

Toward **Universal Network-based** **Speech Translation**

Chai Wutiwiwatchai

Speech and Audio Technology Laboratory
National Electronics and Computer Technology Center

Outline

- **Technology Review**
- **U-STAR Consortium**
 - Brief History
 - Major Activities
- **U-STAR Speech Translation Service**
 - Service architecture
 - Service connection protocol
 - Resource and engine development
- **Evaluation and Issues**
 - Lab and field-testing evaluations
 - Major issues
- **Conclusion**

Outline

- **Technology Review**
- **U-STAR Consortium**
 - Brief History
 - Major Activities
- **U-STAR Speech Translation Service**
 - Service architecture
 - Service connection protocol
 - Resource and engine development
- **Evaluation and Issues**
 - Lab and field-testing evaluations
 - Major issues
- **Conclusion**

Technology Review¹

Confirmation of Feasibility

- **ITU Telecom World 1983**
NEC Corporation performed a demo as a concept exhibit



Extension of Technology

- **1993** ATR (Japan), CMU (USA) and Siemens jointly researched
- **1999 C-STAR** The Consortium for Speech Translation Advanced Research, aiming at a travel planning system using 6 languages (En, Ja, Ge, Ko, It, Fr)

Technology Review¹

Attempts at Practical Systems

- **2000 NESPOLE!** Negotiating through Spoken Language in E-Commerce, funded by NSF
- **2001 IBM** Multilingual Automatic Speech-to-Speech Translator (MASTOR) project funded by DARPA
- **2004 TC-STAR** Technology and Corpora for Speech-to-Speech Translation of European English, European Spanish, and Mandarin Chinese
- **2006 GALE** Global Autonomous Language Exploitation, funded by DARPA for translation Arabic and Chinese speech and text to English
- **2009 TransTac** Spoken Language Communication and Translation Systems for Tactical Use, funded by DARPA for military-used translation devices

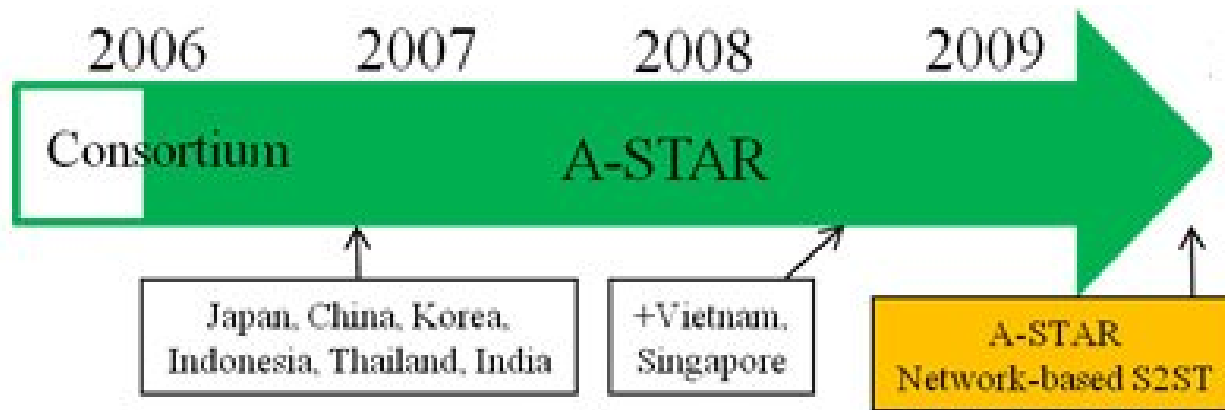
Outline

- **Technology Review**
- **U-STAR Consortium**
 - Brief History
 - Major Activities
- **U-STAR Speech Translation Service**
 - Service architecture
 - Service connection protocol
 - Resource and engine development
- **Evaluation and Issues**
 - Lab and field-testing evaluations
 - Major issues
- **Conclusion**

U-STAR Consortium²

- **2006 : A-STAR Consortium**

Asian Speech Translation Advanced Research



- **Basic Travel Expression Corpus (BTEC)** translated to 8 Asian languages by member countries
- **Speech Translation Marked-up Language (STML)** proposed as a standard connection protocol in APT/ASTAP

U-STAR Consortium²

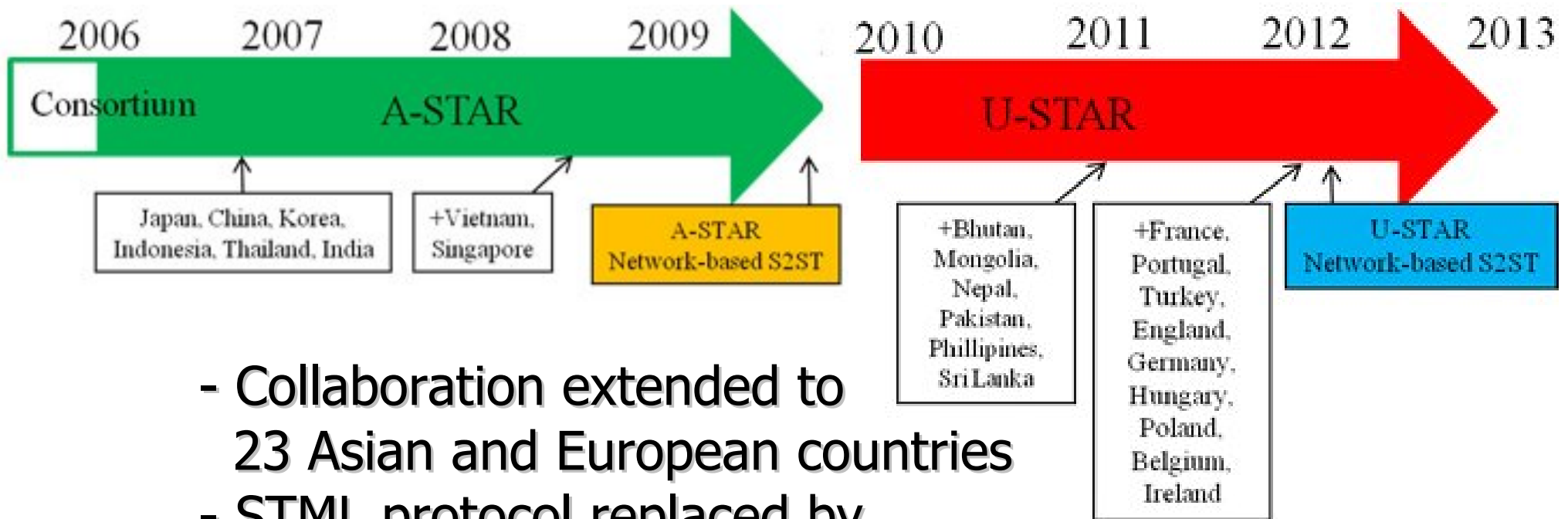
- **2009 : A-STAR S2ST Live Demo**
 - Network-based Multilingual S2ST
 - 8 Asian languages and English
 - Peer-to-peer and Multi-party clients
 - Portable devices (UMPC)



