Oriental COCOSDA 2018

CONFERENCE GUIDE

May 7-8, 2018, Phoenix Seagaia Resort, Miyazaki, Japan

O-COCOSDA 2018
MIYAZAKI

Table of Contents

Message of the O-COCOSDA Convenor

Welcome to Oriental COCOSDA/CASLRE 2018 in Japan. This is the 21th anniversary conference of Oriental COCOSDA, the Oriental chapter of the International **Co**mmittee for **Co**-ordination and **S**tandardisation of Speech **Da**tabases. We celebrated the 20 year anniversary in Korea in 2017, and believe the Oriental COCOSDA 2018 will be the first step towards the next generation of Oriental COCOSDA. It is our great pleasure to organize the conference in Japan this year since the first annual meeting of Oriental COCOSDA 1998 was held in Tsukuba Japan organized by Prof. Itahashi.

In addition, Oriental COCOSDA 2018 will be held jointly with LREC 2018, Language Resources and Evaluation Conference to help boost the research and development in the field of speech technology and further enthusiasm towards speech technology in East and Southeast Asia.

As a non-profit organization that doesn't charge membership fees or other financial resource, Oriental COCOSDA has been having annual meetings since 1998, and become one of the most active organizations in Asia. Some of our long standing members are active also in other well-known international scholarly communities and acting as liaison. This means our members could also be more involved in the international community at large in future.

Over the years, the Oriental COCOSDA has fostered, directly and indirectly, several multi-national research collaborations in Asia. Using a common platform for data collection and providing unrestricted access of collected corpora towards research and development as well as performance assessment is no longer just an aim, but there are projects in progress.

Following the Oriental COCOSDA convention from the very beginning, there will be reports of activity updates from regional members each year. Please attend Oriental COCOSDA country report session on the last day.

All the submissions were peer reviewed. As usual we are working to index oral presentation papers by IEEE Xplore. While indexing is entirely subject to the discretion of each organizer, the action was nonetheless well received by younger colleagues towards travel subsidies and career track records.

The activities and contributions of Oriental COCOSDA as an international scientific community were widely acknowledged and the Antonio Zampolli Award was given to Prof. Itahashi, the founder of Oriental COCOSDA and the first convenor, Prof. Chiu-yu Tseng, the second convenor, and me as the third convenor by ELRA in 2014, and it founded the Oriental COCOSDA ITN Paper Award. The award recipient will be announced later.

I would like to thank Honorary Chair of Oriental COCOSDA 2018, Prof. Shuichi Itahashi, Program Chair, Prof. Nobuaki Minematsu, Local Committee Chair, Prof. Satoshi Tamura and Dr. Tomoko Osuga, and Local Organization Committee Members, Dr. Keiji Yasuda and Ms. Manami Matsuda, for their enormous efforts making the event possible this time in Japan. I deeply appreciate and thank all of the organization committee for the great contributions.

For the sponsorship, we deeply thank Arcadia for their continuous support for Oriental COCOSDA.

As Oriental COCOSDA Convenor, it has been my honor and privilege to serve for the community. We will of course work hard to make the community better, and at the same time we would like to ask for your continued cooperation to Oriental COCOSDA/CASLRE.

Prof. Dr. Satoshi Nakamura
Oriental COCOSDA Convenor

Committee

Organizing Committee

Honorary Chair
Shuichi Itahashi (University of Tsukuba, Japan)

Conference Chair
Satoshi Nakamura (Nara Institute of Science and Technology, Japan)

Program Chair
Nobuaki Minematsu (The University of Tokyo, Japan)

Local Chair
Satoshi Tamura (Gifu University, Japan)

Local Vice Chair
Tomoko Ohsuga (National Institute of Informatics, Japan)

Support Members
Keiji Yasuda (Nara Institute of Science and Technology, Japan)
Manami Matsuda (Nara Institute of Science and Technology, Japan)

Oriental COCOSDA Committee

Convenor
Satoshi Nakamura (Japan)

Steering Committee
Shuichi Itahashi (Japan)
Yong-Ju Lee (Korea)
Aijun Li (China)
Haizhou Li (Singapore)
Luong Chi Mai (Vietnam)
Satoshi Nakamura (Japan)
Agrawal Shyam (India)
Chiu-yu Tseng (Taiwan)
Chai Wutiwiwachai (Thailand)

Scientific Advisory
Hiroya Fujisaki (Japan)
Lin-shan Lee (Taiwan)

Country Representative

– China
Dong Wang
Aijun Li
Thomas Fang Zheng
Dawa Idomuco
Nasirjan Tursun

– East Timor
Borja L. C. Patrocinio Antonino

– Hong Kong
Tan Lee

– India
K. Samudravijaya
Agrawal Shyam

– Indonesia
Hammam Riza
Dessi Puji Lestari

– Ireland
Nick Campbell

– Japan
Shuichi Itahashi
Satoshi Nakamura

– Korea
Yong-Ju Lee
Minhwa Chung

– Malaysia
Zuraida Mohd Don
Pandian Ambigapathy
Syed Zanial Ariff Syed Jamal

– Mongol
Altangerel Ayush
Purev J
Ananlada Chotimongkol

– Myanmar
Win Pa Pa

– Nepal
Regmi Bhim

– Pakistan
Sarmad Hussain
Tania Habib
Dinusha Thilini

– Philippines
Nathaniel Oco
Roxas Rachael
Cu Jocelyn

– Singapore
Haizhou Li
Dong Minghui
Kim-Teng Lua

– Taiwan
Sin-Horng Chen
Chiu-yu Tseng

– Thailand
Wutiwiwachai Chai
Anocha Rugchatjaroen

– Vietnam
Luong Chi Mai

Program Overview

	7-May	8-May	9-May ~ 11-May
09:00 - 09:30	Registration	Session 5	LREC 2018
09:30 - 10:00	Opening Session 1		
10:00 - 10:30		Break	
10:30 - 11:00		Session 6 (Poster)	
11:00 - 11:30	Break Session 2		
11:30 - 12:00			
12:00 - 12:30	Lunch (steering committee meeting)	Lunch (country representatives meeting)	
12:30 - 13:00			
13:00 - 13:30			
13:30 - 14:00	Keynote 1	Keynote 2	
14:00 - 14:30			
14:30 - 15:00	Group Photo Session 3	Break Session 7	
15:00 - 15:30			
15:30 - 16:00			
16:00 - 16:30	Break Session 4	Break Country Report	
16:30 - 17:00			
17:00 - 17:30			
17:30 - 18:00		Closing	
18:00 - 18:30	Banquet at PINE TERRACE		
18:30 - 19:00			
19:00 - 19:30			

DAY-1 May 7, 2018

9:30- 9:40	Opening Remarks
9:40-10:55	Session 1: Corpus
10:55-11:05	Break
11:05-11:55	Session 2: Analysis
12:00-13:30	Lunch
13:30-14:30	Keynote 1
14:30-14:45	Group Photo
14:45-16:05	Session 3: Model & Sysytem
16:05-16:15	Break
16:15-17:30	Session 4: Second Language
18:00-	Banquet

DAY-2 May 8, 2018

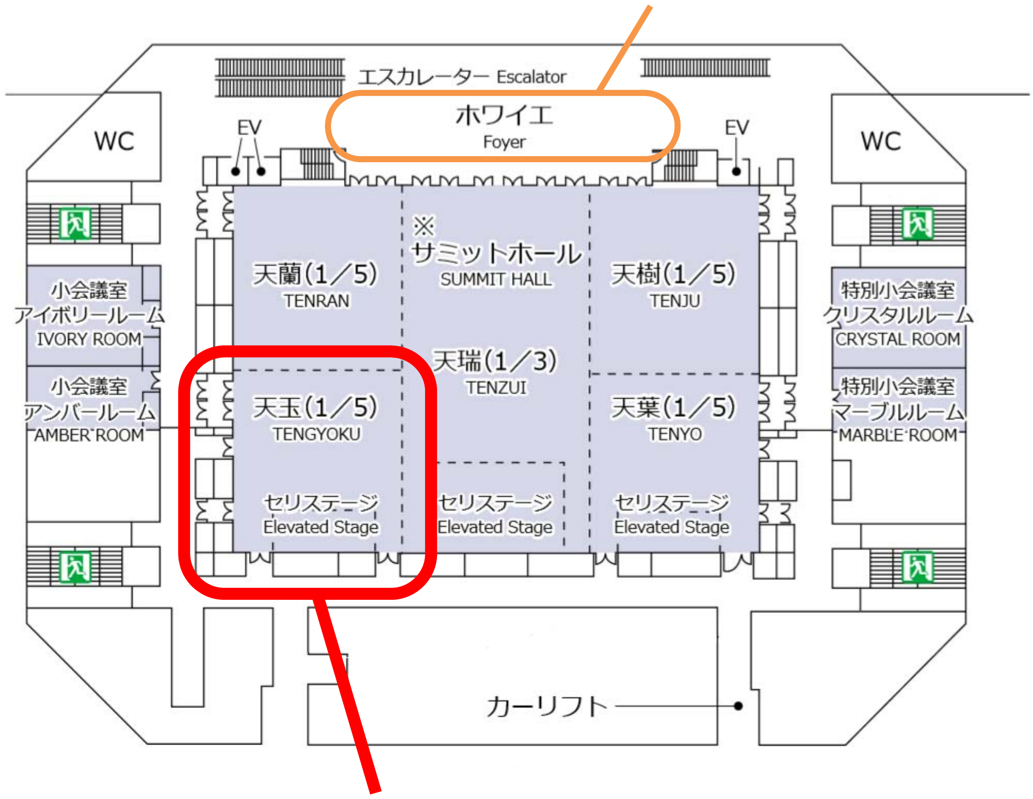
9:00-10:15	Session 5: Corpus
10:15-10:45	Break
10:45-12:00	Session 6: Posters
12:00-13:30	Lunch
13:30-14:30	Keynote 2
14:30-14:40	Break
14:40-15:55	Session 7: Analysis
15:55-16:05	Break
16:05-17:35	Country Report
17:35-17:45	Closing Remarks

Floor Guide

Phoenix Seagaia Convention Center

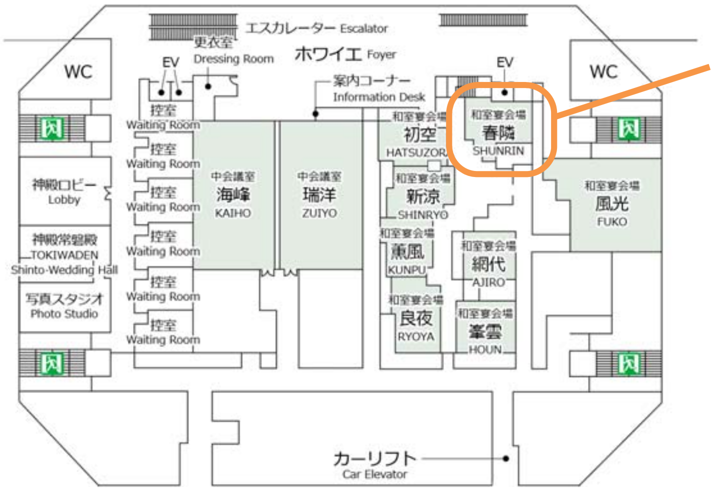
4th Floor

Coffee-break stations



Conference room: TENGYOKU (天玉)

3rd Floor



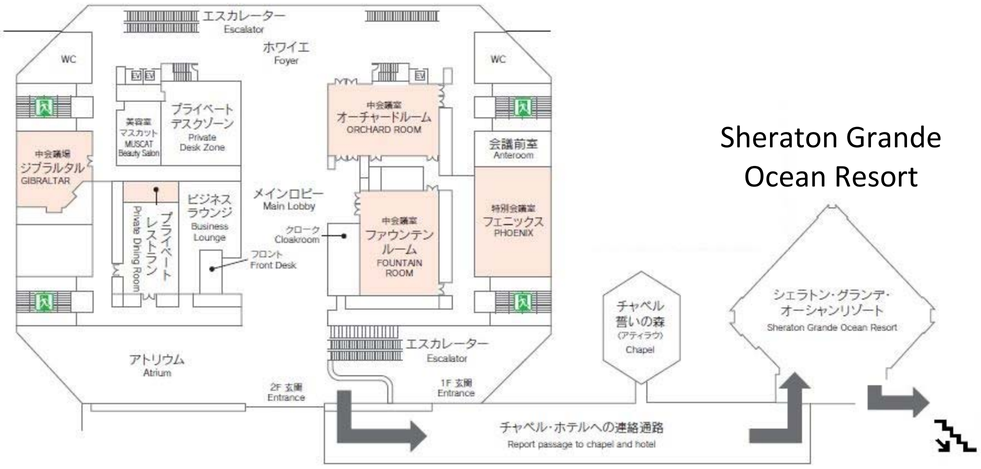
SHUNRIN (春隣)

7-May
Steering committee lunch meeting

8-May
Country representatives lunch meeting

2nd Floor

Convention Center



Sheraton Grande
Ocean Resort

1st Floor

Banquet:
Garden Buffet
PINE TERRACE

Technical Program

DAY-1 May 7, 2018

Opening

9:30 - 9:40 Opening remarks

Session 1 (CORPUS) *Chair: Chai Wutiwiwatchai*

9:40 - 10:05 Development of Text and Speech Corpus for Designing the Multilingual Recognition System
Shweta Bbansal and Shyam S. Agrawal

10:05 - 10:30 Urdu Speech Corpora for Banking Sector in Pakistan
Benazir Mumtaz, Sahar Rauf, Hafsa Qadir, Javairia Khalid, Tania Habib, Sarmad Hussain, Rukhsana Barkat, and Ehsan Ul Haq

10:30 - 10:55 AWA Long-Term Recorded Speech Corpus and Robust Speaker Recognition Method for Session Variability
Satoru Tsuge, Shingo Kuroiwa, Tomoko Ohsuga, and Yuichi Ishimoto

10:55 - 11:05 [Break]

Session 2 (ANALYSIS) *Chair: Aijun Li*

11:05 - 11:30 Parenthetical – A Special Type of Prosodic Reduction in Continuous Speech
Chiu-yu Tseng, Helen Kai-yun Chen, and Yen-Hsing Chen

11:30 - 11:55 Acoustic Comparison of Vowel Articulation When Combined with Different Tone Categories in Mandarin
Chong Cao, Yanlu Xie, and Jinsong Zhang

12:00 - 13:30 [Lunch] *Steering committee meeting at SHUNRIN (3rd floor)

Keynote 1 *Chair: Satoshi Nakamura*

13:30 - 14:30 Linguistic Unit Discovery from Multi-modal Inputs in Unwritten Languages
Odette Scharenborg

Group Photo

14:30 - 14:45

Session 3 (MODEL & SYSTEM) *Chair: Haizhou LI*

14:45 - 15:10 Naso-Articulometry Speech Database for Cleft-Palate Speech Assessment
Chai Wutiwiwatchai, Patcharika Chootrakool, Sawit Kasiriya, Kalyanee Makarabhirom, Nantiya Ooppanasak, and Benjamas Prathanee

15:10 - 15:35 Multi-Modal Multi-Task Deep Learning for Speaker and Emotion Recognition of TV-Series Data
Sashi Novitasari, Quoc Truong Do, Sakriani Sakti, Dessi Lestari, and Satoshi Nakamura

15:35 - 16:05 Mathematical Modeling for Daniel – IPA Vowel System in CSOLP and Actual Application in Quantitative & Dynamic Research on IPA Diphthongs
Qiaoli Feng, Xuan Xiong, Wei He, Ziyu Ye, Min Yu, and Xiaogan Huang

16:05 - 16:15 [Break]

Session 4 (SECOND LANGUAGE) *Chair: Sin-Horng Chen*

16:15 - 16:40 A Typological Study of English Monophthongs Acquisition of EFL Learners in Shandong Dialect Area Region
Yuan Jia, Bin Li, and Aijun Li

16:40 - 17:05 Typology of Convergences and Divergences of English Monophthongs by Chinese Northeastern EFL Learners
Yuan Jia and Yu Wang

17:05 - 17:30 Acoustic Features of Mandarin Diphthongs by Uyghur Learners at Primary Level
Yultuz Rapkat, Gulnur Arkin, Mijit Ablimit, and Askar Hamdulla

Banquet

18:00 - at PINE TERRACE (1st floor of Sheraton Grande Ocean Resort)

DAY-2 May 8, 2018

Session 5 (CORPUS) *Chair: Luong Chi Mai*

9:00 - 9:25 Japanese-English Code-Switching Speech Data Construction
Sahoko Nakayama, Takatomo Kano, Quoc Truong Do, Sakriani Sakti, and Satoshi Nakamura

9:25 - 9:50 Speech Corpora of Under Resourced Languages of North-East India
Barsha Deka, Joyshree Chakraborty, Abhishek Dey, Shikhamoni Nath, Priyankoo Sarmah, S.R. Nirmala, and Samudravijaya K

9:50 - 10:15 Unsupervised Dependency Corpus Annotation for Myanmar Language
Hnin Thu Zar Aye, Win Pa Pa, and Ye Kyaw Thu

10:15 - 10:45 [Break]

Session 6 (POSTERS) *Chair: Shyam S. Agrawal*

10:45 - 12:00

Poster 1: The BLCU-SAIT Speech Corpus of Non-Native Chinese
Wei Wang, Xing Wei, Jiawei Yu, Wei Wei, Yanlu Xie, and Jinsong Zhang

Poster 2: An Enhancement of English-Thai Pronunciation Dictionary
Patcharika Chootrakool, Sittipong Saychum, Chai Wutiwiwatchai, and Anocha Rugchatjaroen

Poster 3: Assessment of Korean Spontaneous Speech Produced by Non-Native Learners: Issues and Methodology
Seung Hee Yang and Minhwa Chung

Poster 4: Acoustic Feature Analysis on the Chinese Mandarin Monophthongs Pronounced by Kazakh College Students
Guljan Alijan, Gulnur Arkin, Dilmurat Tursun, Mijit Ablimit, and Askar Hamdulla

Poster 5: Research on patterns of Unvoiced Fricatives in Uyghur Language
Parizat Keyim, Gulnur Arkin, Mijit Ablimit, and Askar Hamdulla

Poster 6: A High Quality and Phonetic Balanced Speech Corpus for Vietnamese
Phuong Pham Ngoc, Quoc Truong Do, and Luong Chi Mai

Poster 7: Acoustic Analysis of Vowels in Two Southern Angami Dialects
Viyazonuo Terhiija, Priyankoo Sarmah, and Samudravijaya K

Poster 8: Noise-Resistant Telephone Quality Isolated Digit ASR: Towards Application in a Disaster Participatory Toolkit
Emmanuel Malaay, Ronald John Cabatic, Michael Simora, Shrestha Mohanty, Justin Mi, Jonathan Lee, Thanatcha Panpairoj, Sirej Dua, Brandie Nonnecke, Camille Crittenden, Ken Goldberg, Nathaniel Oco, and Rachel Edita Roxas

Poster 9: Utilizing Indonesian Data Resources for Text-to-Speech Using End-to-End Method
Agung Santosa, Asril Jarin, Made Gunawan, Teduh Uliniansyah, Gunarso, Elvira Nurfadhilah, Lyla Ruslana, Fara Ayuningtyas, Harnum Annisa, and Hammam Riza

12:00 - 13:30 [Lunch] *Country representatives meeting at SHUNRIN (3rd floor)

Keynote 2 *Chair: Nobuaki Minematsu*

13:30 - 14:30 Aspects of L2 Learners’ English Speeches: A Study Based on the ICNALE
Shinichiro Ishikawa

