

# **ASEAN Language Speech Translation thru' U-STAR**

ASEAN IVO 2016 Project  
24<sup>th</sup> Nov 2016

*Presented by*

**Prof LI Haizhou  
Ms AW Ai Ti**

# Project Introduction

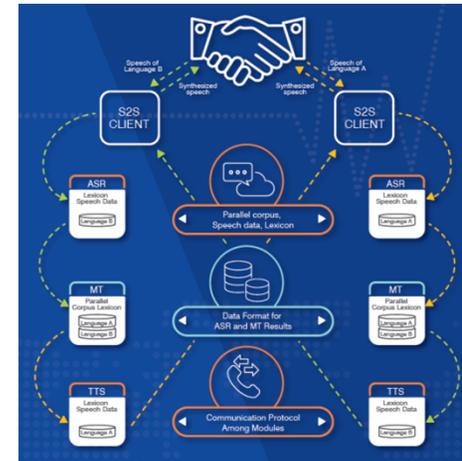
To provide a single and unified platform for multilingual communications across the ASEAN nations through U-STAR

To collaborate and facilitate the development of speech translation technologies, language resources and translation app for ASEAN languages



<http://www.ustar-consortium.com/members.html>

Universal  
Speech  
Translation  
Advanced  
Research



## Project Members (8 members from 7 countries)

National Institute of Posts, Telecommunications, and Information Communication Technology, Cambodia (NIPTICT); Agency for the Assessment and Application of Technology, Indonesia (BPPT); Universiti Teknologi Malaysia, Malaysia (UTM); University of Computer Studies, Yangon, Myanmar (UCSY); Institute for Infocomm Research, Singapore (I<sup>2</sup>R); National Electronics and Computer Technology Center, National Science and Technology Development Agency, Thailand (NECTEC); Hanoi University of Science and Technology, Vietnam (HUST); Institute of Information Technology, Vietnam Academy of Science and Technology, Vietnam (IOIT)

# Project Details

## Duration

1<sup>st</sup> July 2016 to 30<sup>th</sup> June 2019

## Target

- An integrated speech translation service for the ASEAN community.
- A common ASEAN speech translation evaluation language resource for ASEAN speech translation system evaluation.

## **Benefit?**

- A single tool and platform to benefit the technology advancements of any speech translation components within ASEAN and Japan
- A unique and critical resources required by all researchers in ASEAN and beyond working on ASEAN speech translation research and development.

# Leveraging U-STAR network and infrastructure



**U-STAR** - An international consortium for research and collaboration of speech translation technologies

- Provide the platform and global network for enabling real-time, location-free, multi-party communication among different language speakers

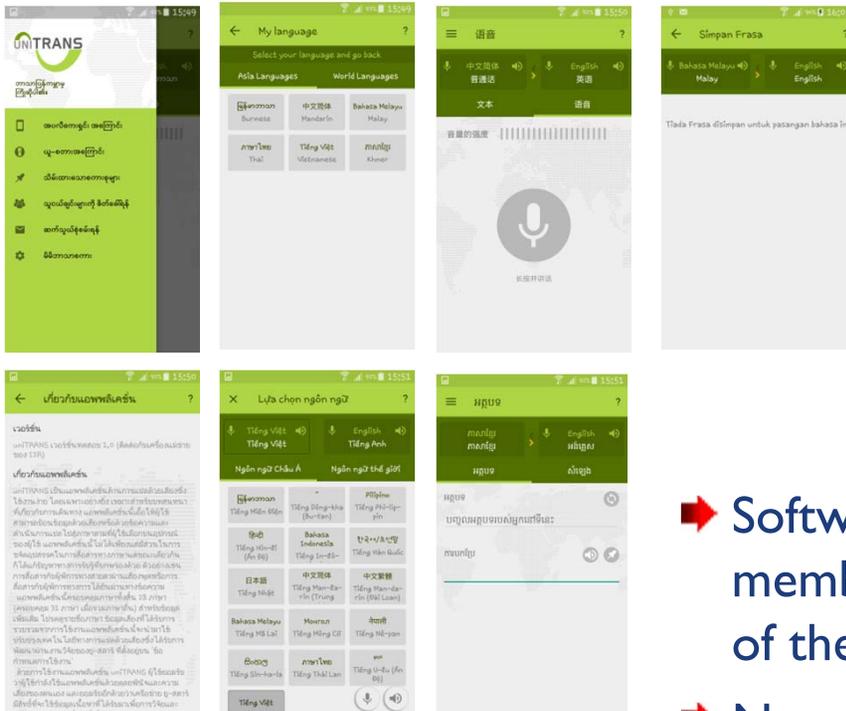
**ASEAN Language Speech Translation** - An regional collaboration to collaborate as well as facilitate the development and use of speech translation components provided through U-STAR

- All members are U-STAR members
- Leverage on the framework, platform and technology provided by U-STAR
- Facilitate the advancement of technology through the development of a common evaluation set