U-STAR Consortium²

- **2010 : U-STAR Consortium**

Universal Speech Translation Advanced Research



- Collaboration extended to 23 Asian and European countries
- STML protocol replaced by **Multimedia Content Marked-up Language (MCML)**, registered as an ITU-T recommendation standard

U-STAR Consortium²

- **2012 : U-STAR S2ST Public Service**

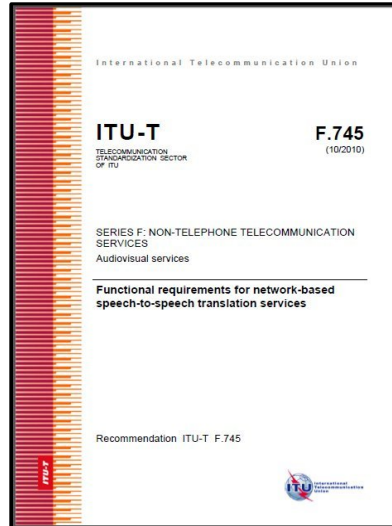
- Network-based Multilingual S2ST in the travel and sport domain
- 23 Asian and European languages supported
- **VoiceTra4U-M**, an iPhone App available freely on the AppStore
- Service launched in Jun 2012, before the opening of London Olympic Games



Outline

- **Technology Review**
- **U-STAR Consortium**
 - Brief History
 - Major Activities
- **U-STAR Speech Translation Service**
 - Service architecture
 - Service connection protocol
 - Resource and engine development
- **Evaluation and Issues**
 - Lab and field-testing evaluations
 - Major issues
- **Conclusion**