14:30 - 14:40 [Break]

Session 7 (ANALYSIS) *Chair: Hammam Riza*

14:40 - 15:05 Phonetic Realization of Information Structures in Chinese English Learners' Reading Texts
Xinyi Wen, Yuan Jia, and Aijun Li

15:05 - 15:30 Examining the Influence of Word Tonality on Pitch Contours when Singing in Mandarin
Yi-Jhe Lee, Bang-Yin Chen, Yun-Ting Lai, Hsueh-Wei Liao, Ting-Chun Liao, Sheng-Lun Kao, Kuan-Yi Kang, Chun-Tang Hsu, and Yi-Wen Liu

15:30 - 15:55 Tonal Target and Peak Delay in Mandarin Neutral Tone
Aijun Li, Zhiqiang Li, Gan Huang, and Liang Zhang

15:55 - 16:05 [Break]

Country Report

16:05 - 17:35

China	*Aijun Li (Chinese Academy of Social Sciences),* *Dong Wang (Tsinghua University)*
Hong Kong	*Tan Lee (The Chinese University of Hong Kong)*
India	*Shyam S. Agrawal (KIIT College of Engineering)*
Indonesia	*Hammam Riza (BPPT)*
Japan	*Satoshi Nakamura (Nara Institute of Science and Technology)*
Korea	*Yong-Ju Lee (SiTEC, Wonkwang University)*
Myanmar	*Win Pa Pa (University of Computer Studies)*
Pakistan	*Tania Habib, Sarmad Hussain* *(University of Engineering and Technology Lahore)*
Philippine	*Nathaniel Oco, Leif Romeritch Syliongka, Rachel Edita Roxas* *(National University)*
Singapore	*Haizhou Li (National University of Singapore)*
Taiwan	*Sin-Horng Chen (National Chiao Tung University),* *Chiu-yu Tseng (Academia Sinica)*
Thailand	*Chai Wutiwiwatchai (NECTEC)*
Timor Leste	*Borja L. C. Patrocinio Antonino (Universidade Nacional de Timor Leste)*
Vietnam	*Luong Chi Mai (Vietnam Academy of Science and Technology)*

Closing

17:35 - 17:45 Closing remarks

Keynote 1

Dr. Odette Scharenborg

*Speech Researcher, M*Modal*

Title

Linguistic Unit Discovery from Multi-modal Inputs in Unwritten Languages

Abstract

Automatic speech recognition (ASR) technologies require a large amount of annotated data for a system to work reasonably well. For many languages in the world, though, not enough speech data is available, or these lack the annotations needed to train an ASR system. In fact, it is estimated that for only about 1% of the world languages the minimum amount of data that is needed to train an ASR is available. The “Speaking Rosetta” JSALT 2017 project laid the foundation for a new research area “Unsupervised multi-modal language acquisition”. It showed that it is possible to build useful speech and language technology (SLT) systems without any textual resources in the language for which the SLT is built, in a way that is similar to that of how infants learn a language. I will present a summary of the accomplishments of the multi-disciplinary “Speaking Rosetta” workshop exploring the computational and scientific issues surrounding the discovery of linguistic units in a language without orthography. I will focus on our efforts on 1) unsupervised discovery of acoustic units from raw speech, and 2) building language and speech technology in which the orthographic transcriptions were replaced by images and/or translated text in a well-resourced language.

Biography

Odette Scharenborg (PhD) is a speech researcher at M*Modal. Previously she was an associate professor at the Centre for Language Studies, Radboud University Nijmegen, The Netherlands, and a research fellow at the Donders Institute for Brain, Cognition and Behaviour at the same university. She is interested in the question where the difference between human and machine recognition performance originates, and whether it is possible to narrow this difference, and investigates these questions using a combination of computational modelling and behavioural experimentation. In 2008, she co-organised the Interspeech 2008 Consonant Challenge, which aimed at promoting comparisons of human and machine speech recognition in noise. She was one of the initiators of the EU Marie Curie Initial Training Network “Investigating Speech Processing In Realistic Environments” (INSPIRE, 2012-2015). In 2017, she led a 6-weeks Frederick Jelinek Memorial Summer Workshop on Speech and Language Technology on the topic of the automatic discovery of grounded linguistic units for languages without orthography. In 2017, she was elected onto the board of the International Speech Communication Association (ISCA).

Keynote 2

Dr. Shinichiro Ishikawa

Professor, Kobe University

Title

Aspects of L2 Learners' English Speeches: A Study Based on the ICNALE

Abstract

The ICNALE (International Corpus Network of Asian Learners of English) is one of the largest learner corpora ever built and it includes more than 10,000 essays and speeches produced by varied L2 learners in Asia. In this talk, I will outline the ICNALE project with a focus on two speech modules: The ICNALE Spoken Dialogue and The ICNALE Spoken Monologue.

Biography

Dr. Shin'ichiro (Shin) Ishikawa is Professor of Applied Linguistics at the School of Languages & Communication, Kobe University, Japan. His research interests cover corpus linguistics, statistical linguistics, TESOL, and SLA. He has published many academic papers and books on branches of applied linguistics. He is a leader in the ICNALE learner corpus project.

DAY-1 May 7, 2018

Session 1 (CORPUS)
9:40 - 10:55
Chair: Chai Wutiwiwatchai

Development of Text and Speech Corpus for Designing the Multilingual Recognition System

Shweta Bbansal and Shyam S. Agrawal

To create the multilingual speech and text corpus manually is very difficult and time-consuming task. This paper presents the overall methodology and experiences of text and speech data collection for three under resourced languages i.e. Hindi, Manipuri and Urdu. The text data collection is done through web crawling in 3 domains i.e. general, news and travel to capture the versatility of database among these languages. The main objective of this project is to collect text and speech database which can be used for training the multilingual spoken language identification systems. In total we collected a text corpus of three million words and audio corpus of 150 speakers (50 native speakers) of each language. Each speaker recorded 300 phonetically rich sentences created through text analysis. The speech utterances were recorded at the rate of 16 kHz through microphone using GOLDWAVE software tool in a sound treated room. The collected speech data sets were annotated manually at phonemic level for each language and made available for development of multilingual recognition system.

Urdu Speech Corpora for Banking Sector in Pakistan

Benazir Mumtaz, Sahar Rauf, Hafsa Qadir, Javairia Khalid, Tania Habib, Sarmad Hussain, Rukhsana Barkat, and Ehsan Ul Haq

This research describes an effort to build Urdu speech corpora for the banking sector in Pakistan. We have designed speech corpora to develop debit card activation ASR and these corpora are comprised of eight types of corpora mainly debit card number corpus, expiry date corpus, last four digit corpus, months' name, date of birth corpus, account type and Urdu-counting corpus. These corpora contain telephone speech in read style obtained from more than 400 speakers specifically in Punjabi accent in both outdoor and indoor environments, including offices, homes, banks, and universities. The speech is automatically annotated and manually verified at sentence tier and reports 98% inter-annotator accuracy. In this paper, we report the design, recording and annotation process of speech corpora that serve as a data development step for ASR, and will be integrated in debit card activation service in banking sector of Pakistan.

AWA Long-Term Recorded Speech Corpus and Robust Speaker Recognition Method for Session Variability

Satoru Tsuge, Shingo Kuroiwa, Tomoko Ohsuga, and Yuichi Ishimoto

Session variability is one of the most important issues in the speaker recognition technology. On the other hand, our scientific interest lies in how individual voice changes as time progresses and where the limit of the changes. From these motivations, we have been constructing “AWA Long-Term Recorded speech corpus (AWA-LTR)” that contains one’s same content speech recorded at morning, noon, and evening once a week for over 10 years using the same microphone in a soundproof chamber.

AWA-LTR first version has been released by Speech Resources Consortium, National Institute of Informatics (NII-SRC), Japan in 2012. In addition, we will release AWA-LTR second version in 2018. Hence, in this paper, we describe the details of AWA-LTR and the data release schedule of this corpus. As an effective application example using the corpus, we propose a robust speaker recognition method for session variability and evaluate the proposed method by the speaker identification experiment in this paper.

Session 2 (ANALYSIS)

11:05 - 11:55
Chair: Aijun Li

Parenthetical – A Special Type of Prosodic Reduction in Continuous Speech

Chiu-yu Tseng, Helen Kai-yun Chen, and Yen-Hsing Chen

The current study investigates parenthetical, a type of prosodic reduction in multi-phrase speech paragraphs. Structurally a modifier of its antecedent to provide supplementary information, such reduction creates a lower level in the prosodic hierarchy nested within a discourse-prosodic unit. Perceptual annotation of parenthetical turned out to be consistent across listeners; their acoustic profiles distinctive. Further calculation of information density in relation to allocation of perceived emphasis also demonstrates that parenthetical triggered prosodic reductions are patterned and accountable. Therefore, in spite of low information standing, their existence in the prosodic hierarchy helps facilitate more precise information expression. In sum, current evidence illustrates how information planning is manifested via both emphases and reductions in global context prosody, why parenthetical caused reductions should be understood from a hierarchical perspective within speech context, and how prosodic reduction also plays a crucial role in contributing to comprehensive understanding toward context prosody.

Acoustic Comparison of Vowel Articulation When Combined with Different Tone Categories in Mandarin

Chong Cao, Yanlu Xie, and Jinsong Zhang

It was found that there existed an interaction between the source (i.e., fundamental frequencies) and the vocal tract filter (i.e., formant frequencies). Previous studies investigated such interaction from a perspective of perception with evidence from Mandarin which uses four tones to distinguish lexical meanings. While few studies examined such interaction from a perspective of production. This study explored differences of formant frequencies in vowel articulation when combined with different fundamental frequency patterns (i.e., tones). We calculated frequencies of the first two formants (i.e., F1, F2) and their distance (i.e., F2-F1) of different vowels with four lexical tones. Results showed that both F1 and F2 values were significantly different when combined with different tones. Moreover, such interaction varied with vowels: high vowels usually presented a contrary correlation pattern compared with other vowels. The finding about the co-variation between formants and fundamental frequencies may help to improve the naturalness of speech synthesis.

Session 3 (MODEL & SYSTEM)

14:45 - 16:05

Chair: Haizhou Li

Naso-Articulometry Speech Database for Cleft-Palate Speech Assessment

Chai Wutiwiwatchai, Patcharika Chootrakool, Sawit Kasiriya, Kalyanee Makarabhirom, Nantiya Ooppanasak, and Benjamas Prathanee

Cleft palate has impact to speech, language and hearing problems. Speech therapy is a common treatment process required after surgery. To improve the assessment efficiency in patient with nasalance and articulation disorders, a novel equipment called *Naso-articulometer* (NASAM) has been introduced to speech and language pathologists (SLP). NASAM has been incrementally developed and used to collect speech data from cleft-palate and normal speakers in word and sentence levels as well as specifically to design for medical assessment. With the proposed new equipment, several issues regarding the assessment process and signal processing are raised to research. This paper was documented the detail of NASAM and speech collection, and addressed important speech processing issues with some preliminary experimental results.

Multi-Modal Multi-Task Deep Learning for Speaker and Emotion Recognition of TV-Series Data

Sashi Novitasari, Quoc Truong Do, Sakriani Sakti, Dessi Lestari, and Satoshi Nakamura

Since paralinguistic aspects must be considered to understand speech, we construct a deep learning framework that utilizes multi-modal features to simultaneously recognize both speakers and emotions. There are three kinds of feature modalities: acoustic, lexical, and facial. To fuse the features from multiple modalities, we experimented on three methods: majority voting, concatenation, and hierarchical fusion. The recognition was done from TV-series dataset that simulate actual conversations.

Mathematical Modeling for Daniel – IPA Vowel System in CSOLP and Actual Application in Quantitative & Dynamic Research on IPA Diphthongs

Qiaoli Feng, Xuan Xiong, Wei He, Ziyu Ye, Min Yu, and Xiaogan Huang

Based on the discovery of the original phoneme /o/ and its main natures, we established the Coordinate System of Linguistic Phonemes (CSOLP). CSOLP can provide a mathematical platform for quantitative and dynamic research on speech sounds [10]. The cardinal vowel system, devised by the famous British phonetician Daniel Jones, was the theoretical basis of the IPA vowel system. As we know, phonetic research based on the IPA vowel diagram could be only qualitative. Taking the original phoneme /o/ as a reference point, this paper attempts to translate Daniel's four basic fixed extreme vowels to the CSOLP and determine their coordinates in it, then constructs a mathematical model for the IPA vowel system to achieve an approach to study linguistic phonemes in natural articulating settings quantitatively and dynamically.

As an actual application paradigm, we set up a mathematical model for the IPA diphthong system with the help of the mathematical model in CSOLP, and illustrate quantitative and dynamic analyses in the pronunciation process of English diphthongs in natural manner.

Session 4 (SECOND LANGUAGE)
16:15 - 17:30
Chair: Sin-Horng Chen

A Typological Study of English Monophthongs Acquisition of EFL Learners in Shandong Dialect Area Region

Yuan Jia, Bin Li, and Aijun Li

This paper aims to investigate the joint effect of dialect and Mandarin on Shandong (SD) English learners' vowel production from a typological perspective. We focus on the acoustic features of English vowels produced by learners from Jinan (JN), Jining (JNI), Weifang (WF) and Yantai (YT), in comparison with those produced by American English speakers. Ten English monophthongs and three similar vowels are selected as target samples and their corresponding F1 and F2 formants are employed as parameters in the study. Specifically, the results of the three similar vowels show that: /i/ is more affected by dialect for all the cities; /u/ is more affected by dialect for JN and WF learners, while for JNI and YT learners, /u/ is closer to Mandarin than to the dialect and /a/ produced by SD learners is similar to that of American speakers. Further, the Speech Learning Model (SLM) is employed to explain the analysis results.

Typology of Convergences and Divergences of English Monophthongs by Chinese Northeastern EFL Learners

Yuan Jia and Yu Wang

The present paper investigates the acoustic features of English vowels by EFL learners (English as a Foreign Language) from Dalian (DL) and Harbin (HRB) dialectal regions, both of which belong to the Chinese Northeastern area. Eleven English monophones, i.e., /i/, /u/, /a/ etc. are selected as target samples and their corresponding F1&F2 formants are employed as parameters to approach the research aim. Through analyzing the acoustic results, this paper focuses on exploring the degree of phonetic transfer of dialects (L1) onto English (L2). The Speech Learning Model (SLM) is adopted to examine the differences caused by the dialectal accent. The results show that, with regard to the tongue position of vowels, EFL learners from these two dialectal regions do show a great divergence from the American (AM) native speakers. As for DL learners, it is difficult for them to make tense-lax contrasts in /i/-/ɪ/, /ɛ/-/æ/ and /u/-/ʊ/. Specifically, /i/ and /u/ are affected by DL dialect, which can be explained by SLM. On the other hand, /ɑ/ produced by DL and HRB learners is similar to that of American speakers. Besides, DL and HRB learners produce longer vowels in duration.

Acoustic Features of Mandarin Diphthongs by Uyghur Learners at Primary Level

Yultuz Rapkat, Gulnur Arkin, Mijit Ablimit, and Askar Hamdulla

From the perspective of experimental phonetics, this paper makes an acoustic comparison analysis of the diphthongs Uyghur and Chinese college speakers, and examines the situation of primary-level Uyghur learners' acquisition of Chinese Mandarin diphthongs. A total of 132 samples (including 9 diphthongs) are extracted from the recorded corpus, and the formants of the vowel are statistically analyzed. The characteristics and the distributions of the formants are analyzed to investigate the acoustic characteristics. Finally, combined with the experimental results, the Uyghur learners' at primary level acquisition of diphthongs will be further discussed and analyzed. The purpose of this paper is to understand the Uyghur college learners' acquisition of Chinese Mandarin diphthongs tracks and to provide the correct reference data for the Computer Assisted Language Learning System of Uyghur Learning Chinese Mandarin.

DAY-2 May 8, 2018

Session 5 (CORPUS)
9:00 - 10:15
Chair: Luong Chi Mai

Japanese-English Code-Switching Speech Data Construction

Sahoko Nakayama, Takatomo Kano, Quoc Truong Do, Sakriani Sakti, and Satoshi Nakamura

As the number of Japanese-English bilingual speakers continues to increase, code-switching phenomena also happen more frequently. The units and locations of switches may vary widely from single word switches to whole phrases (beyond the length of the loanword units). Therefore, speech recognition systems must be developed that can handle not only Japanese or English but also Japanese-English code-switching. Consequently, a large-scale code-switching speech database is required for model training. But collecting natural conversation dialogues of Japanese-English data is both time-consuming and expensive. This paper presents the construction of Japanese-English code-switching speech data by utilizing a Japanese and English text-to-speech system from a bilingual speaker. Various switching units are also investigated including units of words and phrases. As a result, we successfully constructed over 280-k speech utterances of Japanese-English code-switching.

Speech Corpora of Under Resourced Languages of North-East India

Barsha Deka, Joyshree Chakraborty, Abhishek Dey, Shikhamoni Nath, Priyankoo Sarmah,
S.R. Nirmala, and Samudravijaya K

In this paper, we present an account of an ongoing effort in creation of speech corpora of under-resourced languages of North-East India, namely, Assamese, Bengali and Nepali. The speech corpora are being created for development of Automatic Speech Recognition system in Assamese as well as for Language Identification system. The text corpus of Assamese language comprises of 1000 sentences collected from different sources such as story books, novels, proverbs. Speech data are recorded over telephone channel using an interactive voice response system. Speakers were asked to read one or more sets of sentences, each set containing 20 sentences. Speech was simultaneously recorded using a hand-held audio recorder. While significant amount of speech data has been collected for Assamese language, the task has begun for Bengali, Nepali and English spoken by native speakers of these 3 languages. Currently, the Assamese speech database contains more than 5000 utterances by 27 native speakers. Information about the speakers such as dialect, gender, age-group were also collected. We discuss the methodology used in collecting speech samples, and present a descriptive statistics of the speech corpora.

Unsupervised Dependency Corpus Annotation for Myanmar Language

Hnin Thu Zar Aye, Win Pa Pa, and Ye Kyaw Thu

Dependency parsing can provide the connection of linguistic unit (words) by a directed links. This paper presents annota-ting a general domain corpus by using unsupervised approach by applying Universal part-of-speech (U-POS) to build Treebank for unsupervised dependency parsing of Myanmar Language. Up to now it is still hard task to obtain complete syntactic structures for Myanmar Language. Dependency structures of words in Myanmar sentences are also presented of general words and phrases orders and the relations of basic sentence structures. To annotate by using U-POS, UDPipe is used. Moreover, the preliminary results of annotated trees and parsing experiment are presented. Parsing experiments are evaluated by UDPipe in terms of unlabeled and labeled attachment scores: (UAS) and (LAS), which are 93.20%, and 91.21% in test experiment respectively.

Session 6 (POSTERS)

10:45 - 12:00

Chair: Shyam S. Agrawal

P1: The BLCU-SAIT Speech Corpus of Non-Native Chinese

Wei Wang, Xing Wei, Jiawei Yu, Wei Wei, Yanlu Xie, and Jinsong Zhang

A computer-aided pronunciation training (CAPT) system with instructive feedback can hardly develop without large-scale non-native speech corpus with refined error annotation. This paper describes such a corpus: BLCU-SAIT speech corpus, with Chinese as a purpose language. This corpus is composed of four sections which cover most of the Chinese phoneme types and tri-tone types bounded by prosodic boundary using a 103 sentence set. The first phase has been completed with a collection of 302 non-native speakers' data with totally 66% language family of the world, and various proficiency levels were recruited from universities in Beijing and Urumqi. Paper also describes part of the annotation projects in this corpus. There are 156 speakers' bi-syllable were manually labeled to pick out phoneme errors. The total error rate of those data is 16.6%, and abundant error pattern can be found from the annotation data.