- Voice Translation
- Text Translation
- Saved Phrases

# UNITRANS Development



Organization	Country	Language
NIPTICT	Cambodia	Khmer
BPPT	Indonesia	Bahasa Indonesia
UTM	Malaysia	Bahasa Melayu
UCSY	Myanmar	Myanmar
I <sup>2</sup> R	Singapore	Chinese
NECTEC	Thailand	Thai
HUST	Vietnam	Vietnamese
IOIT	Vietnam	Vietnamese

- ➔ Software resource file and images given to the member institutes for collaborative localization of the app
- ➔ New languages to be integrated through the U-STAR platform
  - Myanmar ASR (by UCSY and NICT)
  - Khmer MT (by NIPTICT and NICT)
  - Indonesian MT (by BPPT)

# Project Meeting and Workshop in Singapore

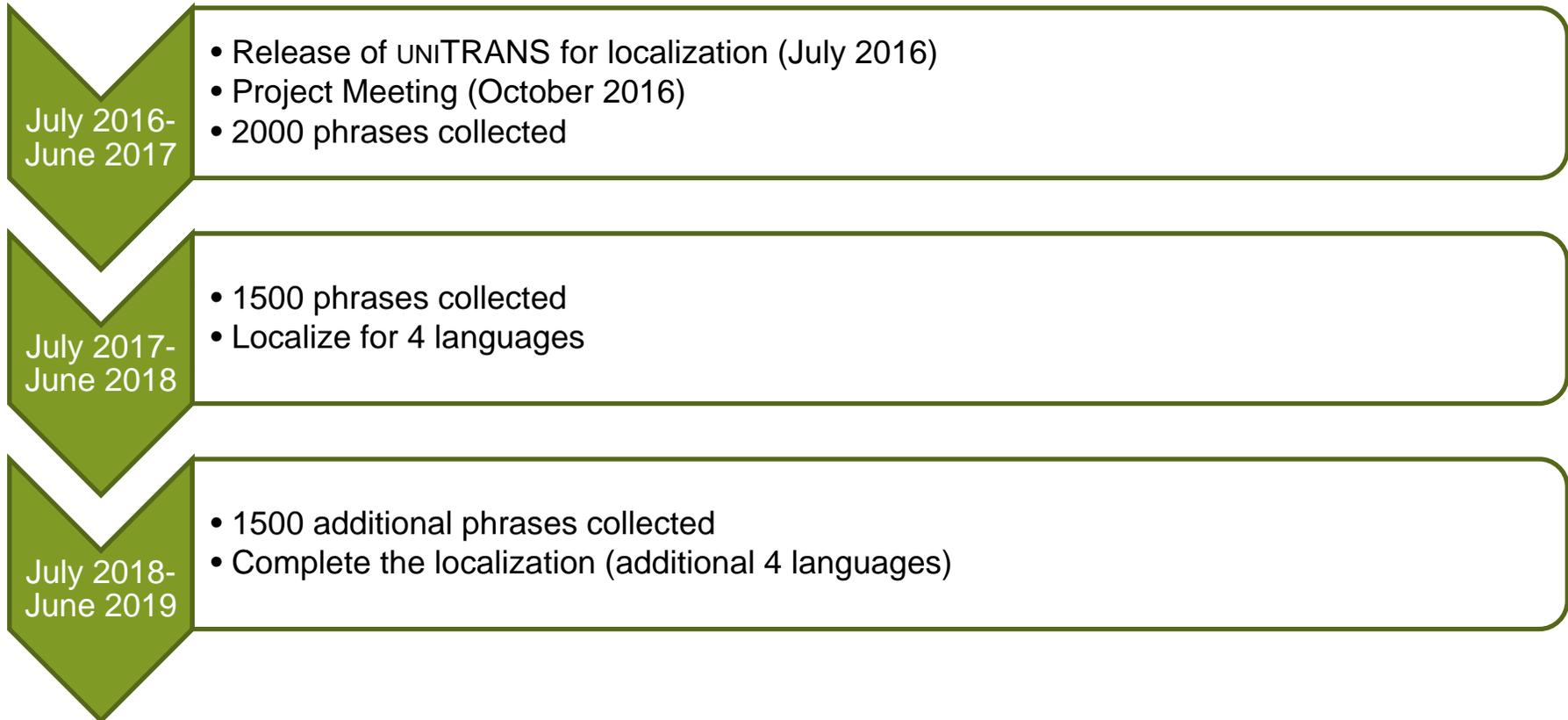


- 4<sup>th</sup> October 2016 during HAI Conference
- 20 participants
- 9 technical papers shared by team members
  - 3 on ASR
  - 2 on system
  - 4 on MT and NLP

- Data Collection Specifications
- Promote UNITRANS within local community
- Invite more ASEAN members to join the project



# Project Timeline



## Speech Translation Evaluation Dataset Collection

- Guidelines to be ready by 30<sup>th</sup> November 2016
- Total of 5000 phrases, including audio, transcription and translation
- Attempt to gather natural, local common expressions used locally rather than obtaining data translated from another language (like English)

# Collaboration with ALT project

- ➡ Use the golden alignment information provided via ALT project to enhance the MT alignment model.
- ➡ Use the ALT corpus to incorporate deeper linguistic features (POS and NER) into MT
- ➡ Incorporate ALT corpus as additional set of training data for MT engine training.

*Thank  
You!*