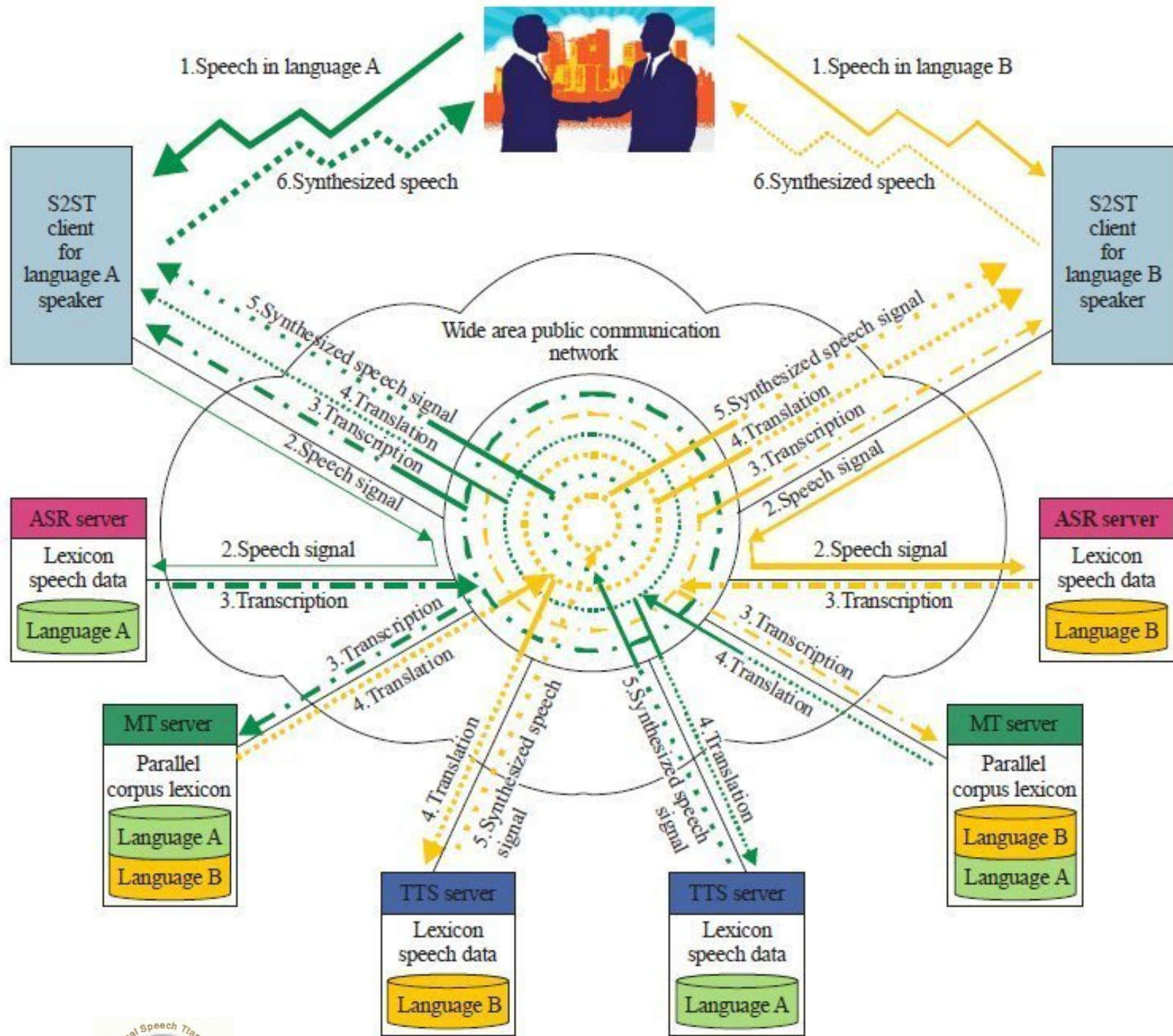
U-STAR S2ST Service Protocol³



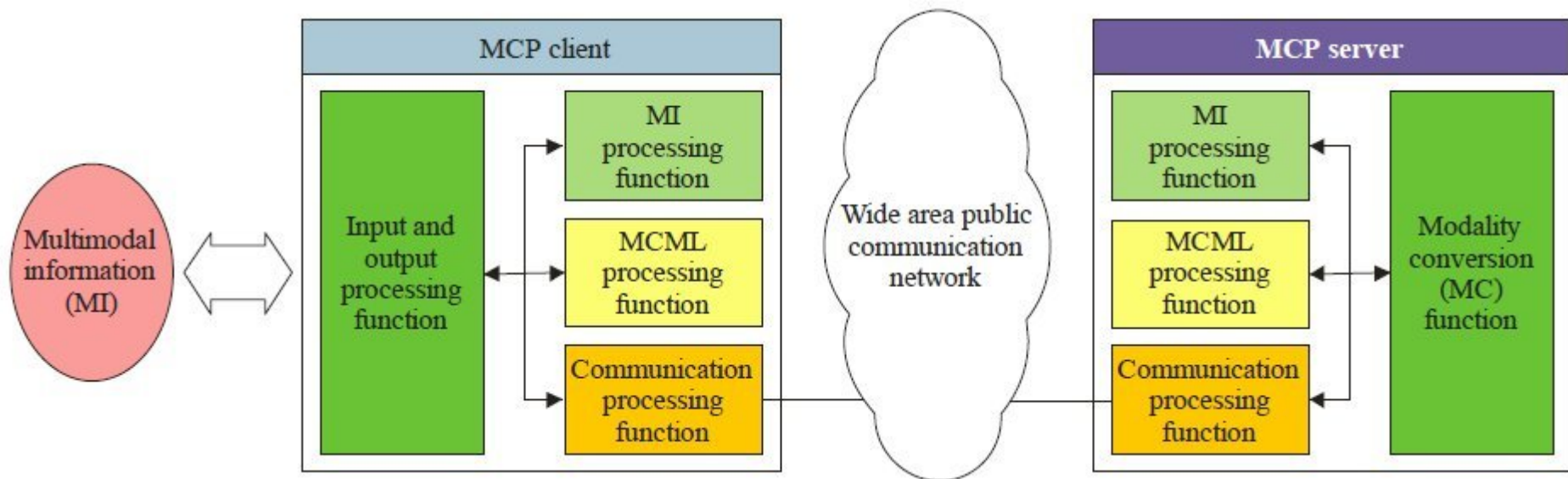
- ITU-T F.745 – Functional requirements for network-based speech-to-speech translation services



- ITU-T H.625 – Architecture for network-based speech-to-speech translation services