P2: An Enhancement of English-Thai Pronunciation Dictionary

Patcharika Chootrakool, Sittipong Saychum, Chai Wutiwiwatchai, and Anocha Rugchatjaroen

This paper presents an enhancement of the capacity of an English-Thai pronunciation dictionary. Thai is a tonal language with syllable-based pronunciation. The proposed dictionary design adds the field of Thai transliterated words with pseudo-syllable boundaries into the traditional pronunciation dictionary field list which consists of English words, English pronunciation, and Thai pronunciation. The words are collected from everyday use. It contains 8,268 words extracted from the news and web boards, and 14,414 words from places of interest such as restaurants, etc. The proposed design aims to increase the performance of pseudo-syllable segmentation and grapheme-to-phoneme conversion tools; hence it was used preliminarily to train and test the pseudo-syllable boundary prediction and grapheme to phoneme conversion system which it obtained with 94.7% accuracy; this can be improved by increasing the dictionary size.

P3: Assessment of Korean Spontaneous Speech Produced by Non-Native Learners: Issues and Methodology

Seung Hee Yang and Minhwa Chung

Manual evaluation of non-native speech is not only valuable data for researches in language acquisition, but is also important for developing an automatic evaluation system. Previous experiments surveyed the criteria for non-native Korean speech evaluation based on literature reviews and four raters evaluated 10,000 utterances produced by fifty learners. However, it was restricted to read speech domain, and its generalizeability to spontaneous speech still remains questionable. This study aims to extend the previous work by proposing new evaluation criteria and conducting evaluation for spontaneous speech. In order to do so, we compare the experiment designs in previous studies. After analyzing corpus characteristics, we define eight evaluation criteria, consisting of six analytical and two holistic criteria. Three native raters of Korean evaluated 2,000 spontaneous speech utterances produced by 50 learners. Not only the evaluation is more comprehensive in scope than previous experiments on Korean spontaneous speech evaluation, we have documented possible issues that may arise in spontaneous speech with specific examples, and how we established detailed guidelines for each case. The research result shows that proficiency has higher correlation with comprehensibility than segmental accuracy. This work will serve as a preliminary work for developing an automatic assessment model for non-native Korean in language learning applications.

P4: Acoustic Feature Analysis on the Chinese Mandarin Monophthongs Pronounced by Kazakh College Students

Guljan Alijan, Gulnur Arkin, Dilmurat Tursun, Mijit Ablimit, and Askar Hamdulla

In order to provide reliable acoustic parameters for the Chinese oral processing basic vowel system, this article experimental analyzed the first two formant frequency and duration parameters of the unitary tones of 15 Kazakh Chinese learners and 10 Mandarin native speakers and compared its similarity, taking comparative analysis of hypothesis and phonetic learning model as the theoretical basis, from the perspective of experimental phonetics for the first time, and the differences and characteristics of vowel pronunciation of the Kazakh Chinese learners and Mandarin native speakers were summarized. As shown by experiment results, Kazakh learners generally encounter difficulty in monophthong pronunciation when leaning mandarin as the difference between two languages in phonetic system, intonation, accent and other features exerts influence on learning of acoustic characteristics in second language. This study first verifies systematic difference in monophthong pronunciation between Kazakh learners of Chinese language and native speaker of mandarin relying on actual experimental data. And the study findings will surely place great reference value on high-natural parameters of the synthetic and high-precise language identification aiming at mandarin learning by Kazakhs.

P5: Research on Patterns of Unvoiced Fricatives in Uyghur Language

Parizat Keyim, Gulnur Arkin, Mijit Ablimit, and Askar Hamdulla

This paper starts from the text analysis module, and uses the “Uyghur speech acoustics parameters Database” to make a statistical analysis about the acoustic parameters of five affricates [f], [s], [ʃ], [x] and [ɦ] at different positions of the words and regularity of distribution in the words, and sum up their distribution modes of formant, sound intensity, duration. At the same time, statistical analysis is made on the fricative resonance peaks of the initial syllable and the final syllable, respectively. This paper also summarizes the fricative pattern of the initial and final syllable and analyzes the difference between them in detail, then draws the formant pattern of Uighur fricatives. The aim is to provide a natural language regularity mode and scientific basis for Uyghur language teaching, scientific research, phonetic synthesis, phonetic recognition, corpus information processing, and so on.

P6: A High Quality and Phonetic Balanced Speech Corpus for Vietnamese

Phuong Pham Ngoc, Quoc Truong Do, and Luong Chi Mai

This paper presents a high quality Vietnamese speech corpus that can be used for analyzing Vietnamese speech characteristic as well as building speech synthesis models. The corpus consists of 5400 clean-speech utterances spoken by 12 speakers including 6 males and 6 females. The corpus is designed with phonetic balanced in mind so that it can be used for speech synthesis, especially, speech adaptation approaches. Specifically, all speakers utter a common dataset contains 250 phonetic balanced sentences. To increase the variety of speech context, each speaker also utters another 200 non-shared, phonetic-balanced sentences. The speakers are selected to cover a wide range of age and come from different regions of the North of Vietnam. The audios are recorded in a soundproof studio room, they are sampling at 48 kHz, 16 bits PCM, mono channel.

P7: Acoustic Analysis of Vowels in Two Southern Angami Dialects

Viyazonuo Terhiija, Priyankoo Sarmah, and Samudravijaya K

This paper describes the acoustic analysis of vowels in two southern Angami dialects and the creation of the database. While, attention has been given to the standard variety of Angami, studies on its varieties are few in number. The goal of the paper is to study the segmental features of the dialects and account for the variations that exist. Kigwema and Viswema are considered to be the oldest villages of the Angamis to have established in the Kohima district. Numerous villages have branched out from these two ancestral villages and there are varying degrees of mutual intelligibility among the villages. This paper explores the variation that exists among two dialects originating from the Kigwema and Viswema villages.The findings of the study show that dialectal variation is well represented by vowel variations in the two dialects.

P8: Noise-Resistant Telephone Quality Isolated Digit ASR: Towards Application in a Disaster Participatory Toolkit

Emmanuel Malaay, Ronald John Cabatic, Michael Simora, Shrestha Mohanty, Justin Mi, Jonathan Lee, Thanatcha Panpairoj, Sirej Dua, Brandie Nonnecke, Camille Crittenden, Ken Goldberg, Nathaniel Oco, and Rachel Edita Roxas

We present our work on developing an isolated digit Automatic Speech Recognizer (ASR) covering 5 languages spoken in the Philippines: Filipino, Ilocano, Cebuano, English, and Spanish-borrowed cardinal numbers. The ASR recognizes quantitative responses for a disaster participatory toolkit called Malasakit (a Filipino term which means “sincere care”). To make the toolkit inclusive, the ASR was designed to be employed in an Interactive Voice Response (IVR) by integrating it in Twilio, a web service API that can receive calls and connect to the Malasakit database. The speech corpus of the ASR was collected from 296 speakers with a total duration of 8 hours, 53 minutes, and 48 seconds which were decimated from a wideband quality (16 kHz) to a telephone quality (8 kHz). To make the ASR noise-resistant, the telephone quality corpus was contaminated with channel and background noises (eg. busy road, marketplace, construction). For future work, researchers are suggested to use these methods for continuous speech recognition using telephone quality speech corpora which can be used towards analyzing qualitative responses.

P9: Utilizing Indonesian Data Resources for Text-to-Speech Using End-to-End Method

Agung Santosa, Asril Jarin, Made Gunawan, Teduh Uliniansyah, Gunarso, Elvira Nurfadhilah, Lyla Ruslana, Fara Ayuningtyas, Harnum Annisa, and Hammam Riza

The Agency for the Assessment and Application of Technology (BPPT), has developed a speech corpus by recording the voices of an adult female and an adult male with 15,645 utterances each. The recorded audio data has duration of approximately 28 hours for female speaker while the male speaker is 30 hours. With increasing popularity of deep learning techniques in natural language processing, we used this data to build an experimental Indonesian text-to-speech (TTS) based on an end-to-end method. Subjective evaluation for naturalness aspect showed a score of 3.5 for male voice and 4 for female voice. For intelligibility aspect, the accuracy was 64.63% for male voice and 71.28% for female voice. Objective evaluation represented by RMSE value of fundamental frequency f0 was 0.514 for male voice and 0.515 for female voice. This experiment is part of ongoing development for achieving the best TTS for Bahasa Indonesia.

Session 7 (ANALYSIS)

14:45 - 16:05

Chair: Hammam Riza

Phonetic Realization of Information Structures in Chinese English Learners' Reading Texts

Xinyi Wen, Yuan Jia, and Aijun Li

The present study aims to investigate the phonetic realization of information structure in L2, by comparing the productions of English discourse from Beijing English learners and from native English speakers. Phonetic and statistical analyses are conducted on English reading texts selected from Asian English Speech cOrpus Project (AESOP). The main findings include: Beijing English learners do not distinguish the given and new information with pitch range as native English speakers do, which is the main difference between the two speaker groups; the slight differences found on duration and mean pitch value might result from other factors rather than phonetic strategies utilized in information packaging. Besides, the difference between Beijing English learners' performance in lexical and referential levels mainly lies in the duration of accessible information.

Examining the Influence of Word Tonality on Pitch Contours when Singing in Mandarin

Yi-Jhe Lee, Bang-Yin Chen, Yun-Ting Lai, Hsueh-Wei Liao, Ting-Chun Liao, Sheng-Lun Kao, Kuan-Yi Kang, Chun-Tang Hsu, and Yi-Wen Liu

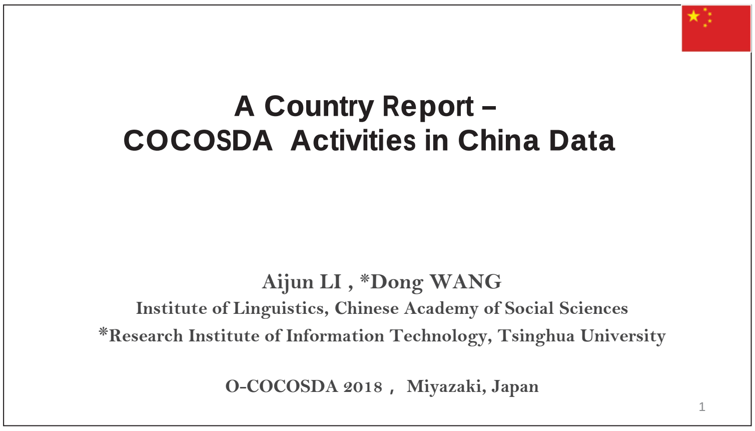
In Mandarin, word meanings are differentiated by tones. Therefore, when a Mandarin song is sung according to its musical melody, word meanings could potentially be misunderstood. In this research, we intend to investigate whether or not a singer would adjust the pitch contour so as to best convey word meanings. A Mandarin singing dataset is currently being manually parsed into single words, phonetics of which are manually transcribed (including the tones), and for each word the pitch contour is calculated by the YIN algorithm. Afterwards, the distance between arbitrary pairs of contours can be calculated by a dynamic time warping-based method. By comparing average same-tone distances with the distances calculated without distinguishing the tones, one can measure the extent to which a singer modifies his/her pitch inflection, consciously or not, according to the actual tone of the word. Some mixed results are reported.

Tonal Target and Peak Delay in Mandarin Neutral Tone

Aijun Li, Zhiqiang Li, Gan Huang, and Liang Zhang

We examined the tonal target of the neutral tone syllable, F0 peak delay and the F0 preplanning process in production of Mandarin neutral tone by manipulating the number of neutral tone syllables and the preceding tonal contexts. The results showed that 1) the tonal target of neutral tone was L; 2) its realization was greatly influenced by the number of neutral tone syllables, as well as the prosodic structure; 3) the F0 pattern of the neutral tone depended on the tonal target of the proceeding non-neutral syllable. Specifically, the interpolation rule realized between the end position of F0 in T1 and T4, or the Peak Delay in T2 and T3, and the target position of neutral tone; and 4) with the increasing number of neutral syllables, the initial F0 in the prosodic unit also increased accordingly, indicating that our pitch preplanning ability was closely related to the prosodic structure.

Country Report - China

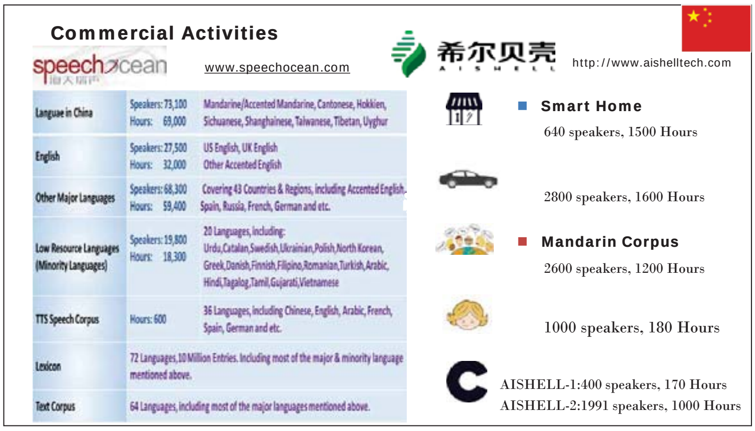


A Country Report – COCOSDA Activities in China Data

Aijun LI , *Dong WANG
Institute of Linguistics, Chinese Academy of Social Sciences
*Research Institute of Information Technology, Tsinghua University

O-COCOSDA 2018， Miyazaki, Japan

1



Commercial Activities

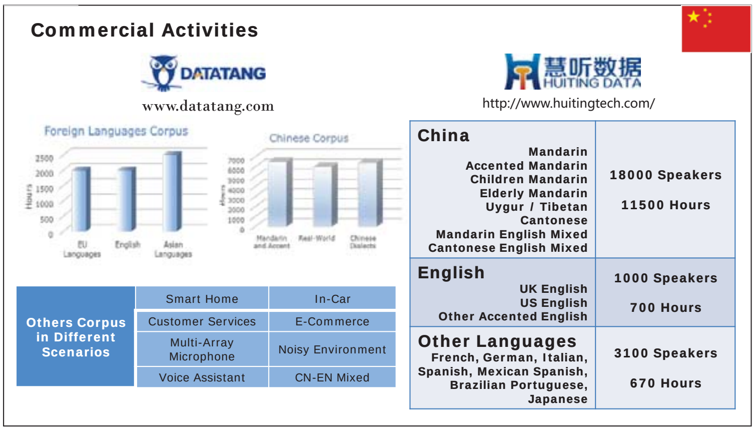
www.speechocean.com

Languae in China	Speakers: 73,100 Hours: 69,000	Mandarine/Accented Mandarine, Cantonese, Hokkien, Sichuanese, Shanghainese, Taiwanese, Tibetan, Uyghur
English	Speakers: 27,500 Hours: 32,000	US English, UK English Other Accented English
Other Major Languages	Speakers: 68,300 Hours: 59,400	Covering 43 Countries & Regions, including Accented English, Spain, Russia, French, German and etc.
Low Resource Languages (Minority Languages)	Speakers: 19,800 Hours: 18,300	20 Languages, including: Urdu,Catalan,Swedish,Ukrainian,Polish,North Korean, Greek,Danish,Finnish,Filipino,Romanian,Turkish,Arabic, Hindi,Tagalog,Tamil,Gujarati,Vietnamese
TTS Speech Corpus	Hours: 600	36 Languages, including Chinese, English, Arabic, French, Spain, German and etc.
Lexicon	72 Languages,10 Million Entries. Including most of the major & minority language mentioned above.	
Text Corpus	64 Languages, including most of the major languages mentioned above.	

希尔贝壳 AISHELL http://www.aishelltech.com

- Smart Home
  640 speakers, 1500 Hours
  2800 speakers, 1600 Hours
- Mandarin Corpus
  2600 speakers, 1200 Hours
  1000 speakers, 180 Hours

AISHELL-1:400 speakers, 170 Hours
AISHELL-2:1991 speakers, 1000 Hours



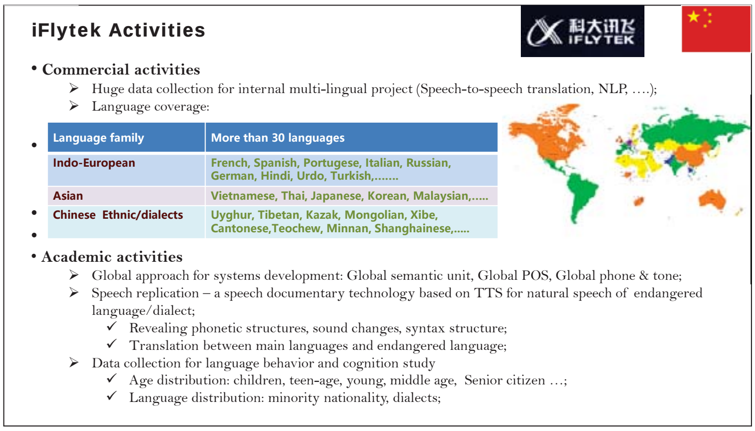
Commercial Activities

DATATANG
www.datatang.com

Others Corpus in Different Scenarios	Smart Home	In-Car
	Customer Services	E-Commerce
	Multi-Array Microphone	Noisy Environment
	Voice Assistant	CN-EN Mixed

慧听数据 HUITING DATA
http://www.huitingtech.com/

China Mandarin Accented Mandarin Children Mandarin Elderly Mandarin Uygur / Tibetan Cantonese Mandarin English Mixed Cantonese English Mixed	18000 Speakers 11500 Hours
English UK English US English Other Accented English	1000 Speakers 700 Hours
Other Languages French, German, Italian, Spanish, Mexican Spanish, Brazilian Portuguese, Japanese	3100 Speakers 670 Hours



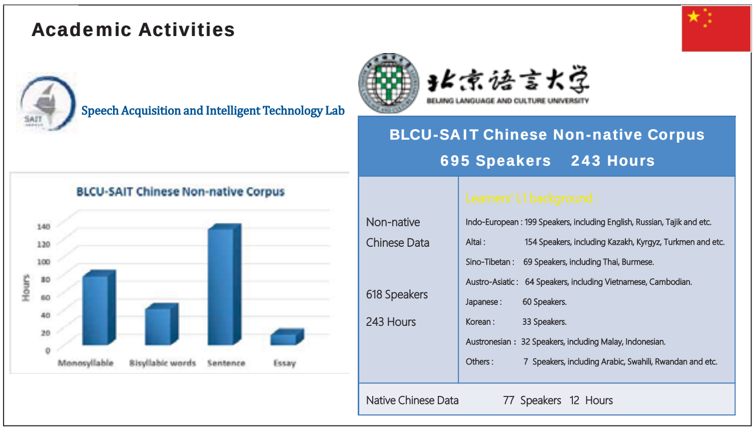
iFlytek Activities

科大讯飞 IFLYTEK

- Commercial activities
  - Huge data collection for internal multi-lingual project (Speech-to-speech translation, NLP, ….);
  - Language coverage:

Language family	More than 30 languages
Indo-European	French, Spanish, Portugese, Italian, Russian, German, Hindi, Urdo, Turkish,…….
Asian	Vietnamese, Thai, Japanese, Korean, Malaysian,…..
Chinese Ethnic/dialects	Uyghur, Tibetan, Kazak, Mongolian, Xibe, Cantonese,Teochew, Minnan, Shanghainese,…..

