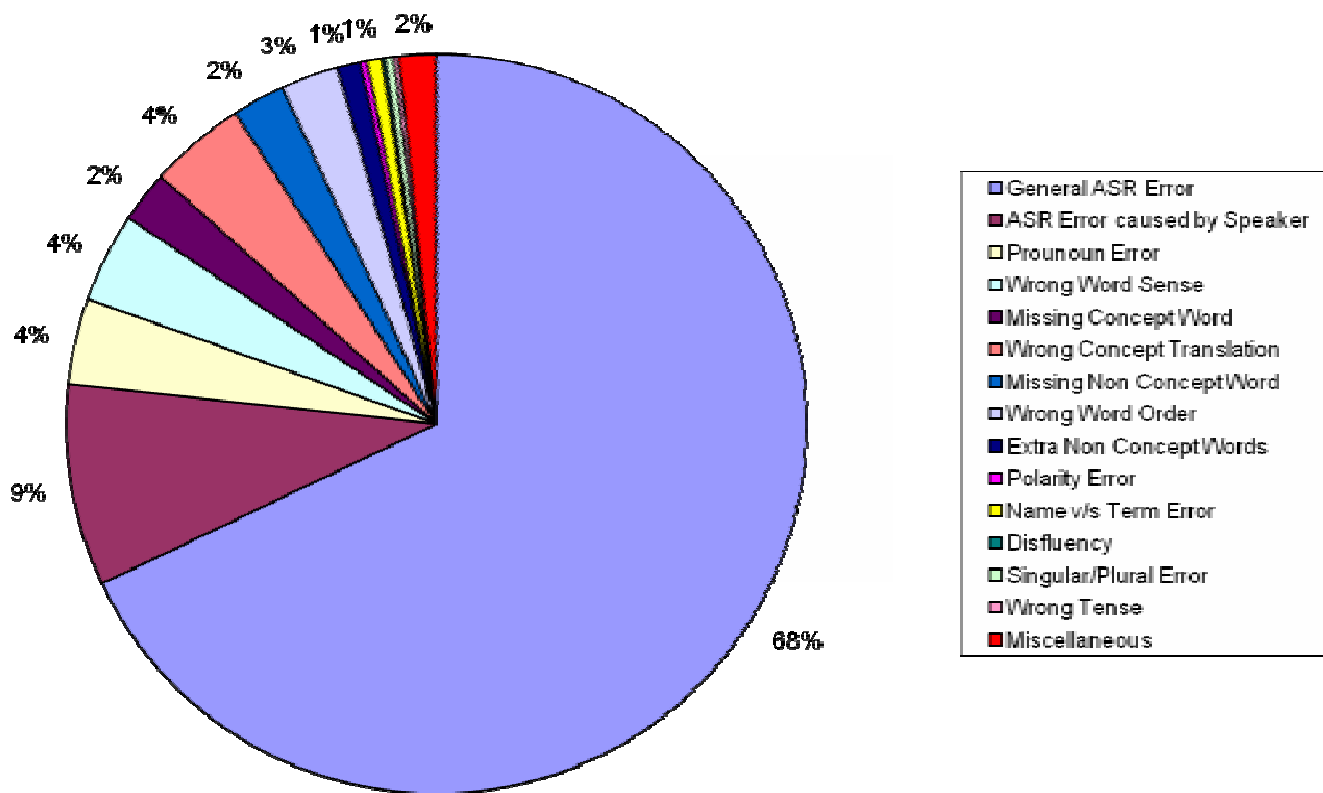- **Intuitively more "minor" errors are the least damaging**

BBN TECHNOLOGIES

Speech-to-Text (S2T) Error Dominated by ASR Error– E2I

Total Fractional Likert Error for Each Category on July 07 Offline Iraqi Set

Legend:
- General ASR Error
- Pronoun Error
- Wrong Concept Translation
- Wrong Word Sense
- Wrong Word Order
- Missing Non Concept Word
- Missing Concept Word
- ASR error caused by Speaker
- Singular/Plural Error
- Extra Non Concept Words
- Affix Error
- Miscellaneous

Pie chart values: 50%, 13%, 9%, 8%, 4%, 3%, 5%, 2%, 1%, 1%, 1%, 3%

IEEE ASRU 2007 Panel "Problems and Future Directions of Speech Translation Technology"

BBN TECHNOLOGIES

S2T Error Dominated by ASR Error – I2E

Total Fractional Likert Error for Each Category on July 07 Offline Iraqi Set

Legend:
- General ASR Error
- ASR Error caused by Speaker
- Prounoun Error
- Wrong Word Sense
- Missing Concept Word
- Wrong Concept Translation
- Missing Non Concept Word
- Wrong Word Order
- Extra Non Concept Words
- Polarity Error
- Name vs Term Error
- Disfluency
- Singular/Plural Error
- Wrong Tense
- Miscellaneous

IEEE ASRU 2007 Panel "Problems and Future Directions of Speech Translation Technology"

BBN TECHNOLOGIES

Future Directions for Algorithmic Improvements

- **Improve robustness to ASR errors**
  - Tighter coupling of ASR and MT beyond feed-forward 1-best integration
  - Noise-robust acoustic modeling
  - Improve adaptation to speaker, environment, topics

- **Incorporate context-awareness in ASR and MT**
  - Need to use context at multiple levels (utterance, dialog, etc.)
  - Major error categories such as Word Sense, Wrong Concept, and Pronoun Errors can be mitigated with use of context

- **Develop a methodology for optimizing end-to-end system performance**

BBN TECHNOLOGIES