U-STAR S2ST Service Protocol³

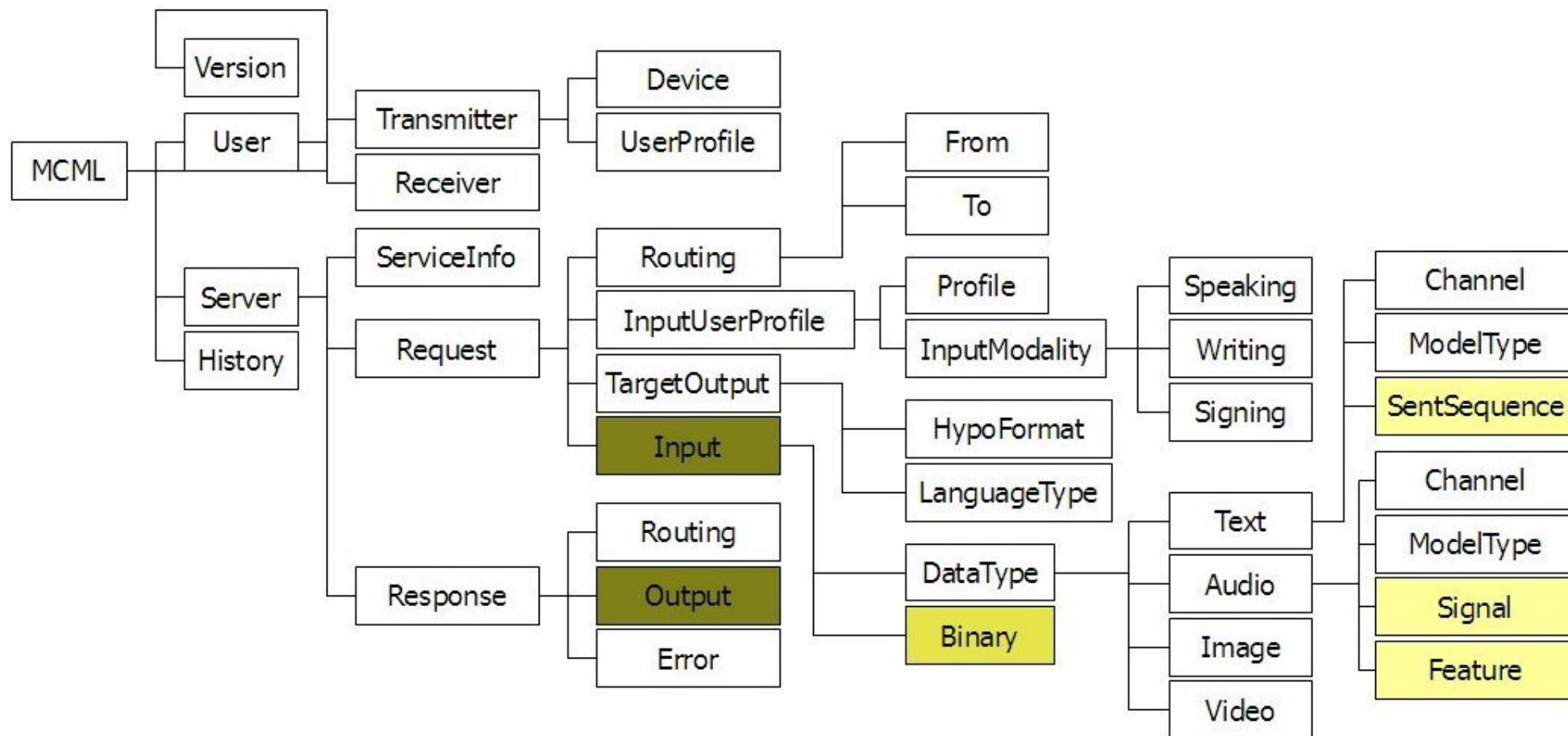


Modality Conversion Protocol (MCP)

- Multimodal Information (MI) transferred to/from a MCP client, i.e. a S2ST client
- MCP client communicates with MCP server using **Modality Conversion Marked-up Language (MCML)**
- MCP server includes ASR, MT, and TTS servers

U-STAR S2ST Service Protocol³

- A part of MCML structure



U-STAR S2ST Development

- **Common Language Resources**

- **Basic Travel Expression Corpus (BTEC)** has been used to translate to member languages since A-STAR
- To extend the service for users during London Olympic Games, an Olympic expression corpus by **Harbin Institute of Technology (HIT)** has been acquired and distributed to translate
- A **Named Entity (NE) list** of words related to Olympic expressions has also been collected from member countries
- Parallel corpora have been NE tagged for class-based language modeling

U-STAR S2ST Development

- **Thai Language Resources⁴**

Corpus	Purpose	Characteristic
Read speech (LOTUS PB)	Acoustic model initialization	48 speakers 15,600 utterances
Multi-conditioned read speech (NECTEC-ATR)	Acoustic model re-estimation	50 hours, 48 speakers 128,768 utterances
Telephone speech (LOTUS-CELL)	Acoustic model adaptation	20 hours, 162 speakers 35,851 utterances
Travel-domain text (BTEC)	Language modeling	87,355 sentences 14,685 vocabulary
Sport-domain text (HIT)	Language modeling	56,460 sentences 9,576 vocabulary

U-STAR S2ST Development

- **NE Category⁴**

No.	Category	Example
1	SPORT	archery, badminton, basketball
2	PERSON	Chai, John, Gihan, Eiichiro
3	TRANSPORT	bus, bicycle, airplane, car
4	COUNTRY	Australia, Belgium, Brazil, India
5	CURRENCY	Dong, US Dollar, Manat, Baht

U-STAR S2ST Development

- **Examples of Engines⁵**

Language	ASR	MT	TTS
English (En)	HMnet (SSS)		Concatenative
Hindi (Hi)		SMT (Cleopatra)	HMM
Indonesian (Id)	HMnet (SSS)	SMT (Moses)	HMM
Japanese (Ja)	HMnet (SSS)	SMT (Cleopatra)	Concatenative
Korean (Ko)	FST	RBMT (Parser)	HMM
Malay (Ms)	HMM	RBMT (Pyramid)	HMM
Thai (Th)	HMM	SMT (Moses)	HMM
Vietnamese (Vi)	HMM	SMT (Moses)	HMM
Chinese (Zh)	HMnet (SSS)	SMT (Cleopatra)	Concatenative

U-STAR S2ST Client

- **VoiceTra4U-M iPhone App**



- Peer-to-peer communication
- Multi-party chatting
- Available freely on AppStore for worldwide field-testing



U-STAR S2ST Client

- 23 languages supported
- 17 languages speech input enabled

Language	Input		Output	
	Speech	Text	Speech	Text
Dutch	✓	✓		✓
Dzongkha		✓		✓
English	✓	✓	✓	✓
French	✓	✓		✓
German	✓	✓		✓
Hindi	✓	✓	✓	✓
Hungarian	✓	✓	✓	✓
Indonesian	✓	✓	✓	✓
Japanese	✓	✓	✓	✓
Korean	✓	✓	✓	✓
Malay	✓	✓	✓	✓
Mandarin	✓	✓	✓	✓
Mongolian		✓		✓
Nepali		✓		✓
Polish	✓	✓	✓	✓
Portuguese	✓	✓	✓	✓
Russian	✓	✓		✓
Sinhala		✓		✓
Tagalog		✓		✓
Thai	✓	✓	✓	✓
Turkish	✓	✓	✓	✓
Urdu		✓	✓	✓
Vietnamese	✓	✓	✓	✓