- Academic activities
  - Global approach for systems development: Global semantic unit, Global POS, Global phone & tone;
  - Speech replication – a speech documentary technology based on TTS for natural speech of endangered language/dialect;
    - ✓ Revealing phonetic structures, sound changes, syntax structure;
    - ✓ Translation between main languages and endangered language;
  - Data collection for language behavior and cognition study
    - ✓ Age distribution: children, teen-age, young, middle age, Senior citizen …;
    - ✓ Language distribution: minority nationality, dialects;



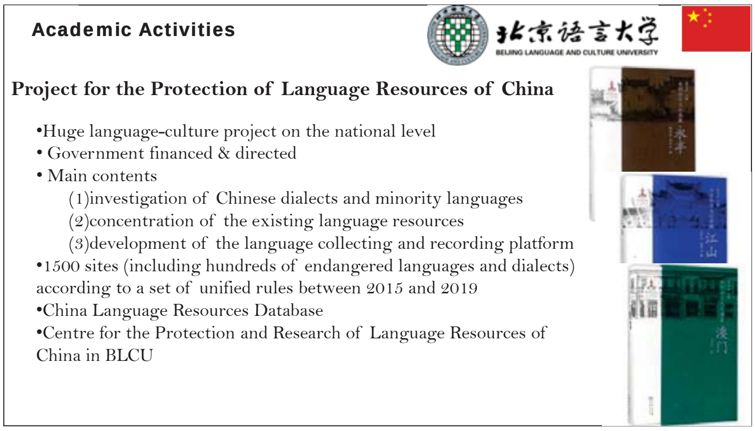
Academic Activities

Speech Acquisition and Intelligent Technology Lab

北京语言大学 BEIJING LANGUAGE AND CULTURE UNIVERSITY

BLCU-SAIT Chinese Non-native Corpus
695 Speakers 243 Hours

Non-native Chinese Data 618 Speakers 243 Hours	Learners' L1 background : Indo-European : 199 Speakers, including English, Russian, Tajik and etc. Altai : 154 Speakers, including Kazakh, Kyrgyz, Turkmen and etc. Sino-Tibetan : 69 Speakers, including Thai, Burmese. Austro-Asiatic : 64 Speakers, including Vietnamese, Cambodian. Japanese : 60 Speakers. Korean : 33 Speakers. Austronesian : 32 Speakers, including Malay, Indonesian. Others : 7 Speakers, including Arabic, Swahili, Rwandan and etc.
Native Chinese Data	77 Speakers 12 Hours



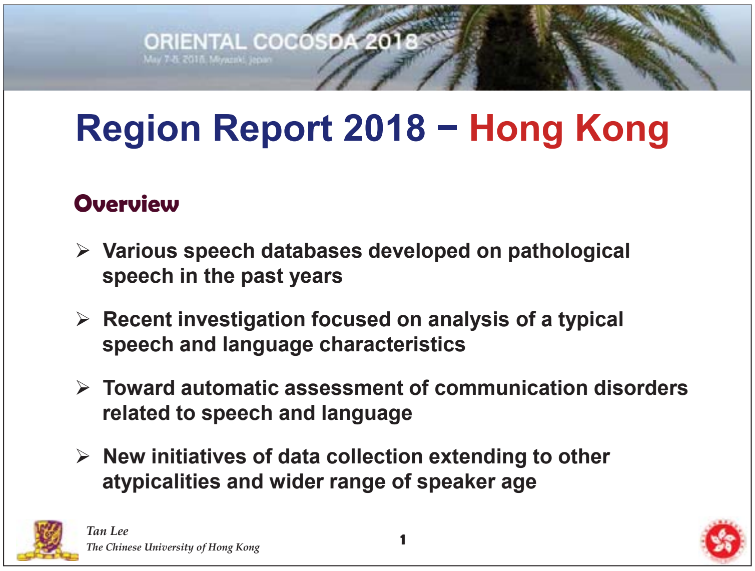
Academic Activities

北京语言大学 BEIJING LANGUAGE AND CULTURE UNIVERSITY

Project for the Protection of Language Resources of China

- •Huge language-culture project on the national level
- Government financed & directed
- Main contents
  - (1)investigation of Chinese dialects and minority languages
  - (2)concentration of the existing language resources
  - (3)development of the language collecting and recording platform
- •1500 sites (including hundreds of endangered languages and dialects) according to a set of unified rules between 2015 and 2019
- •China Language Resources Database
- •Centre for the Protection and Research of Language Resources of China in BLCU

Country Report - Hong Kong

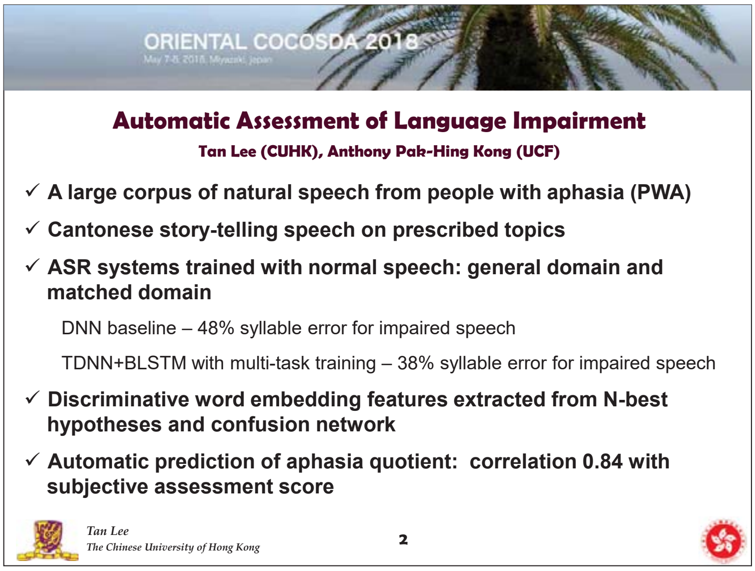

ORIENTAL COCOSDA 2018
May 7-8, 2018, Miyazaki, Japan

Region Report 2018 – Hong Kong

Overview

- Various speech databases developed on pathological speech in the past years
- Recent investigation focused on analysis of a typical speech and language characteristics
- Toward automatic assessment of communication disorders related to speech and language
- New initiatives of data collection extending to other atypicalities and wider range of speaker age

Tan Lee
The Chinese University of Hong Kong
1

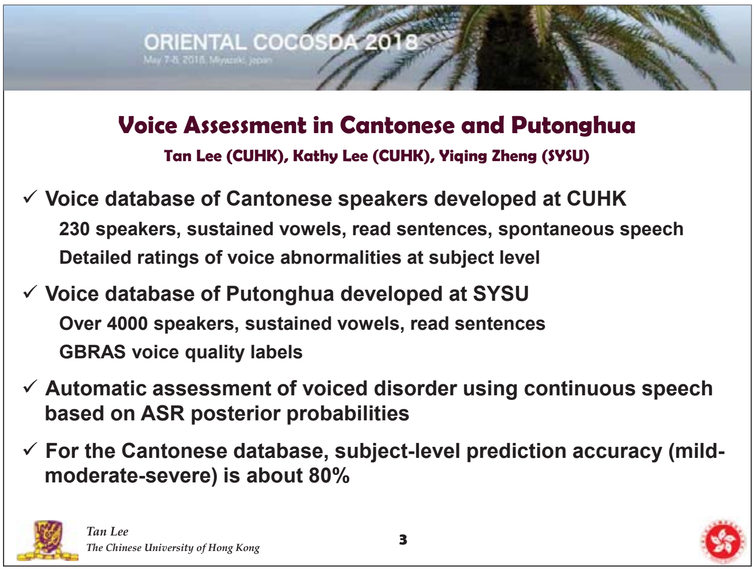

ORIENTAL COCOSDA 2018
May 7-8, 2018, Miyazaki, Japan

Automatic Assessment of Language Impairment

Tan Lee (CUHK), Anthony Pak-Hing Kong (UCF)

- ✓ A large corpus of natural speech from people with aphasia (PWA)
- ✓ Cantonese story-telling speech on prescribed topics
- ✓ ASR systems trained with normal speech: general domain and matched domain
  - DNN baseline – 48% syllable error for impaired speech
  - TDNN+BLSTM with multi-task training – 38% syllable error for impaired speech
- ✓ Discriminative word embedding features extracted from N-best hypotheses and confusion network
- ✓ Automatic prediction of aphasia quotient: correlation 0.84 with subjective assessment score

Tan Lee
The Chinese University of Hong Kong
2

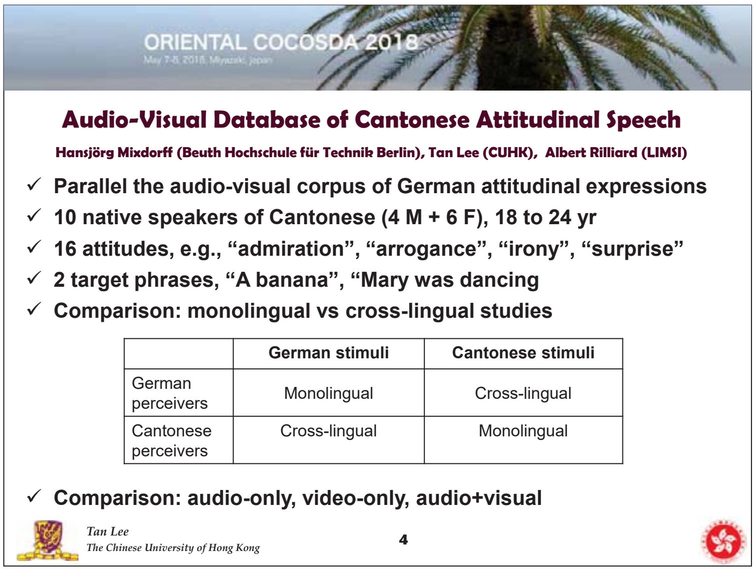

ORIENTAL COCOSDA 2018
May 7-8, 2018, Miyazaki, Japan

Voice Assessment in Cantonese and Putonghua

Tan Lee (CUHK), Kathy Lee (CUHK), Yiqing Zheng (SYSU)

- ✓ Voice database of Cantonese speakers developed at CUHK
  - 230 speakers, sustained vowels, read sentences, spontaneous speech
  - Detailed ratings of voice abnormalities at subject level
- ✓ Voice database of Putonghua developed at SYSU
  - Over 4000 speakers, sustained vowels, read sentences
  - GBRAS voice quality labels
- ✓ Automatic assessment of voiced disorder using continuous speech based on ASR posterior probabilities
- ✓ For the Cantonese database, subject-level prediction accuracy (mild-moderate-severe) is about 80%

Tan Lee
The Chinese University of Hong Kong
3


ORIENTAL COCOSDA 2018
May 7-8, 2018, Miyazaki, Japan

Audio-Visual Database of Cantonese Attitudinal Speech

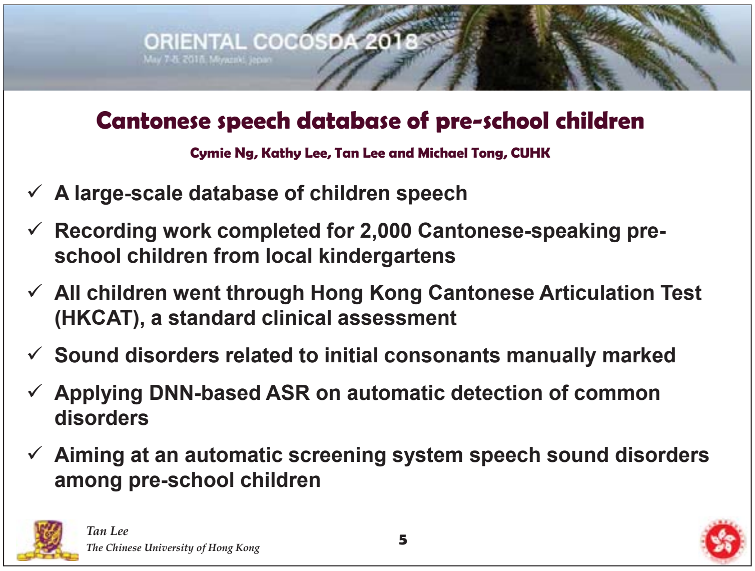
Hansjörg Mixdorff (Beuth Hochschule für Technik Berlin), Tan Lee (CUHK), Albert Rilliard (LIMSI)

- ✓ Parallel the audio-visual corpus of German attitudinal expressions
- ✓ 10 native speakers of Cantonese (4 M + 6 F), 18 to 24 yr
- ✓ 16 attitudes, e.g., “admiration”, “arrogance”, “irony”, “surprise”
- ✓ 2 target phrases, “A banana”, “Mary was dancing
- ✓ Comparison: monolingual vs cross-lingual studies

	German stimuli	Cantonese stimuli
German perceivers	Monolingual	Cross-lingual
Cantonese perceivers	Cross-lingual	Monolingual

- ✓ Comparison: audio-only, video-only, audio+visual

Tan Lee
The Chinese University of Hong Kong
4

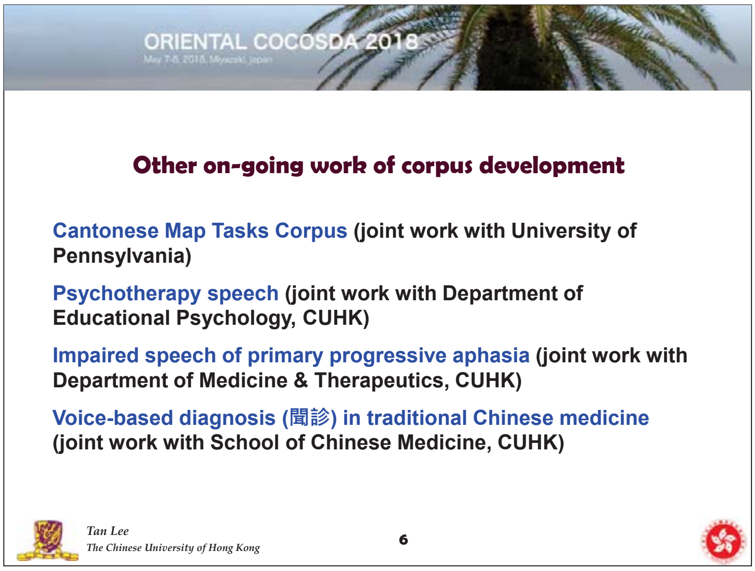

ORIENTAL COCOSDA 2018
May 7-8, 2018, Miyazaki, Japan

Cantonese speech database of pre-school children

Cymie Ng, Kathy Lee, Tan Lee and Michael Tong, CUHK

- ✓ A large-scale database of children speech
- ✓ Recording work completed for 2,000 Cantonese-speaking pre-school children from local kindergartens
- ✓ All children went through Hong Kong Cantonese Articulation Test (HKCAT), a standard clinical assessment
- ✓ Sound disorders related to initial consonants manually marked
- ✓ Applying DNN-based ASR on automatic detection of common disorders
- ✓ Aiming at an automatic screening system speech sound disorders among pre-school children

Tan Lee
The Chinese University of Hong Kong
5


ORIENTAL COCOSDA 2018
May 7-8, 2018, Miyazaki, Japan

Other on-going work of corpus development

Cantonese Map Tasks Corpus (joint work with University of Pennsylvania)

Psychotherapy speech (joint work with Department of Educational Psychology, CUHK)

Impaired speech of primary progressive aphasia (joint work with Department of Medicine & Therapeutics, CUHK)

Voice-based diagnosis (聞診) in traditional Chinese medicine (joint work with School of Chinese Medicine, CUHK)

Tan Lee
The Chinese University of Hong Kong
6

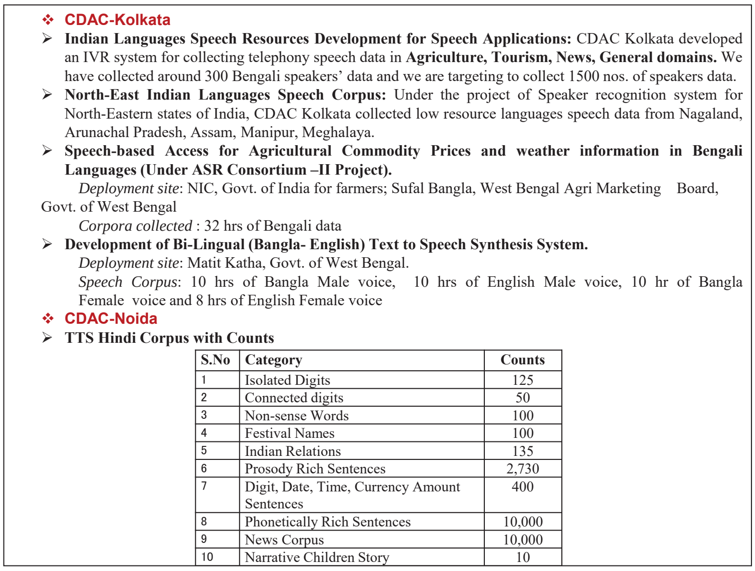
Country Report - India

International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques

Country Report -India

O-COCOSDA 2018

S. S. Agrawal
ss_agrawal@hotmail.com



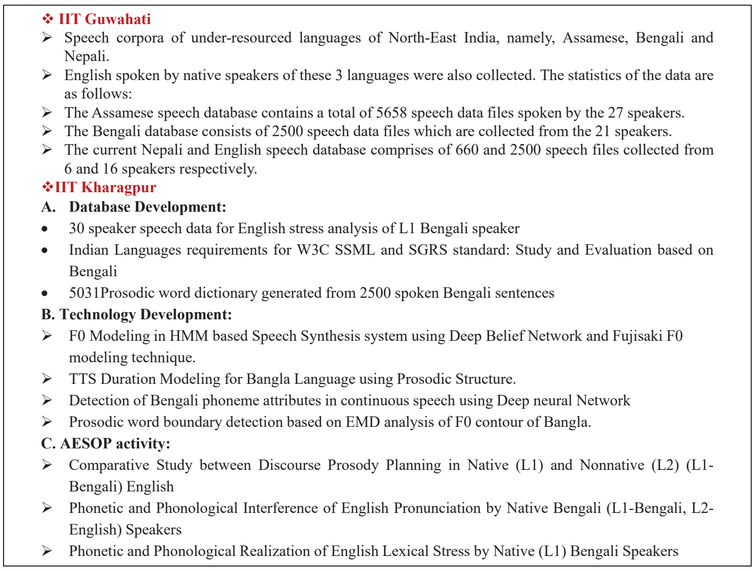
❖ **CDAC-Kolkata**

- ➢ **Indian Languages Speech Resources Development for Speech Applications:** CDAC Kolkata developed an IVR system for collecting telephony speech data in **Agriculture, Tourism, News, General domains.** We have collected around 300 Bengali speakers' data and we are targeting to collect 1500 nos. of speakers data.
- ➢ **North-East Indian Languages Speech Corpus:** Under the project of Speaker recognition system for North-Eastern states of India, CDAC Kolkata collected low resource languages speech data from Nagaland, Arunachal Pradesh, Assam, Manipur, Meghalaya.
- ➢ **Speech-based Access for Agricultural Commodity Prices and weather information in Bengali Languages (Under ASR Consortium –II Project).**
  Deployment site: NIC, Govt. of India for farmers; Sufal Bangla, West Bengal Agri Marketing Board, Govt. of West Bengal
  Corpora collected : 32 hrs of Bengali data
- ➢ **Development of Bi-Lingual (Bangla- English) Text to Speech Synthesis System.**
  Deployment site: Matit Katha, Govt. of West Bengal.
  Speech Corpus: 10 hrs of Bangla Male voice, 10 hrs of English Male voice, 10 hr of Bangla Female voice and 8 hrs of English Female voice

❖ **CDAC-Noida**

- ➢ **TTS Hindi Corpus with Counts**

S.No	Category	Counts
1	Isolated Digits	125
2	Connected digits	50
3	Non-sense Words	100
4	Festival Names	100
5	Indian Relations	135
6	Prosody Rich Sentences	2,730
7	Digit, Date, Time, Currency Amount Sentences	400
8	Phonetically Rich Sentences	10,000
9	News Corpus	10,000
10	Narrative Children Story	10



❖ **IIT Guwahati**

- ➢ Speech corpora of under-resourced languages of North-East India, namely, Assamese, Bengali and Nepali.
- ➢ English spoken by native speakers of these 3 languages were also collected. The statistics of the data are as follows:
- ➢ The Assamese speech database contains a total of 5658 speech data files spoken by the 27 speakers.
- ➢ The Bengali database consists of 2500 speech data files which are collected from the 21 speakers.
- ➢ The current Nepali and English speech database comprises of 660 and 2500 speech files collected from 6 and 16 speakers respectively.

❖**IIT Kharagpur**

A. Database Development:

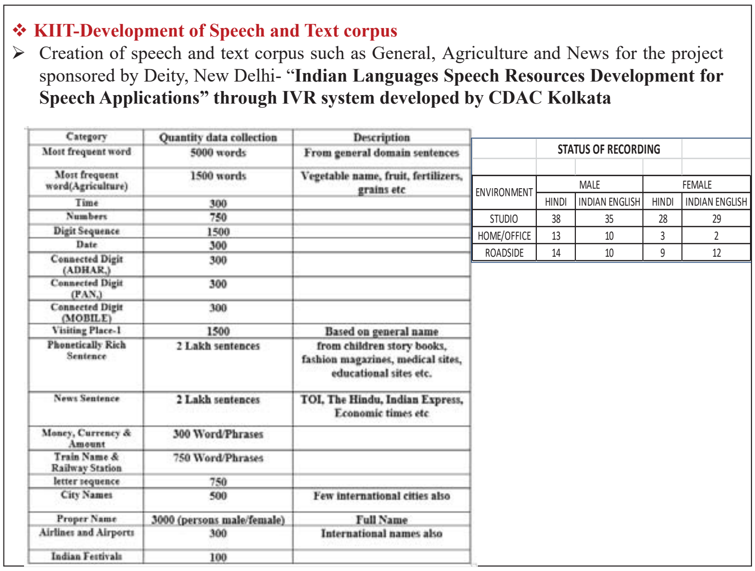
- 30 speaker speech data for English stress analysis of L1 Bengali speaker
- Indian Languages requirements for W3C SSML and SGRS standard: Study and Evaluation based on Bengali
- 5031Prosodic word dictionary generated from 2500 spoken Bengali sentences

B. Technology Development:

- ➢ F0 Modeling in HMM based Speech Synthesis system using Deep Belief Network and Fujisaki F0 modeling technique.
- ➢ TTS Duration Modeling for Bangla Language using Prosodic Structure.
- ➢ Detection of Bengali phoneme attributes in continuous speech using Deep neural Network
- ➢ Prosodic word boundary detection based on EMD analysis of F0 contour of Bangla.

C. AESOP activity:

- ➢ Comparative Study between Discourse Prosody Planning in Native (L1) and Nonnative (L2) (L1-Bengali) English
- ➢ Phonetic and Phonological Interference of English Pronunciation by Native Bengali (L1-Bengali, L2-English) Speakers
- ➢ Phonetic and Phonological Realization of English Lexical Stress by Native (L1) Bengali Speakers



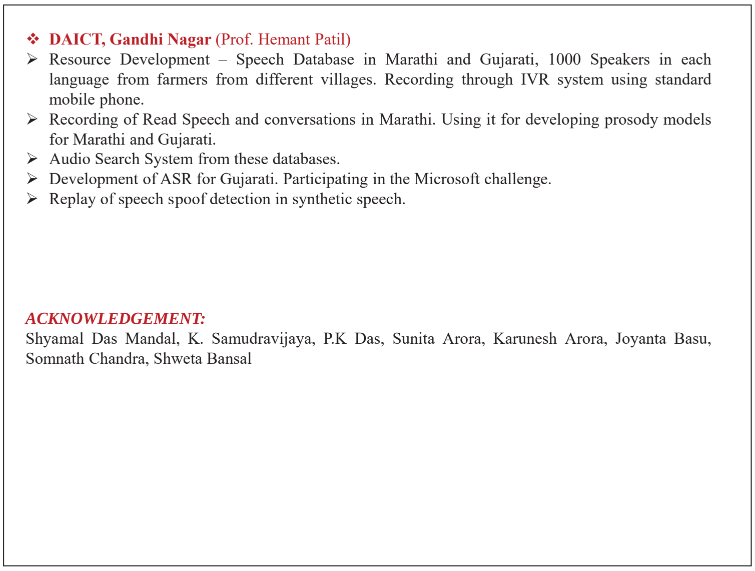
❖ **KIIT-Development of Speech and Text corpus**

- ➢ Creation of speech and text corpus such as General, Agriculture and News for the project sponsored by Deity, New Delhi- "**Indian Languages Speech Resources Development for Speech Applications" through IVR system developed by CDAC Kolkata**

Category	Quantity data collection	Description
Most frequent word	5000 words	From general domain sentences
Most frequent word(Agriculture)	1500 words	Vegetable name, fruit, fertilizers, grains etc
Time	300	
Numbers	750	
Digit Sequence	1500	
Date	300	
Connected Digit (ADHAR,)	300	
Connected Digit (PAN,)	300	
Connected Digit (MOBILE)	300	
Visiting Place-1	1500	Based on general name
Phonetically Rich Sentence	2 Lakh sentences	from children story books, fashion magazines, medical sites, educational sites etc.
News Sentence	2 Lakh sentences	TOI, The Hindu, Indian Express, Economic times etc
Money, Currency & Amount	300 Word/Phrases	
Train Name & Railway Station	750 Word/Phrases	
letter sequence	750	
City Names	500	Few international cities also
Proper Name	3000 (persons male/female)	Full Name
Airlines and Airports	300	International names also
Indian Festivals	100	

STATUS OF RECORDING

ENVIRONMENT	MALE		FEMALE	
	HINDI	INDIAN ENGLISH	HINDI	INDIAN ENGLISH
STUDIO	38	35	28	29
HOME/OFFICE	13	10	3	2
ROADSIDE	14	10	9	12



❖ **DAICT, Gandhi Nagar** (Prof. Hemant Patil)

- ➢ Resource Development – Speech Database in Marathi and Gujarati, 1000 Speakers in each language from farmers from different villages. Recording through IVR system using standard mobile phone.
- ➢ Recording of Read Speech and conversations in Marathi. Using it for developing prosody models for Marathi and Gujarati.
- ➢ Audio Search System from these databases.
- ➢ Development of ASR for Gujarati. Participating in the Microsoft challenge.
- ➢ Replay of speech spoof detection in synthetic speech.

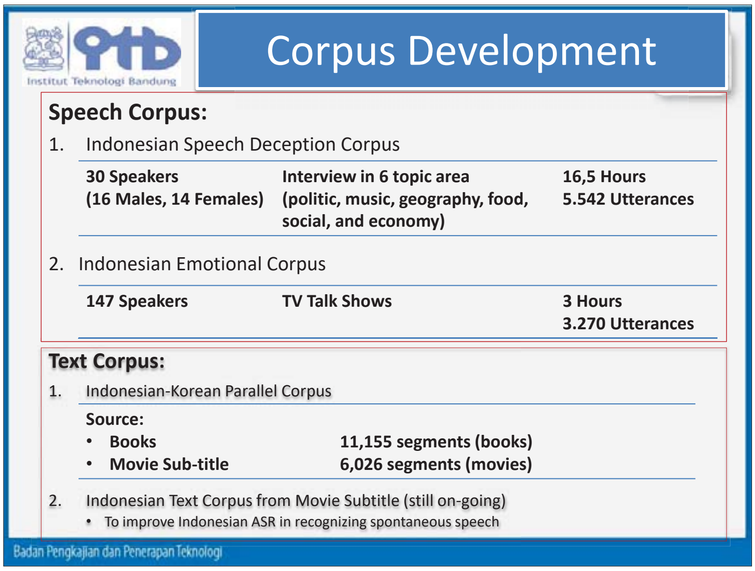
ACKNOWLEDGEMENT:
Shyamal Das Mandal, K. Samudravijaya, P.K Das, Sunita Arora, Karunesh Arora, Joyanta Basu, Somnath Chandra, Shweta Bansal

O-COCOSDA 2018

Indonesia Country Report

Badan Pengkajian dan Penerapan Teknologi

Hammam Riza
Agency for the Assessment and Application of Technology (BPPT)



Corpus Development

Speech Corpus:

1. Indonesian Speech Deception Corpus

30 Speakers (16 Males, 14 Females)	**Interview in 6 topic area (politic, music, geography, food, social, and economy)**	**16,5 Hours 5.542 Utterances**

2. Indonesian Emotional Corpus

147 Speakers	**TV Talk Shows**	**3 Hours 3.270 Utterances**

Text Corpus:

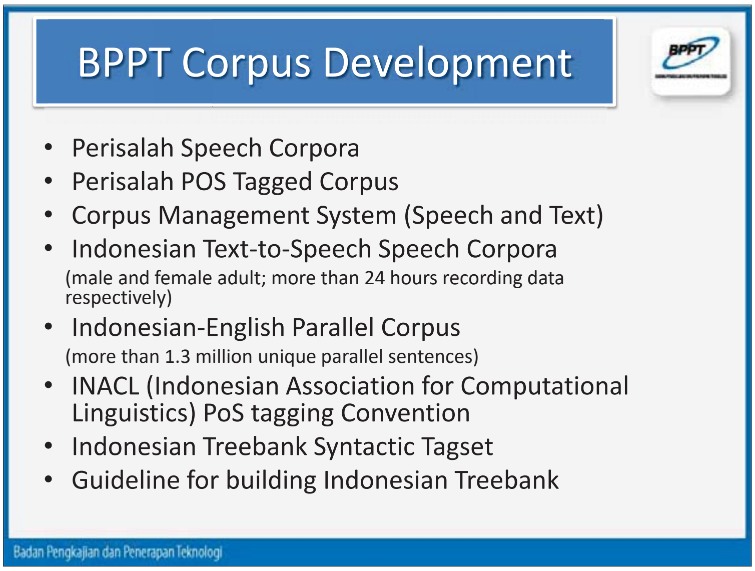
1. Indonesian-Korean Parallel Corpus

Source:

• **Books**	**11,155 segments (books)**
• **Movie Sub-title**	**6,026 segments (movies)**

2. Indonesian Text Corpus from Movie Subtitle (still on-going)
   - To improve Indonesian ASR in recognizing spontaneous speech

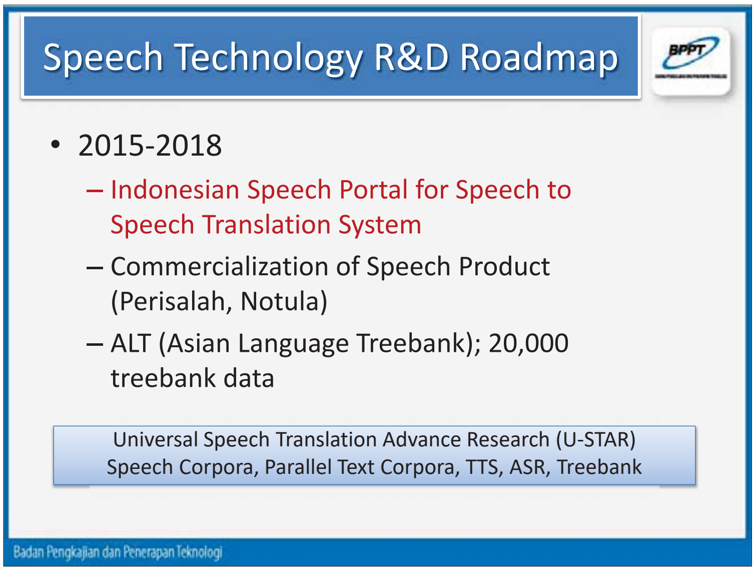
Badan Pengkajian dan Penerapan Teknologi



BPPT Corpus Development

- Perisalah Speech Corpora
- Perisalah POS Tagged Corpus
- Corpus Management System (Speech and Text)
- Indonesian Text-to-Speech Speech Corpora
  (male and female adult; more than 24 hours recording data respectively)
- Indonesian-English Parallel Corpus
  (more than 1.3 million unique parallel sentences)
- INACL (Indonesian Association for Computational Linguistics) PoS tagging Convention
- Indonesian Treebank Syntactic Tagset
- Guideline for building Indonesian Treebank

Badan Pengkajian dan Penerapan Teknologi

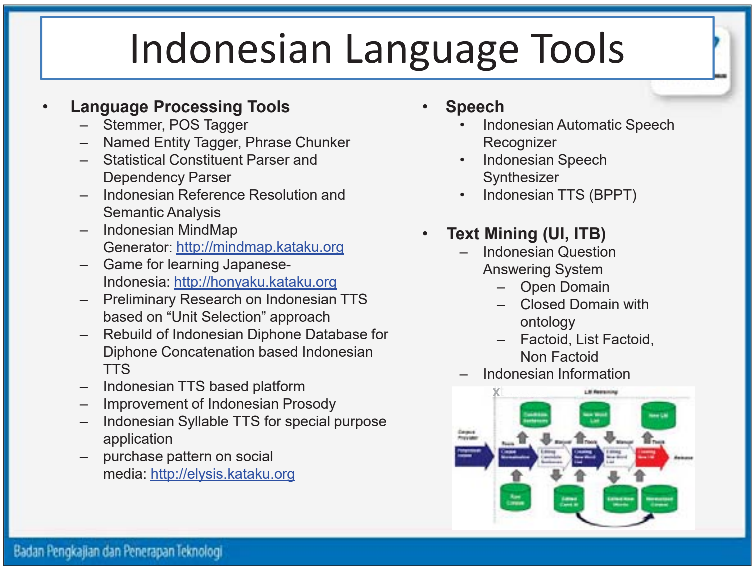


Speech Technology R&D Roadmap

- 2015-2018
  - – Indonesian Speech Portal for Speech to Speech Translation System
  - – Commercialization of Speech Product (Perisalah, Notula)
  - – ALT (Asian Language Treebank); 20,000 treebank data

Universal Speech Translation Advance Research (U-STAR)
Speech Corpora, Parallel Text Corpora, TTS, ASR, Treebank

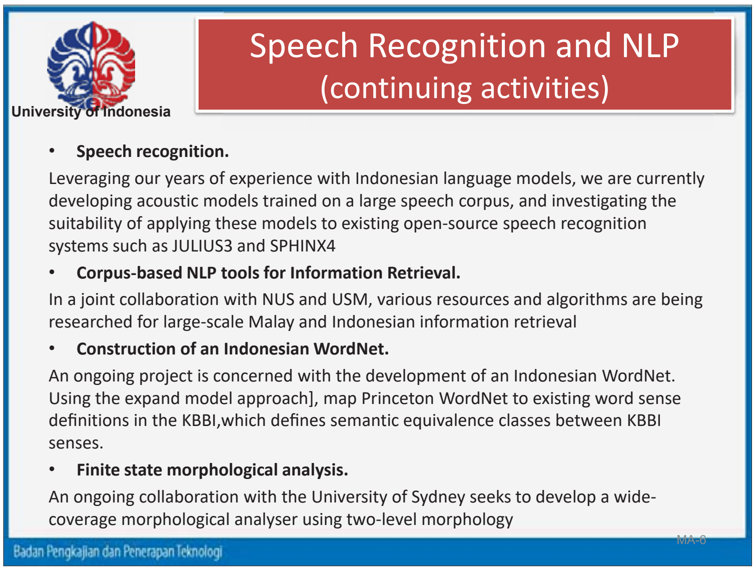
Badan Pengkajian dan Penerapan Teknologi



Indonesian Language Tools

- **Language Processing Tools**
  - – Stemmer, POS Tagger
  - – Named Entity Tagger, Phrase Chunker
  - – Statistical Constituent Parser and Dependency Parser
  - – Indonesian Reference Resolution and Semantic Analysis
  - – Indonesian MindMap Generator: http://mindmap.kataku.org
  - – Game for learning Japanese-Indonesia: http://honyaku.kataku.org
  - – Preliminary Research on Indonesian TTS based on “Unit Selection” approach
  - – Rebuild of Indonesian Diphone Database for Diphone Concatenation based Indonesian TTS
  - – Indonesian TTS based platform
  - – Improvement of Indonesian Prosody
  - – Indonesian Syllable TTS for special purpose application
  - – purchase pattern on social media: http://elysis.kataku.org
- **Speech**
  - Indonesian Automatic Speech Recognizer
  - Indonesian Speech Synthesizer
  - Indonesian TTS (BPPT)
- **Text Mining (UI, ITB)**
  - – Indonesian Question Answering System
    - – Open Domain
    - – Closed Domain with ontology
    - – Factoid, List Factoid, Non Factoid
  - – Indonesian Information

Badan Pengkajian dan Penerapan Teknologi



University of Indonesia

Speech Recognition and NLP (continuing activities)

- **Speech recognition.**

Leveraging our years of experience with Indonesian language models, we are currently developing acoustic models trained on a large speech corpus, and investigating the suitability of applying these models to existing open-source speech recognition systems such as JULIUS3 and SPHINX4

- **Corpus-based NLP tools for Information Retrieval.**

In a joint collaboration with NUS and USM, various resources and algorithms are being researched for large-scale Malay and Indonesian information retrieval

- **Construction of an Indonesian WordNet.**

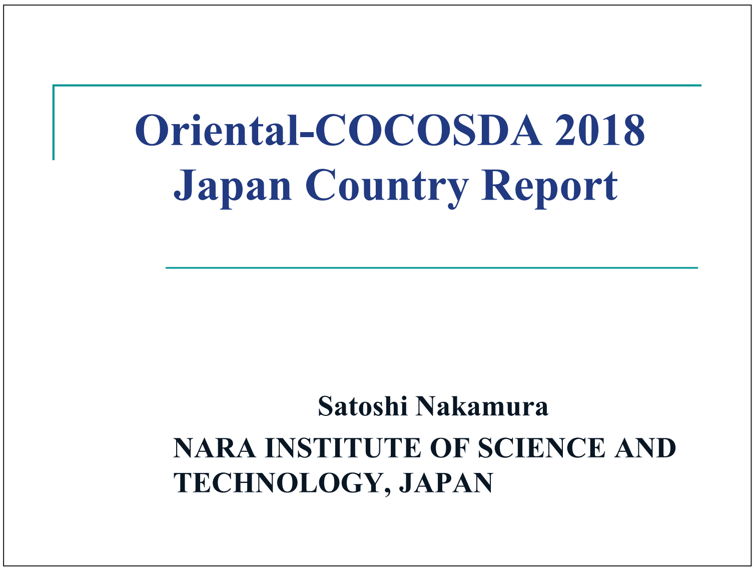
An ongoing project is concerned with the development of an Indonesian WordNet. Using the expand model approach], map Princeton WordNet to existing word sense definitions in the KBBI,which defines semantic equivalence classes between KBBI senses.