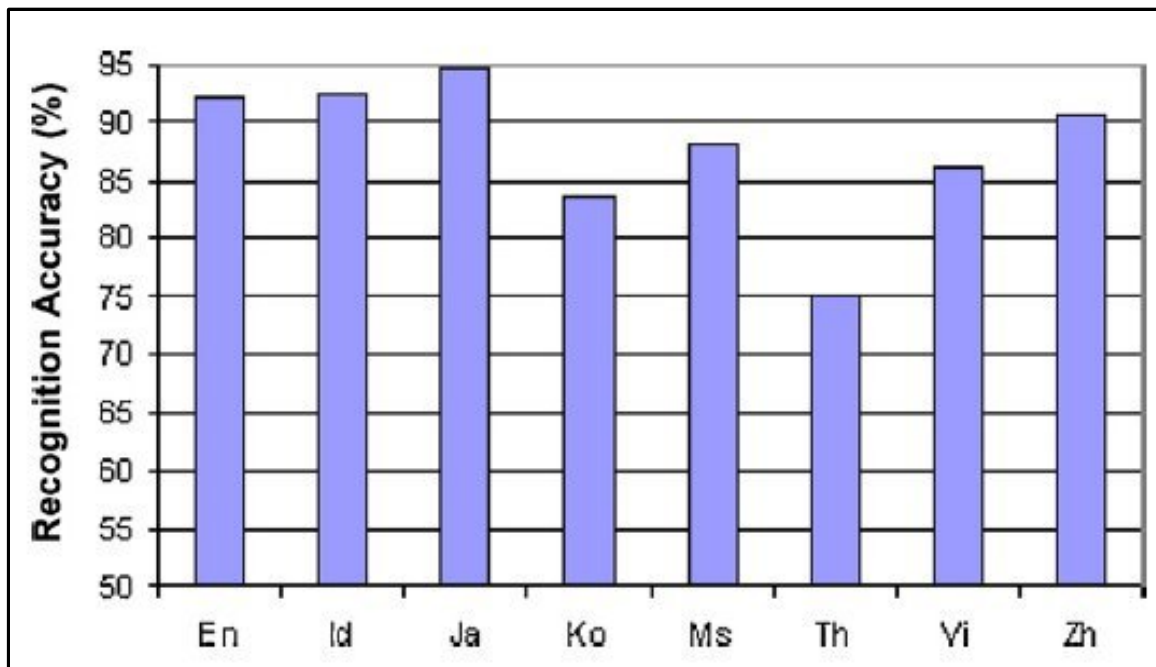
Outline

- **Technology Review**
- **U-STAR Consortium**
 - Brief History
 - Major Activities
- **U-STAR Speech Translation Service**
 - Service architecture
 - Service connection protocol
 - Resource and engine development
- **Evaluation and Issues**
 - Lab and field-testing evaluations
 - Major issues
- **Conclusion**