- **Finite state morphological analysis.**

An ongoing collaboration with the University of Sydney seeks to develop a wide-coverage morphological analyser using two-level morphology

MA-6

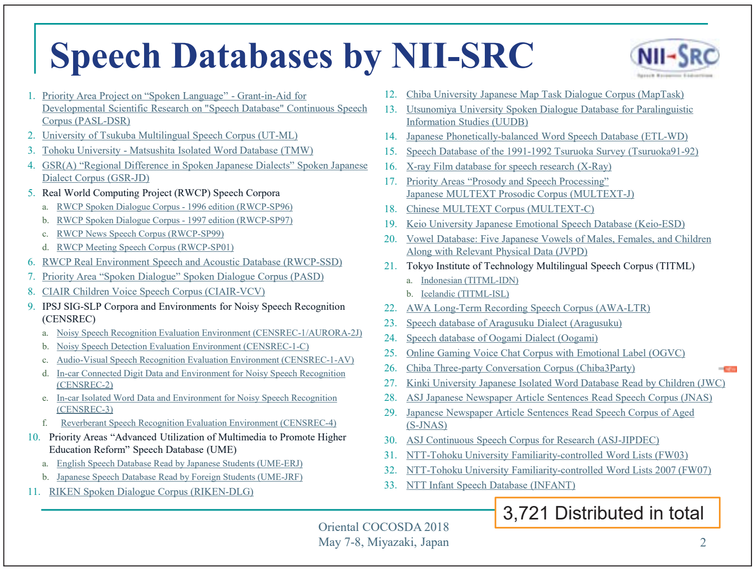
Badan Pengkajian dan Penerapan Teknologi



Oriental-COCOSDA 2018 Japan Country Report

Satoshi Nakamura

NARA INSTITUTE OF SCIENCE AND TECHNOLOGY, JAPAN

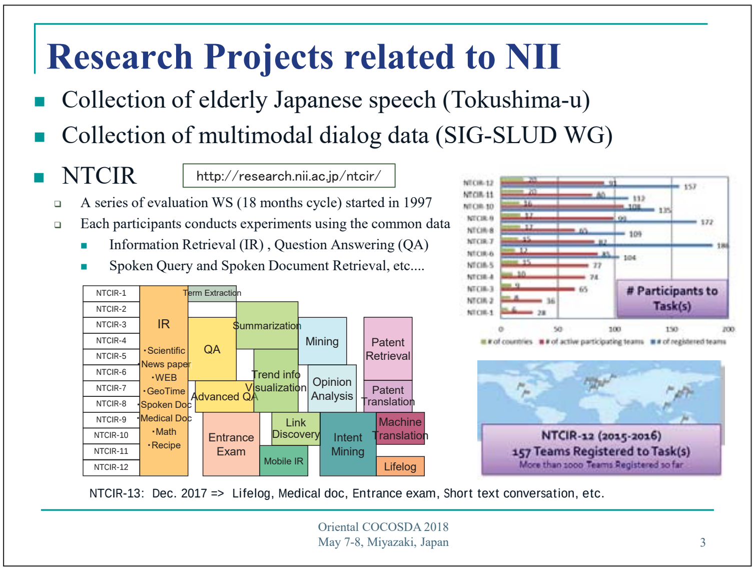


Speech Databases by NII-SRC

1. Priority Area Project on "Spoken Language" - Grant-in-Aid for Developmental Scientific Research on "Speech Database" Continuous Speech Corpus (PASL-DSR)
2. University of Tsukuba Multilingual Speech Corpus (UT-ML)
3. Tohoku University - Matsushita Isolated Word Database (TMW)
4. GSR(A) "Regional Difference in Spoken Japanese Dialects" Spoken Japanese Dialect Corpus (GSR-JD)
5. Real World Computing Project (RWCP) Speech Corpora
   a. RWCP Spoken Dialogue Corpus - 1996 edition (RWCP-SP96)
   b. RWCP Spoken Dialogue Corpus - 1997 edition (RWCP-SP97)
   c. RWCP News Speech Corpus (RWCP-SP99)
   d. RWCP Meeting Speech Corpus (RWCP-SP01)
6. RWCP Real Environment Speech and Acoustic Database (RWCP-SSD)
7. Priority Area "Spoken Dialogue" Spoken Dialogue Corpus (PASD)
8. CIAIR Children Voice Speech Corpus (CIAIR-VCV)
9. IPSJ SIG-SLP Corpora and Environments for Noisy Speech Recognition (CENSREC)
   a. Noisy Speech Recognition Evaluation Environment (CENSREC-1/AURORA-2J)
   b. Noisy Speech Detection Evaluation Environment (CENSREC-1-C)
   c. Audio-Visual Speech Recognition Evaluation Environment (CENSREC-1-AV)
   d. In-car Connected Digit Data and Environment for Noisy Speech Recognition (CENSREC-2)
   e. In-car Isolated Word Data and Environment for Noisy Speech Recognition (CENSREC-3)
   f. Reverberant Speech Recognition Evaluation Environment (CENSREC-4)
10. Priority Areas "Advanced Utilization of Multimedia to Promote Higher Education Reform" Speech Database (UME)
    a. English Speech Database Read by Japanese Students (UME-ERJ)
    b. Japanese Speech Database Read by Foreign Students (UME-JRF)
11. RIKEN Spoken Dialogue Corpus (RIKEN-DLG)
12. Chiba University Japanese Map Task Dialogue Corpus (MapTask)
13. Utsunomiya University Spoken Dialogue Database for Paralinguistic Information Studies (UUDB)
14. Japanese Phonetically-balanced Word Speech Database (ETL-WD)
15. Speech Database of the 1991-1992 Tsuruoka Survey (Tsuruoka91-92)
16. X-ray Film database for speech research (X-Ray)
17. Priority Areas "Prosody and Speech Processing" Japanese MULTEXT Prosodic Corpus (MULTEXT-J)
18. Chinese MULTEXT Corpus (MULTEXT-C)
19. Keio University Japanese Emotional Speech Database (Keio-ESD)
20. Vowel Database: Five Japanese Vowels of Males, Females, and Children Along with Relevant Physical Data (JVPD)
21. Tokyo Institute of Technology Multilingual Speech Corpus (TITML)
    a. Indonesian (TITML-IDN)
    b. Icelandic (TITML-ISL)
22. AWA Long-Term Recording Speech Corpus (AWA-LTR)
23. Speech database of Aragusuku Dialect (Aragusuku)
24. Speech database of Oogami Dialect (Oogami)
25. Online Gaming Voice Chat Corpus with Emotional Label (OGVC)
26. Chiba Three-party Conversation Corpus (Chiba3Party)
27. Kinki University Japanese Isolated Word Database Read by Children (JWC)
28. ASJ Japanese Newspaper Article Sentences Read Speech Corpus (JNAS)
29. Japanese Newspaper Article Sentences Read Speech Corpus of Aged (S-JNAS)
30. ASJ Continuous Speech Corpus for Research (ASJ-JIPDEC)
31. NTT-Tohoku University Familiarity-controlled Word Lists (FW03)
32. NTT-Tohoku University Familiarity-controlled Word Lists 2007 (FW07)
33. NTT Infant Speech Database (INFANT)

3,721 Distributed in total

Oriental COCOSDA 2018
May 7-8, Miyazaki, Japan 2

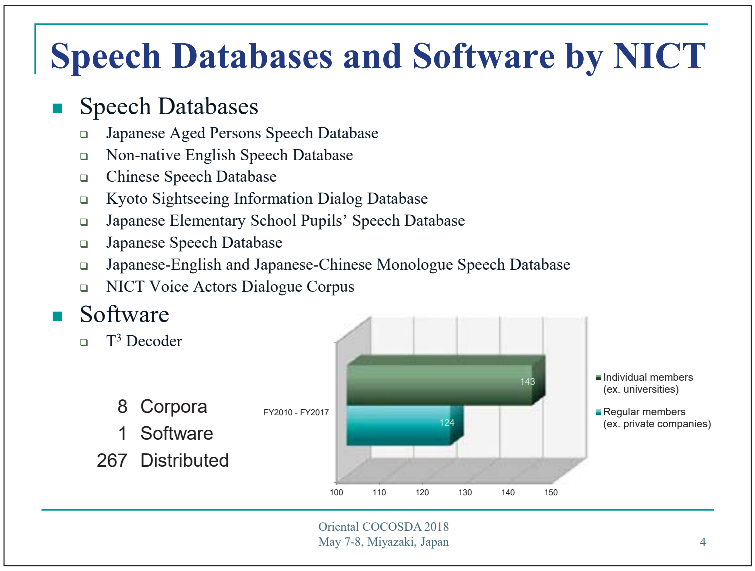


Research Projects related to NII

- Collection of elderly Japanese speech (Tokushima-u)
- Collection of multimodal dialog data (SIG-SLUD WG)
- NTCIR http://research.nii.ac.jp/ntcir/
  - A series of evaluation WS (18 months cycle) started in 1997
  - Each participants conducts experiments using the common data
    - Information Retrieval (IR) , Question Answering (QA)
    - Spoken Query and Spoken Document Retrieval, etc....

NTCIR-13: Dec. 2017 => Lifelog, Medical doc, Entrance exam, Short text conversation, etc.

Oriental COCOSDA 2018
May 7-8, Miyazaki, Japan 3



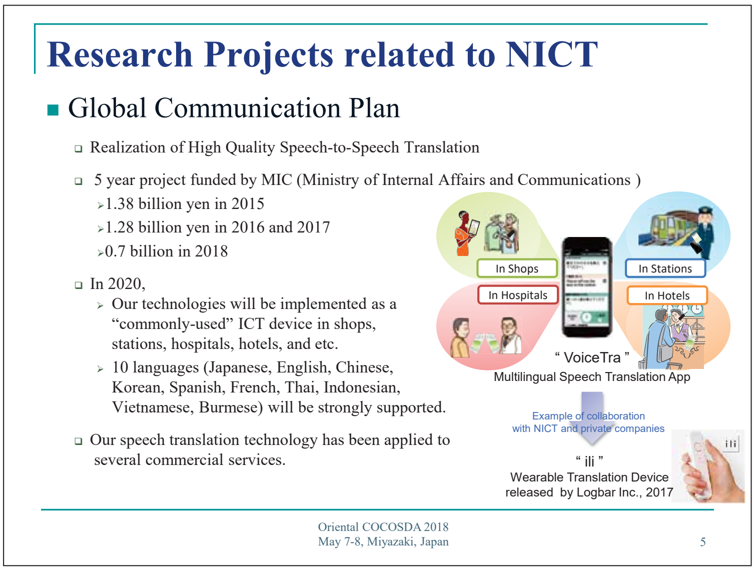
Speech Databases and Software by NICT

- Speech Databases
  - Japanese Aged Persons Speech Database
  - Non-native English Speech Database
  - Chinese Speech Database
  - Kyoto Sightseeing Information Dialog Database
  - Japanese Elementary School Pupils' Speech Database
  - Japanese Speech Database
  - Japanese-English and Japanese-Chinese Monologue Speech Database
  - NICT Voice Actors Dialogue Corpus
- Software
  - T^3 Decoder

8 Corpora
1 Software
267 Distributed

FY2010 - FY2017
143
124
100
110
120
130
140
150
Individual members (ex. universities)
Regular members (ex. private companies)

Oriental COCOSDA 2018
May 7-8, Miyazaki, Japan 4



Research Projects related to NICT

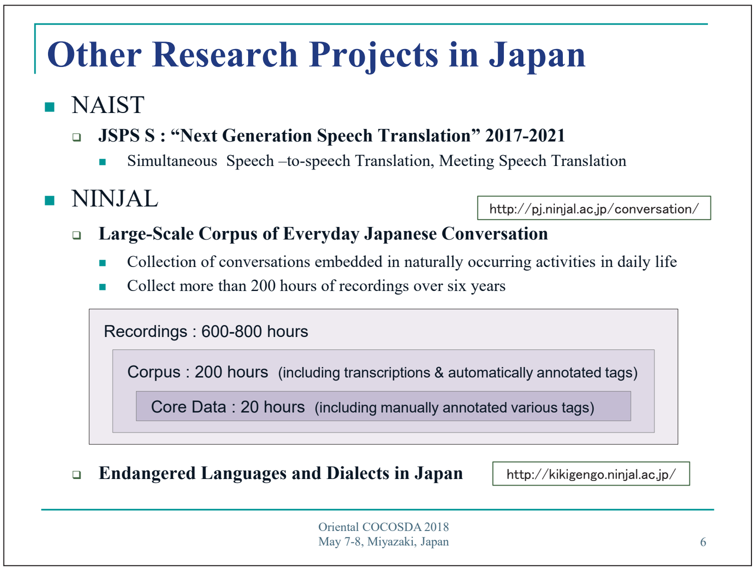
- Global Communication Plan
  - Realization of High Quality Speech-to-Speech Translation
  - 5 year project funded by MIC (Ministry of Internal Affairs and Communications)
    - 1.38 billion yen in 2015
    - 1.28 billion yen in 2016 and 2017
    - 0.7 billion in 2018
  - In 2020,
    - Our technologies will be implemented as a "commonly-used" ICT device in shops, stations, hospitals, hotels, and etc.
    - 10 languages (Japanese, English, Chinese, Korean, Spanish, French, Thai, Indonesian, Vietnamese, Burmese) will be strongly supported.
  - Our speech translation technology has been applied to several commercial services.

" VoiceTra "
Multilingual Speech Translation App

Example of collaboration with NICT and private companies

" ili "
Wearable Translation Device released by Logbar Inc., 2017

Oriental COCOSDA 2018
May 7-8, Miyazaki, Japan 5



Other Research Projects in Japan

- NAIST
  - **JSPS S : "Next Generation Speech Translation" 2017-2021**
    - Simultaneous Speech –to-speech Translation, Meeting Speech Translation
- NINJAL http://pj.ninjal.ac.jp/conversation/
  - **Large-Scale Corpus of Everyday Japanese Conversation**
    - Collection of conversations embedded in naturally occurring activities in daily life
    - Collect more than 200 hours of recordings over six years

Recordings : 600-800 hours

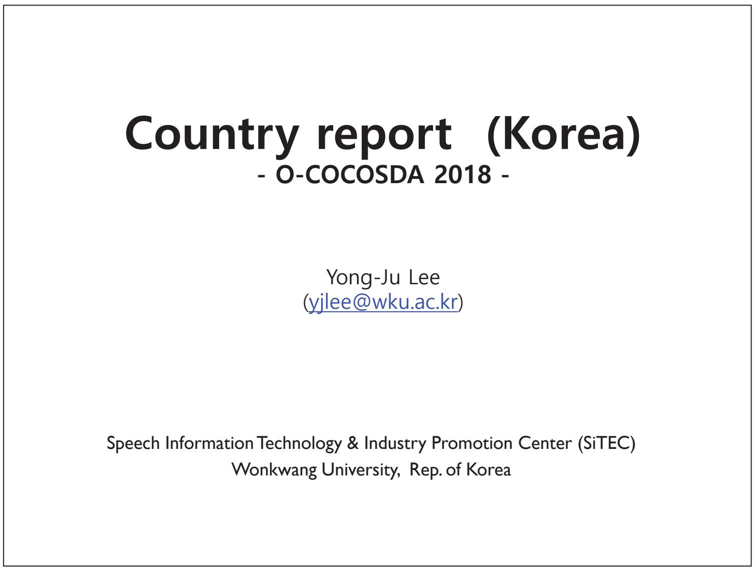
Corpus : 200 hours (including transcriptions & automatically annotated tags)

Core Data : 20 hours (including manually annotated various tags)

  - **Endangered Languages and Dialects in Japan** http://kikigengo.ninjal.ac.jp/

Oriental COCOSDA 2018
May 7-8, Miyazaki, Japan 6

Country Report - Korea

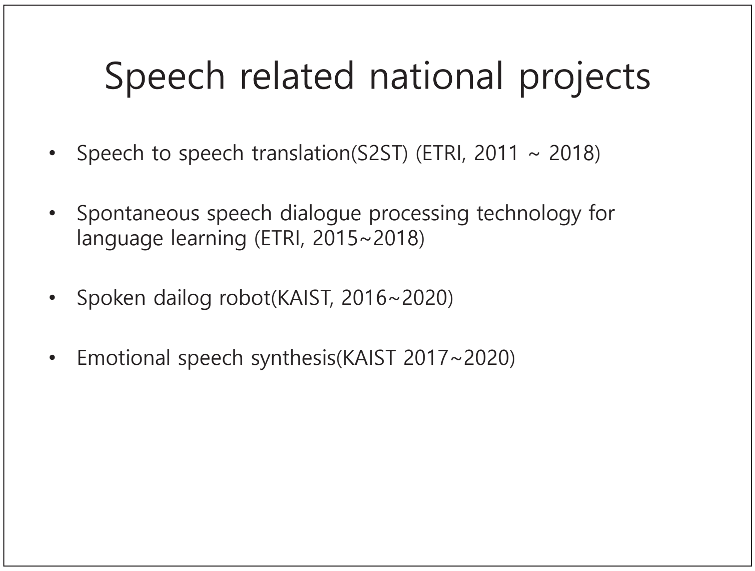


Country report (Korea)

- O-COCOSDA 2018 -

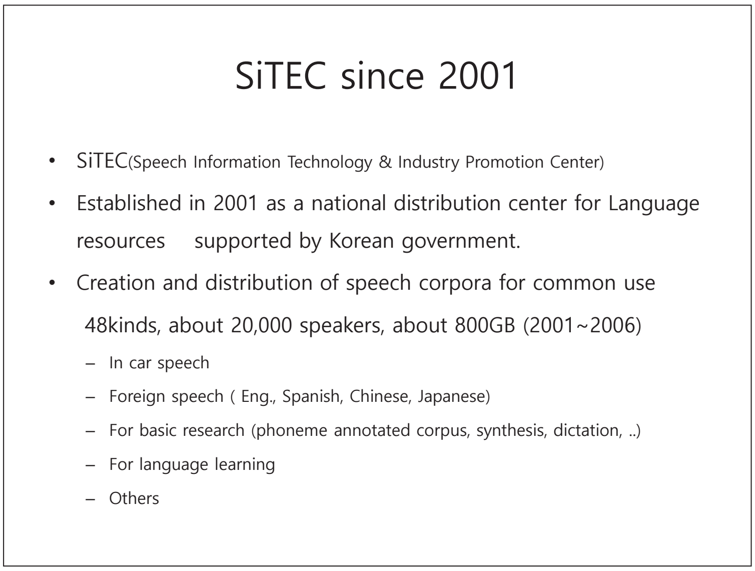
Yong-Ju Lee
(yjlee@wku.ac.kr)

Speech Information Technology & Industry Promotion Center (SiTEC)
Wonkwang University, Rep. of Korea



Speech related national projects

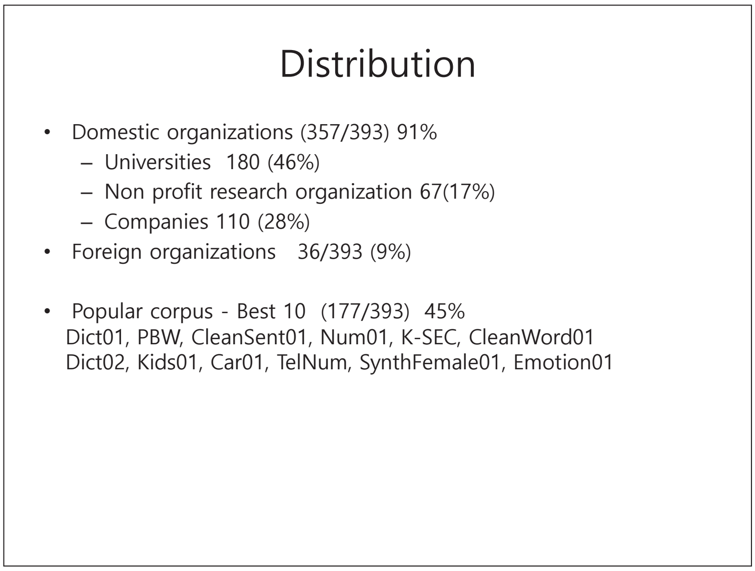
- Speech to speech translation(S2ST) (ETRI, 2011 ~ 2018)
- Spontaneous speech dialogue processing technology for language learning (ETRI, 2015~2018)
- Spoken dailog robot(KAIST, 2016~2020)
- Emotional speech synthesis(KAIST 2017~2020)



SiTEC since 2001

- SiTEC(Speech Information Technology & Industry Promotion Center)
- Established in 2001 as a national distribution center for Language resources supported by Korean government.
- Creation and distribution of speech corpora for common use

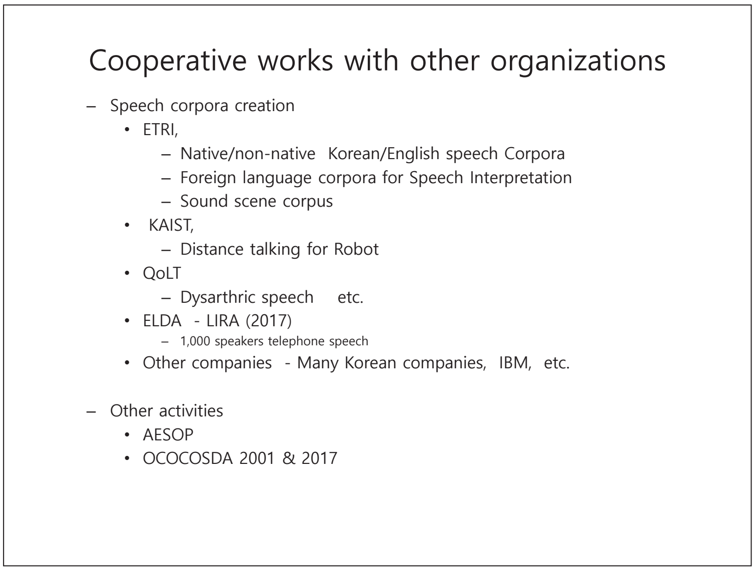
  48kinds, about 20,000 speakers, about 800GB (2001~2006)
  - In car speech
  - Foreign speech (Eng., Spanish, Chinese, Japanese)
  - For basic research (phoneme annotated corpus, synthesis, dictation, ..)
  - For language learning
  - Others



Distribution

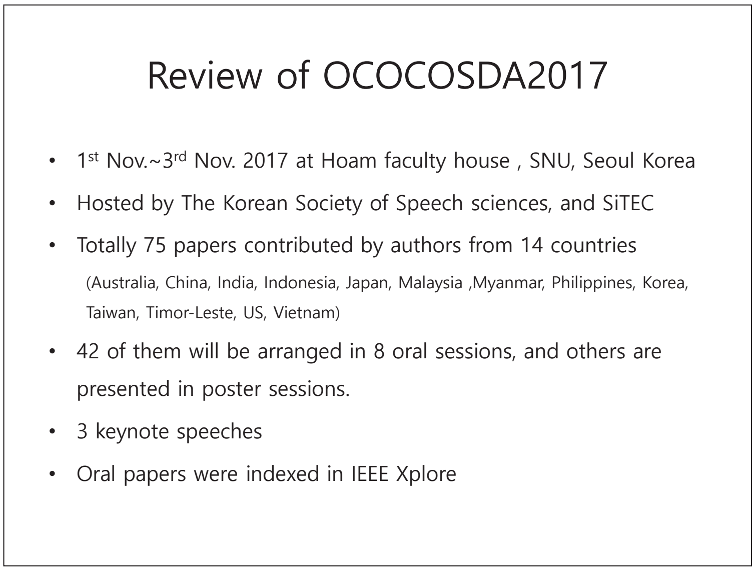
- Domestic organizations (357/393) 91%
  - Universities 180 (46%)
  - Non profit research organization 67(17%)
  - Companies 110 (28%)
- Foreign organizations 36/393 (9%)
- Popular corpus - Best 10 (177/393) 45%

  Dict01, PBW, CleanSent01, Num01, K-SEC, CleanWord01
  Dict02, Kids01, Car01, TelNum, SynthFemale01, Emotion01



Cooperative works with other organizations

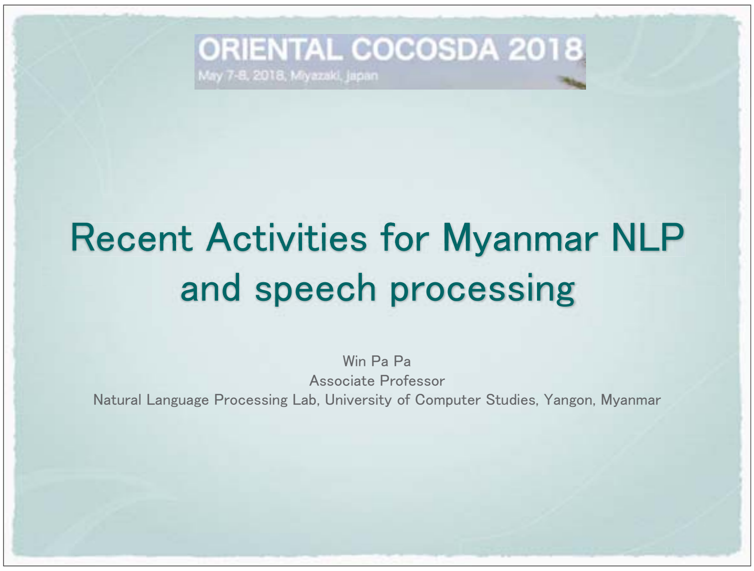
- Speech corpora creation
  - ETRI,
    - Native/non-native Korean/English speech Corpora
    - Foreign language corpora for Speech Interpretation
    - Sound scene corpus
  - KAIST,
    - Distance talking for Robot
  - QoLT
    - Dysarthric speech etc.
  - ELDA - LIRA (2017)
    - 1,000 speakers telephone speech
  - Other companies - Many Korean companies, IBM, etc.
- Other activities
  - AESOP
  - OCOCOSDA 2001 & 2017



Review of OCOCOSDA2017

- 1st Nov.~3rd Nov. 2017 at Hoam faculty house , SNU, Seoul Korea
- Hosted by The Korean Society of Speech sciences, and SiTEC
- Totally 75 papers contributed by authors from 14 countries

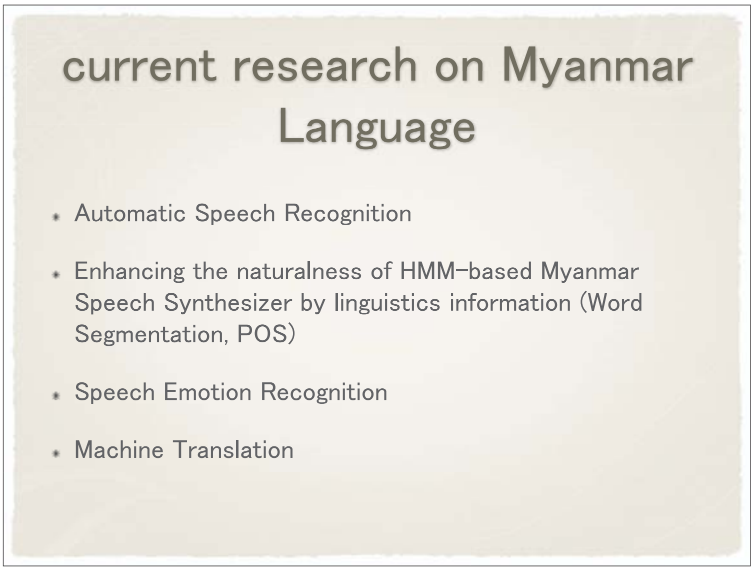
  (Australia, China, India, Indonesia, Japan, Malaysia ,Myanmar, Philippines, Korea, Taiwan, Timor-Leste, US, Vietnam)
- 42 of them will be arranged in 8 oral sessions, and others are presented in poster sessions.
- 3 keynote speeches
- Oral papers were indexed in IEEE Xplore



ORIENTAL COCOSDA 2018
May 7-8, 2018, Miyazaki, Japan

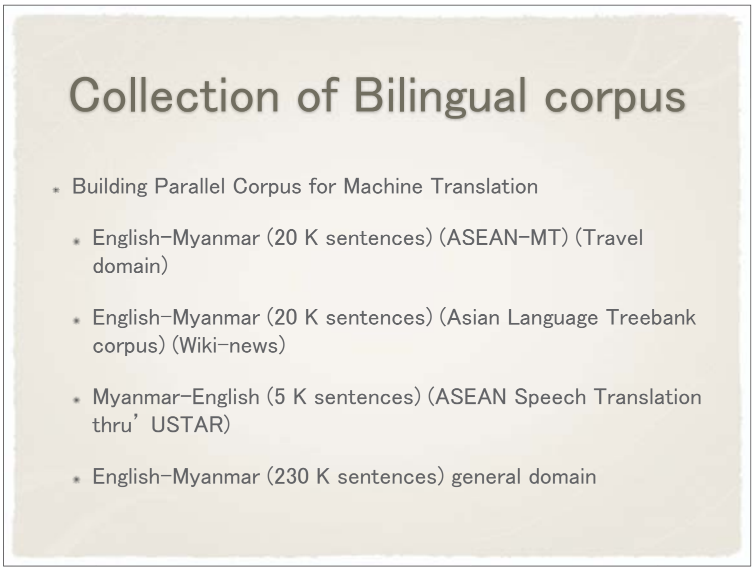
Recent Activities for Myanmar NLP and speech processing

Win Pa Pa
Associate Professor
Natural Language Processing Lab, University of Computer Studies, Yangon, Myanmar



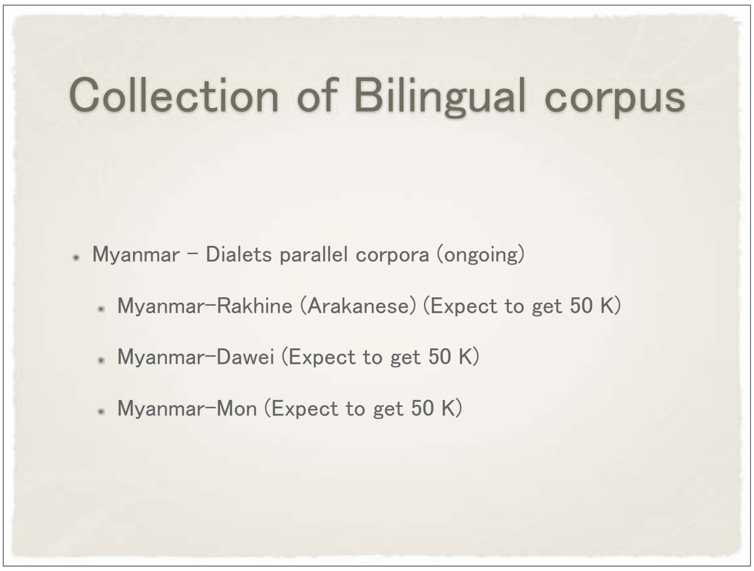
current research on Myanmar Language

- Automatic Speech Recognition
- Enhancing the naturalness of HMM-based Myanmar Speech Synthesizer by linguistics information (Word Segmentation, POS)
- Speech Emotion Recognition
- Machine Translation



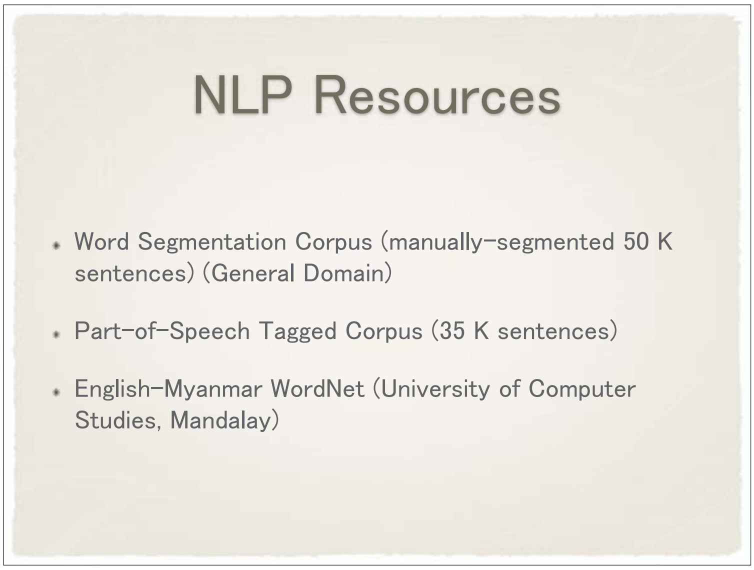
Collection of Bilingual corpus

- Building Parallel Corpus for Machine Translation
  - English-Myanmar (20 K sentences) (ASEAN-MT) (Travel domain)
  - English-Myanmar (20 K sentences) (Asian Language Treebank corpus) (Wiki-news)
  - Myanmar-English (5 K sentences) (ASEAN Speech Translation thru' USTAR)
  - English-Myanmar (230 K sentences) general domain



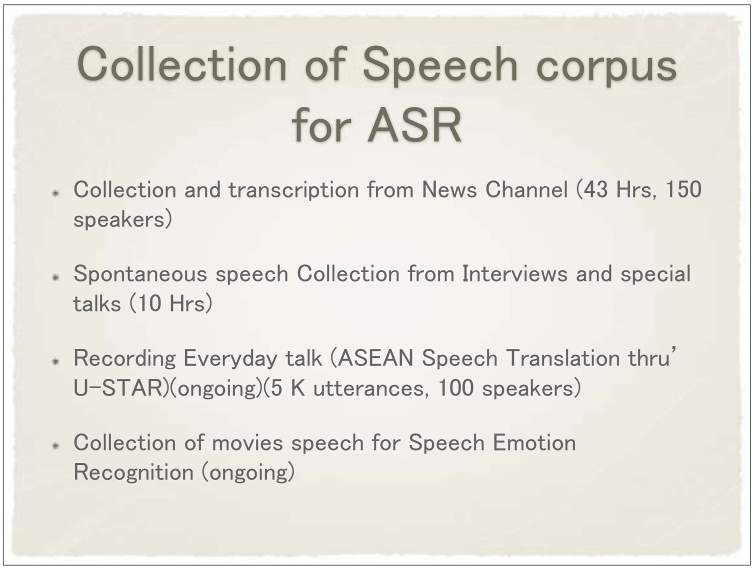
Collection of Bilingual corpus

- Myanmar – Dialets parallel corpora (ongoing)
  - Myanmar-Rakhine (Arakanese) (Expect to get 50 K)
  - Myanmar-Dawei (Expect to get 50 K)
  - Myanmar-Mon (Expect to get 50 K)



NLP Resources

- Word Segmentation Corpus (manually-segmented 50 K sentences) (General Domain)
- Part-of-Speech Tagged Corpus (35 K sentences)
- English-Myanmar WordNet (University of Computer Studies, Mandalay)



Collection of Speech corpus for ASR

- Collection and transcription from News Channel (43 Hrs, 150 speakers)
- Spontaneous speech Collection from Interviews and special talks (10 Hrs)
- Recording Everyday talk (ASEAN Speech Translation thru' U-STAR)(ongoing)(5 K utterances, 100 speakers)
- Collection of movies speech for Speech Emotion Recognition (ongoing)

O-COCOSDA Country Report of Pakistan

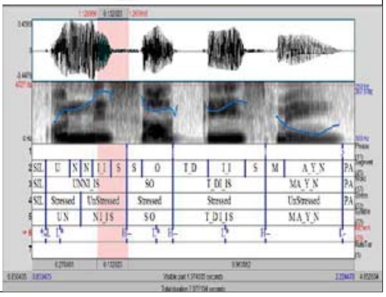
Progress reported till May, 2018

Tania Habib
Sarmad Hussain

Center for Language Engineering
Al-Khawarizmi Institute of Computer Science
University of Engineering and Technology Lahore, Pakistan

Intonation Marked Urdu Speech Corpus (1/1)

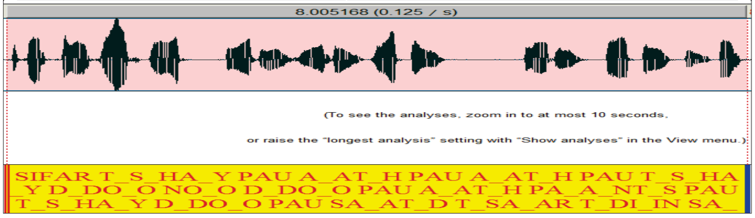
- Collaborative project: UET, DHA Suffa university and Konstanz university
- Recording Style: Read speech Corpus
- Corpus size: 3 hrs of speech recorded by 3 speakers (2 males and 1 female) in an anechoic chamber
- Sampling Rate: 48kHz
- Tiers annotated: Stress, break index and intonation [1]



[1] Mumtaz, B., Urooj, S., Hussain, S. and Ehsan Ul Haq. "Break Index (BI) Annotated Speech Corpus for Urdu TTS", in the Proceedings of 19th Oriental COCOSDA Conference 2016, Bali, Indonesia. (URL: http://www.ococosda2016.org/)

URDU SPEECH CORPORA FOR BANKING SECTOR IN PAKISTAN (1/2)

City	Lahore
Accents	Urdu, Punjabi and others
Age	18-50
Environment	Indoor & Outdoor
Channel	Mobile Phones (Ufone, Warid-Jazz, Telenor, Zong) & landline (PTCL)
Recording Style	Read Speech
File Format	.wav
Sampling Rate	8kHz



(To see the analyses, zoom in to at most 10 seconds,
or raise the "longest analysis" setting with "Show analyses" in the View menu.)

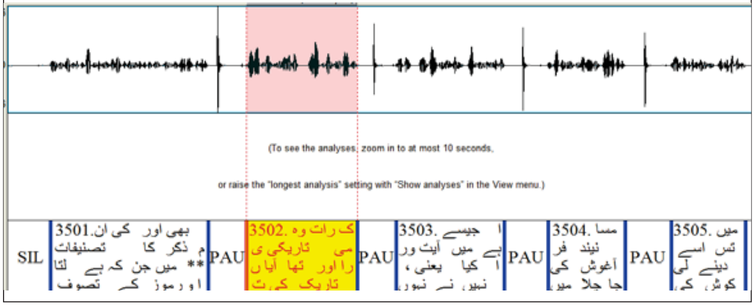
SIFAR T_S_HA_Y PAU A_AT_H PAU A_AT_H PAU T_S_HA _Y D_DO_O NO_O D_DO_O PAU A_AT_H PA_A_NT_S PAU T_S_HA_Y D_DO_O PAU SA_AT_D T_SA_AR T_DI_IN SA_

URDU SPEECH CORPORA FOR BANKING SECTOR IN PAKISTAN (2/2)

Corpora	No. of Speakers	Vocabulary	Duration (minutes)
Debit card number (DCN) corpus	230	Urdu digits 0-9	258
Date of birth (DOB) corpus	230	120	98
Debit card expiry date (DCED) corpus	235	101 Urdu digits	55
Debit card last four digit (DCLFD) corpus	233	107 Urdu digits	47
Month's names (Urdu/English) corpus (MsN)	200	English & Urdu months names	44
Special words corpus	200	1st to 31st & jəkəm	23
Yes/no account type (YN/AT) corpus	200	Urdu Words	21
Counting (Urdu/English) corpus	200	English & Urdu counting	15 Total = 9 hours

Urdu ASR Speech Corpus

- Corpus Size:10500 phonetically rich sentences (9 hours of speech data)
- Recording Channel: laptop and Mic.
- Speakers: Data recorded from 30 speakers (male/female)
- Annotation: Speech files automatically segmented through utility



(To see the analyses, zoom in to at most 10 seconds,
or raise the "longest analysis" setting with "Show analyses" in the View menu.)