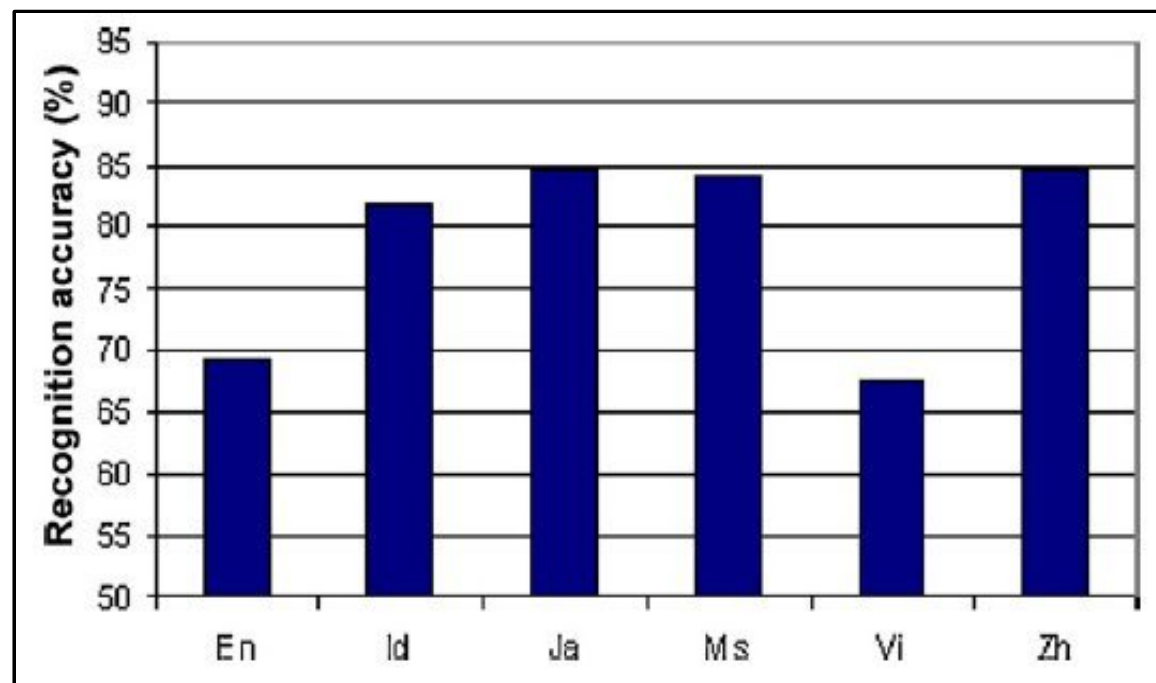
Evaluations

- **ASR Performance⁵ (2010)**

Test set taken from BTEC



Test set including dialog scenarios



Evaluations

- **MT Performance (2010)⁵**

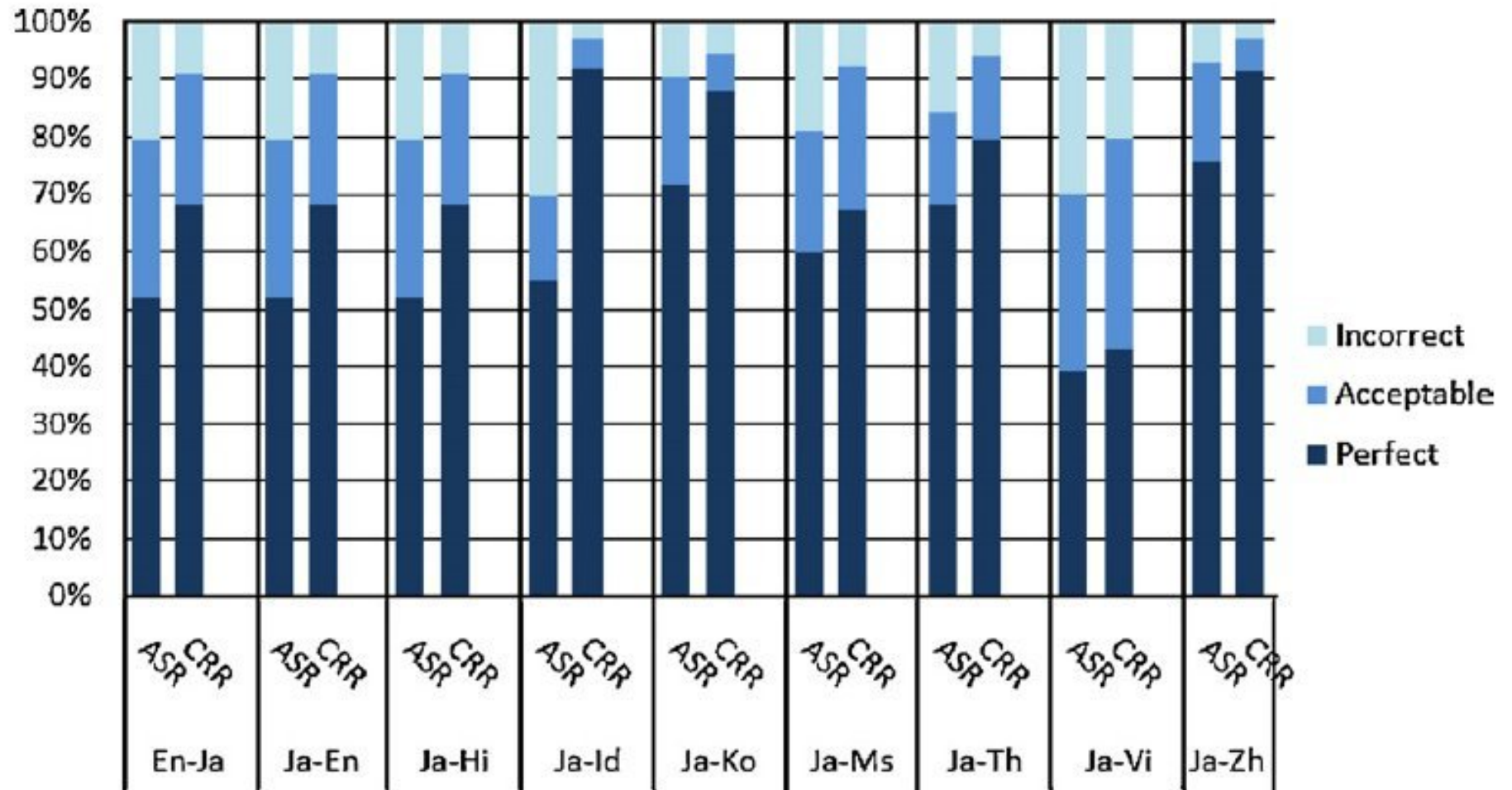
Pair	ASR Result	Correct Recognition Result
En-Ja	27.8	42.8
Ja-En	38.1	43.0
Ja-Hi	10.9	12.2
Ja-Id	25.1	28.2
Ja-Ko	19.2	23.0
Ja-Ms	33.7	34.9
Ja-Th	32.4	37.7
Ja-Vi	25.1	28.0
Ja-Zh	35.1	41.1

Test set including dialog scenarios

Measured in BLEU score

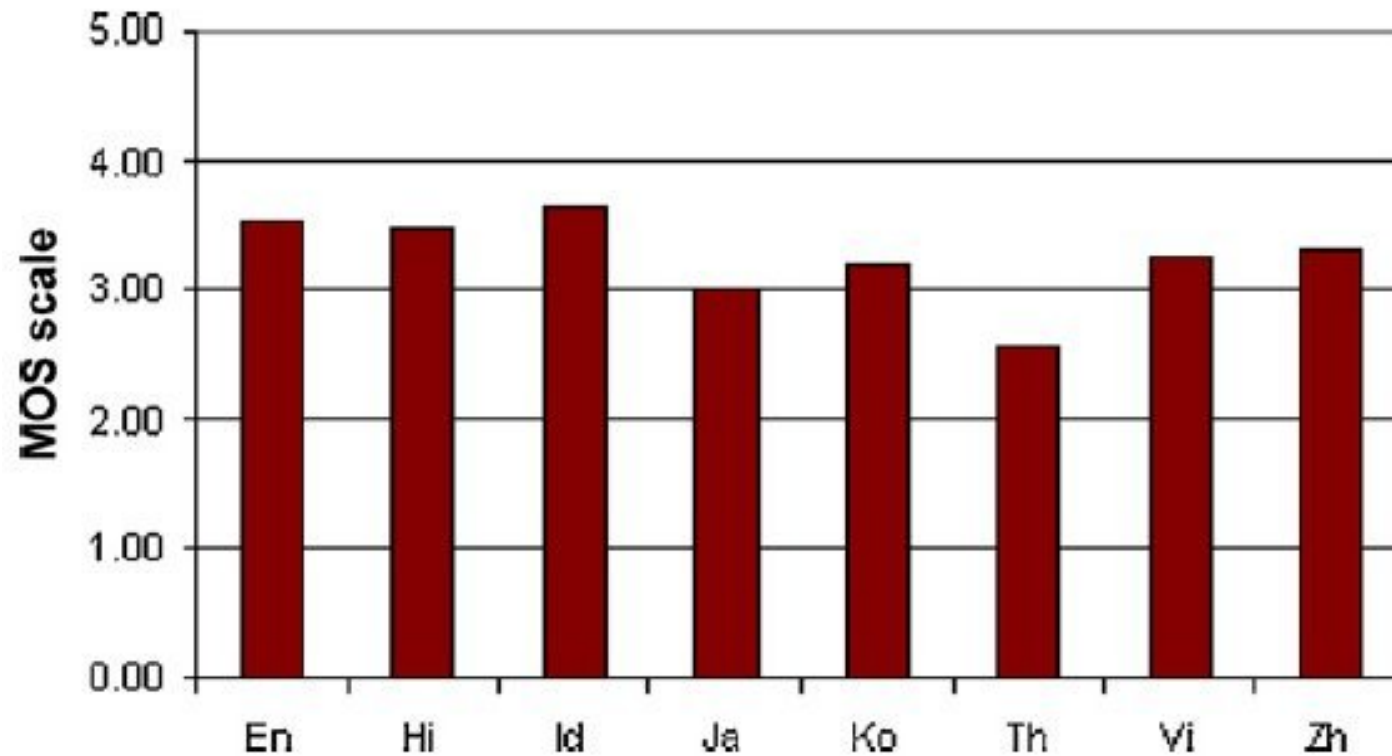
Evaluations

- **MT Performance (2010)⁵**



Evaluations

- **TTS Performance (2010)⁵**



Evaluations

- **No. of Downloads (Jul-Oct 2012) – 15,645 in total**

Country	Total
Japan	7266
Thailand	5984
United States	1098
Brazil	253
Taiwan	177
Singapore	165
China	154
United Kingdom	76
Australia	56
Russian Federation	45
France	40
Canada	29
Hong Kong	27

Country	Total
Germany	20
Indonesia	18
Poland	16
Malaysia	12
Belgium	11
Israel	11
India	11
Viet Nam	10
Spain	9
Korea	9
Latvia	9
Italy	8
Netherlands	8

Evaluations

- **No. of Transactions (2012) – 26,882 in total**

Language	Utterances
Japanese	10333
English	5179
Thai	7104
Chinese	965
English (UK)	833
Korean	496
Hindi	181
French	263
German	1362
Indonesian	122
Polish	22
Hungarian	10
Portuguese	7
Malay	5

Evaluations

- **Analysis of Thai ASR Speech Input**
 - 2,480 utterances during Jul 2012

Useful		Garbage	
Normal	52.6%	Incorrect language used	4.8%
Noisy	23.1%	Silence only	16.4%
Left or right chopped	3.1%		

- **Thai Language Model Interpolation**
 - Using another 1,000 utterances from real services

Interpolation Weight		WER (%)	PP	OOV (%)
BTEC+HIT	1,000 Utt.			
1.0	0.0	51.7	77.5	3.4
0.25	0.75	59.3	13.9	1.4

Issues

- **Named-Entities (NE)**

- NE words are often language specific
- When there is no direct translation of a given NE word
 - 1) Using a compound or descriptive word
 - 2) Using transliteration
- Compounds occurred e.g. in Thai have often made confusion with common words in class-based language modeling

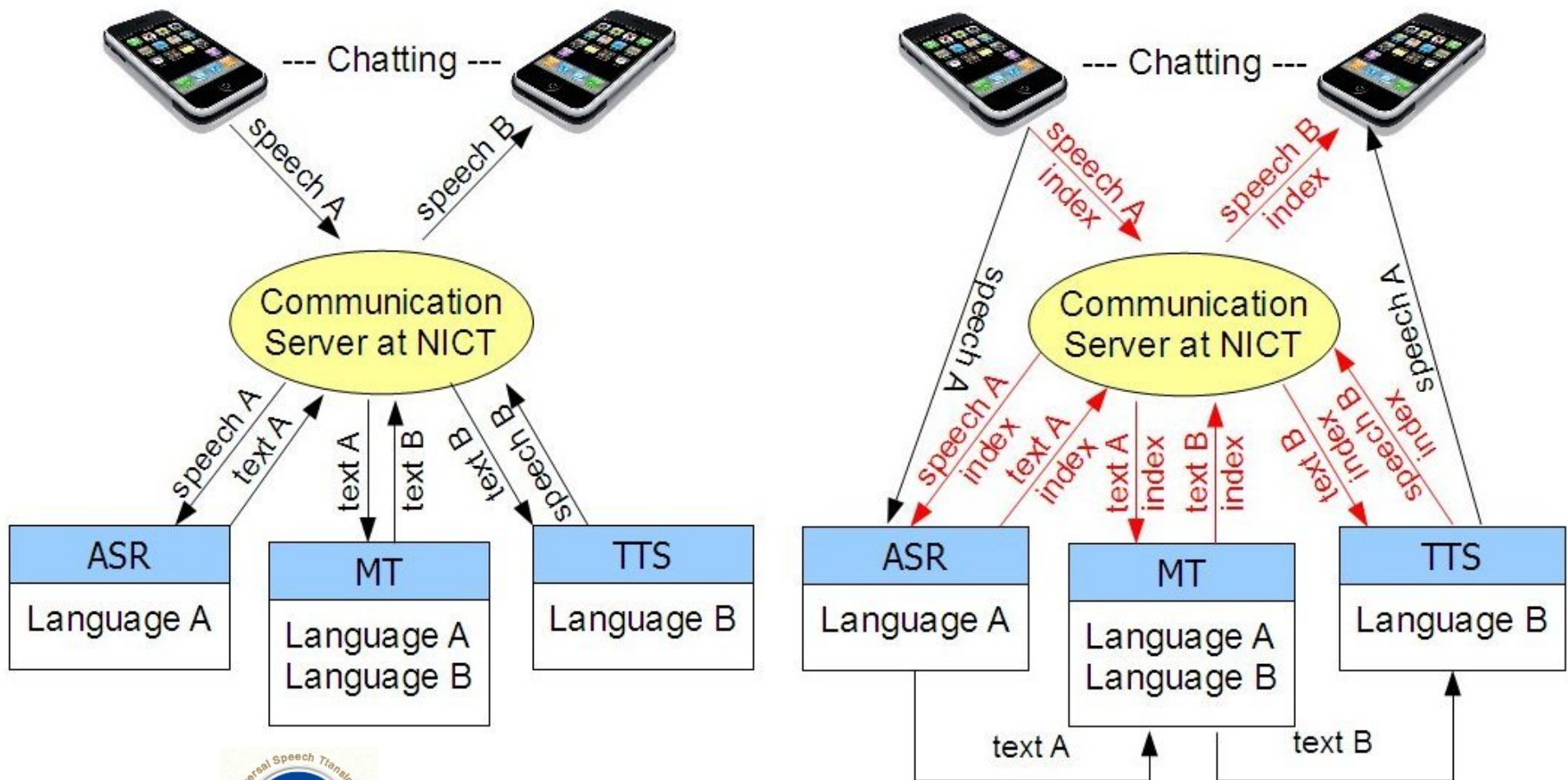
- **Scalability and Extensibility**

- Service capacity requires continuous maintenance of all language services locating in member countries
- Improving service performance by enlarging service lexicon and training data gathered from real usage
- Extending to new domains and languages

Issues

- **Service Latency**

- The condition of network is the key
- Setting communication mirror servers

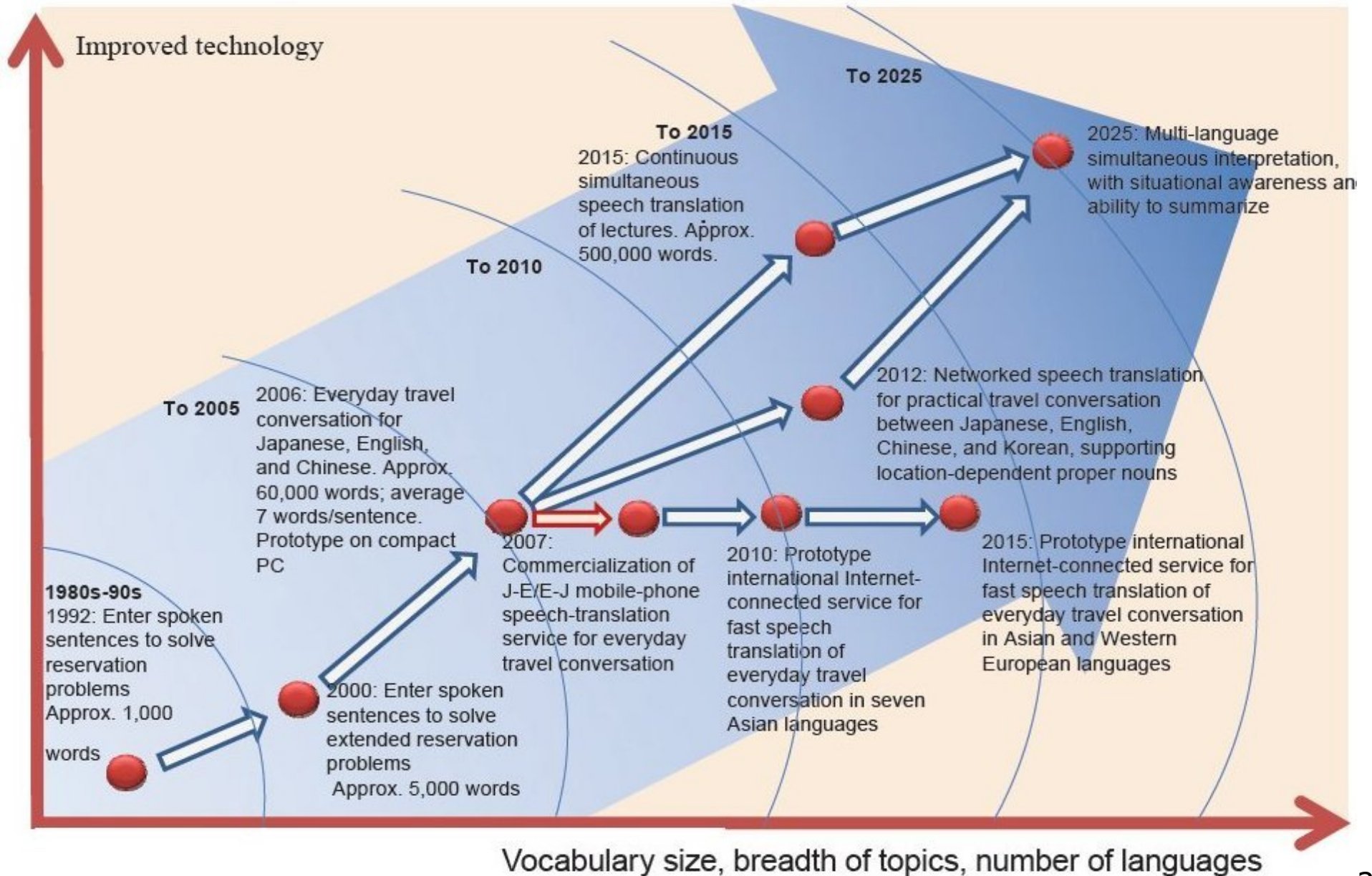


Outline

- **Technology Review**
- **U-STAR Consortium**
 - Brief History
 - Major Activities
- **U-STAR Speech Translation Service**
 - Service architecture
 - Service connection protocol
 - Resource and engine development
- **Evaluation and Issues**
 - Lab and field-testing evaluations
 - Major issues
- **Conclusion**

Conclusion

• Future of S2ST¹



Conclusion

- **Advantages of U-STAR Framework**

- U-STAR provides a service infrastructure with the ITU-T recommendation connection standard available for extending to Universal S2ST, where individual language engines are flexible
- Language resources and tools sharing in the network are useful for future research and innovation

- **Near Future Direction**

- Service improvement in term of accuracy and latency
- Service extension to new member languages

References

- ¹ S. Nakamura, 2009. Overcoming the language barrier with speech translation technology. Science and Technology Trends – Quarterly Review No. 31, Apr 2009, pp. 35-48.
- ² U-STAR consortium, <http://ustar-consortium.com/>
- ³ ITU-T standard, <http://www.itu-t.int/>
- ⁴ C. Wutiwiwatchai, K. Thangthai, P. Sertsi, 2012. Thai ASR development for network-based speech translation. To be printed in Proc. of O-COCOSDA 2012.
- ⁵ S. Sakti, M. Paul, A. Finch, S. Sakai, T. T. Vu, N. Kimura, C. Hori, E. Sumita, S. Nakamura, J. Park, C. Wutiwiwatchai, B. Xu, H. Riza, K. Arora, C. M. Luong, H. Li, 2011. Toward translating Asian spoken languages. Computer Speech and Language (2011), doi:10.1016/j.csl.2011.07.001.



<http://ustar-consortium.com/>