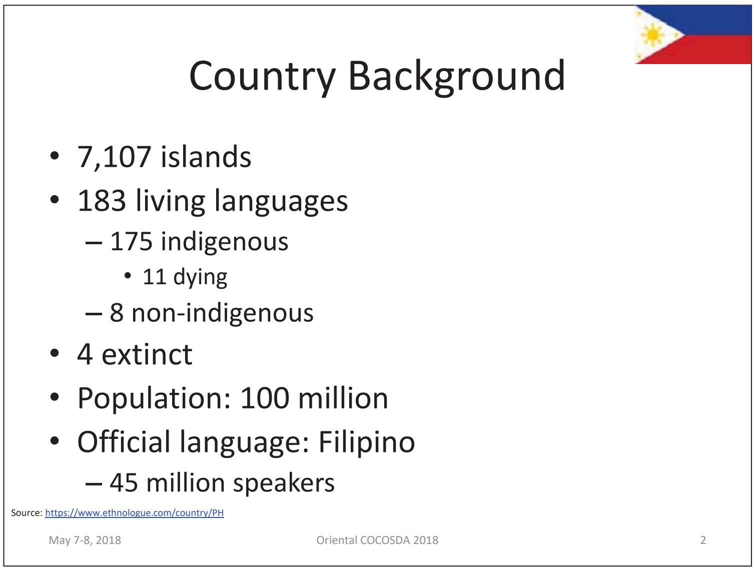
Urdu Large Vocabulary Continuous Speech Recognition System

- Open domain
- Training Corpus Statistics
  - Duration: 28.36 hours
  - Lexicon: 106K
  - LM 3-gram : 35 million text corpus
- Testing Corpus
  - Male : 17 speakers (29 minutes)
  - Female: 17 Speakers (41 minutes)
- Developed using KALDI toolkit
- WER: 21 %

2018 Philippine Country Report

Nathaniel Oco
Leif Romeritch Syliongka
Rachel Edita Roxas

National University

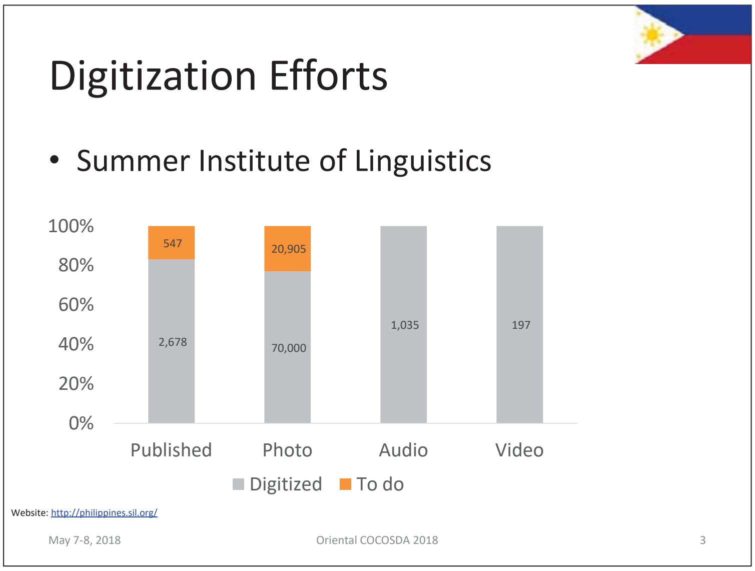


Country Background

- 7,107 islands
- 183 living languages
  - 175 indigenous
    - 11 dying
  - 8 non-indigenous
- 4 extinct
- Population: 100 million
- Official language: Filipino
  - 45 million speakers

Source: https://www.ethnologue.com/country/PH

May 7-8, 2018 Oriental COCOSDA 2018 2

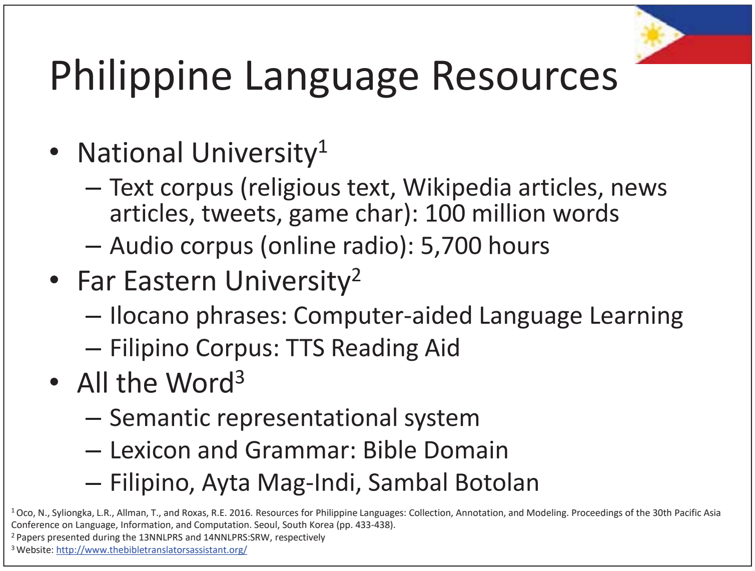


Digitization Efforts

- Summer Institute of Linguistics

Website: http://philippines.sil.org/

May 7-8, 2018 Oriental COCOSDA 2018 3



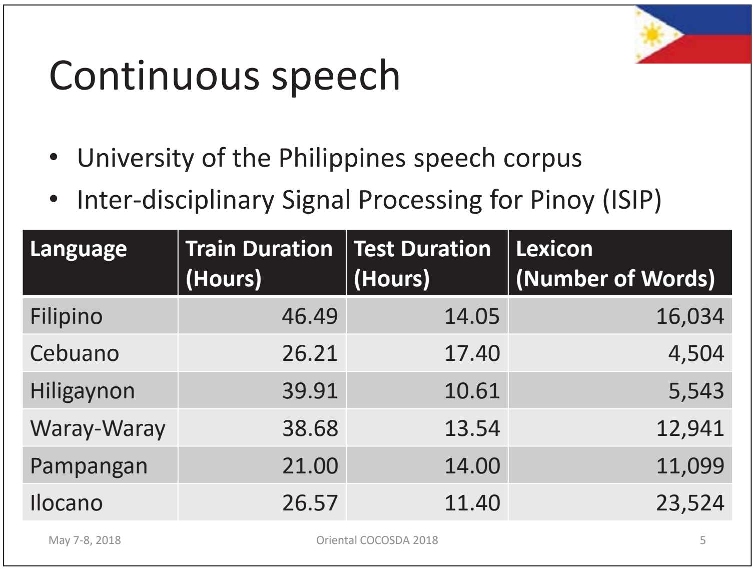
Philippine Language Resources

- National University[1]
  - Text corpus (religious text, Wikipedia articles, news articles, tweets, game char): 100 million words
  - Audio corpus (online radio): 5,700 hours
- Far Eastern University[2]
  - Ilocano phrases: Computer-aided Language Learning
  - Filipino Corpus: TTS Reading Aid
- All the Word[3]
  - Semantic representational system
  - Lexicon and Grammar: Bible Domain
  - Filipino, Ayta Mag-Indi, Sambal Botolan

[1] Oco, N., Syliongka, L.R., Allman, T., and Roxas, R.E. 2016. Resources for Philippine Languages: Collection, Annotation, and Modeling. Proceedings of the 30th Pacific Asia Conference on Language, Information, and Computation. Seoul, South Korea (pp. 433-438).

[2] Papers presented during the 13NNLPRS and 14NNLPRS:SRW, respectively

[3] Website: http://www.thebibletranslatorsassistant.org/



Continuous speech

- University of the Philippines speech corpus
- Inter-disciplinary Signal Processing for Pinoy (ISIP)

Language	Train Duration (Hours)	Test Duration (Hours)	Lexicon (Number of Words)
Filipino	46.49	14.05	16,034
Cebuano	26.21	17.40	4,504
Hiligaynon	39.91	10.61	5,543
Waray-Waray	38.68	13.54	12,941
Pampangan	21.00	14.00	11,099
Ilocano	26.57	11.40	23,524

May 7-8, 2018 Oriental COCOSDA 2018 5

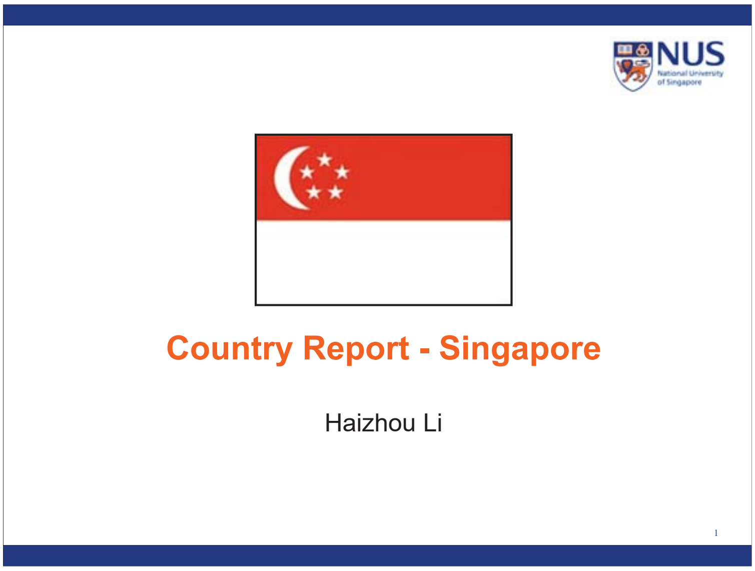
Isolated speech: digits 0-9

- E-Participation 2.0: Connecting Diverse Philippine Populations for Disaster Risk Management with a Toolkit Integrating Text and Speech Analytics
  - National University and University of California, Berkeley
- Duration: **8 hrs 41 min 19 sec**
- Speakers: **517 individuals**
- Languages: Filipino, English, Cebuano, Ilocano, Spanish-borrowed cardinal numbers
- Noise corpus

Project website: http://eparticipation.national-u.edu.ph/

May 7-8, 2018 Oriental COCOSDA 2018 6

Country Report - Singapore

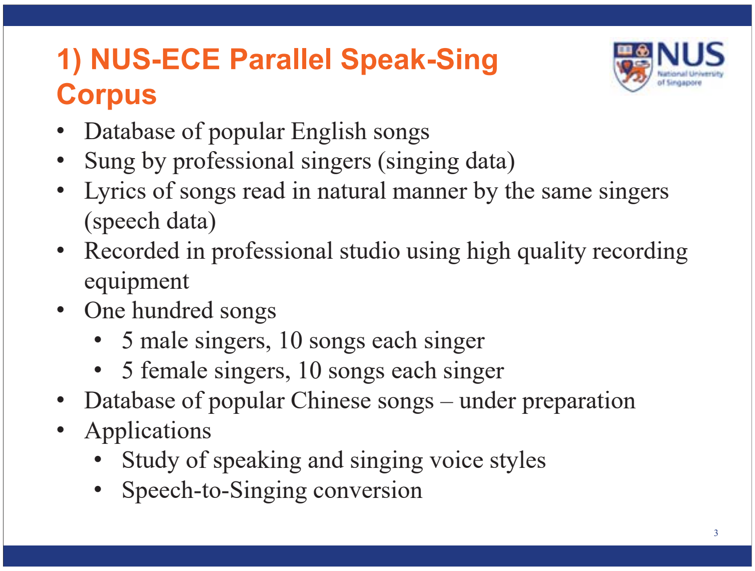


Country Report - Singapore

Haizhou Li

Contributing Organizations

1) Human Language Technology Lab, National University of Singapore
2) The Speech and Language Technology Lab, Nanyang Technological University
3) Institute for Infocomm Research, A*STAR



1) NUS-ECE Parallel Speak-Sing Corpus

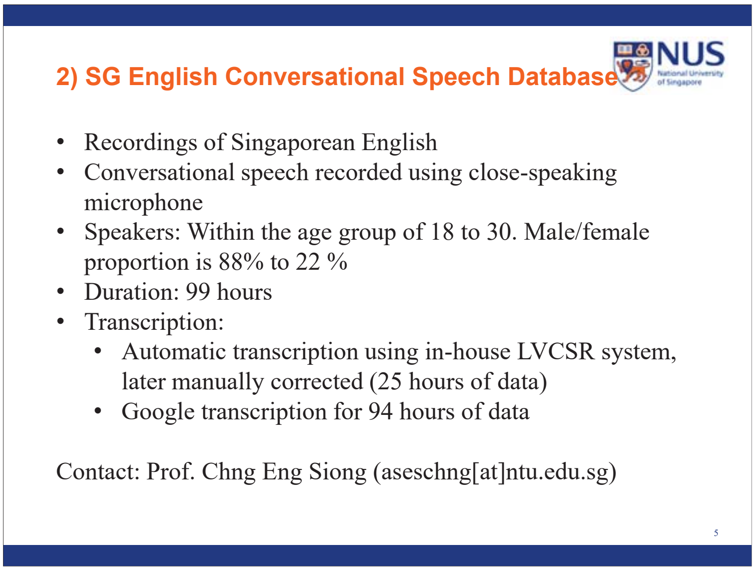
- Database of popular English songs
- Sung by professional singers (singing data)
- Lyrics of songs read in natural manner by the same singers (speech data)
- Recorded in professional studio using high quality recording equipment
- One hundred songs
  - 5 male singers, 10 songs each singer
  - 5 female singers, 10 songs each singer
- Database of popular Chinese songs – under preparation
- Applications
  - Study of speaking and singing voice styles
  - Speech-to-Singing conversion

1) NUS-ECE Parallel Speak-Sing Corpus

Related Publications

- K. Vijayan, M. Dong, and H. Li, “A dual alignment scheme for improved speech-to-singing voice conversion,” in Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA-ASC), December 2017
- K. Vijayan and H. Li, “Parallel speak-sing corpus of English and Chinese songs for speech-to-singing voice conversion,” in Workshop on Asian Language Resources, Language Resources and Evaluation Conference(LREC), May 2018.

Contact: Karthika Vijayan (vijayan.Karthika[at]nus.edu.sg)



2) SG English Conversational Speech Database

- Recordings of Singaporean English
- Conversational speech recorded using close-speaking microphone
- Speakers: Within the age group of 18 to 30. Male/female proportion is 88% to 22 %
- Duration: 99 hours
- Transcription:
  - Automatic transcription using in-house LVCSR system, later manually corrected (25 hours of data)
  - Google transcription for 94 hours of data

Contact: Prof. Chng Eng Siong (aseschng[at]ntu.edu.sg)

3) SG English radio and Youtube bloggers Speech Database

- Source: Singapore Radio and Youtube
- Conversational English and Singaporean English
- Duration: About 880 hours
- Transcription: Google/ Youtube

Contact: Prof. Chng Eng Siong (aseschng[at]ntu.edu.sg)

Country Report - Taiwan

Oriental COCOSDA – Country Report 2018
Language Resources Developed in Taiwan

Sin-Horng CHEN, National Chiao Tung University
Chiu-yu TSENG, Academia Sinica, Taiwan

NCTU Preschool Children Speech Corpus Corpora – Prof. Yih-Ru Wang

- **Purpose:** to develop a computer-aided language learning system for preschool children with developmental articulation disorders
- Approved by the Research Ethics Committee of the NTU Hospital, HsinChu Branch (Approval Number: 105-024- F)
- **Corpus Contain**
  - **Text source:** text is from Jing-Yi Jeng "Manual of Mandarin Speech Test for Children"; 40 phonetic balanced words of 1 ~ 2 syllables
  - **Speakers:** 80 speakers, age of *3 ~ 5*
  - **Size:** each speaker recorded 32 words and 8 short sentences
  - **Format:** Microphone read speech/16kHz sample rate

1

Release & Distribution Update
Speech Corpora Academia Sinica—Chiu-yu Tseng

Corpora	Sinica Cospro (Sinica Continuous Speech Prosody Corpora & Toolkit)		AESOP (Asian English Speech cOrpus Project)-ILAS Corpora
Type	read speech mic		read speech mic
Size	**10.5 GB** (7.7 GB preprocessed) 114 spkrs total		**14 GB total** **AESOP-ILAS 1:** 500 spkrs, 8.64 GB **AESOP-ILAS 2:** 40 spkrs, 5.42 GB
Language	**L1 Mandarin**		**L1 English & Mandarin L2 English**
Content	**9 sets of speech corpora:** Prosodic features of continuous speech **Toolkit:** self-developed software		**AESOP-ILAS 1:** English segmental and supra-segmental features (words, sentences, "The North Wind & the Sun") **AESOP-ILAS 2:** English phonotactics, focus/prominence and discourse features (words, sentences, "Cinderalla")
Annotation	1. HTK Force Aligned Segmental labeling 2. Spot Checked & Adjusted HTK files (Manual) 3. Perceived Boundaries & Prominence (Manual)		
Access	ACLCLP (Association of Computational Linguistics & Chinese Language Processing)		
	http://www.aclclp.org.tw/use_mat.php#cospro		http://www.aclclp.org.tw/use_mat.php#aesop
Fees	Academic Use Only		
	Domestic	NT1,000	NT1,000
	International	USD100	USD100

Academia Sinica – Audio-Visual Corpus

- **PI:** Prof. Yu Tsao, Academia Sinica
- **Purpose**: Audio-visual bi-modal speech enhancement
- **Size:** 2 speakers (1 male, 1 female)
  - Each speaker has 320 utterances
- **Recording Setting:**
  - Audio: stereo, 48kHz
  - Video: 1920x1080 at 29.97 frames per second (fps)
  - Camcorder (JVC Everio GZ-HD520BU) in a seminar room.
- **Text sources:**
  - Sentences selected from Taiwan Mandarin hearing in noise test (TMHINT)
  - Each sentence contains 10 characters
- **Applications:**
  - Has been used to develop a CNN-based audio-visual speech enhancement system

3

NCKU-MHMC Interview Corpus

- **PI:** Prof. Chung-Hsien Wu, NCKU
- **Purpose**: to develop an interview coaching system
- **Statistics of the Corpus**

	Total
Number of subjects	4
Number of dialogs	260
Number of turns	3016
Number of normal / follow-up dialog turns	1754 / 1262
Average number of turns	10.7
Average number of normal / follow up turns	5.74 / 4.96
Average number of sentences in each answer	3.84
Interview time (minute)	20

Cevkqp
Qr gpkpi
Gzr gtkgpeg *Cevkqp$_3$+
J cdkv *Cevkqp$_4$+
Oqvkxcvkqp"*Cevkqp$_5$+
Unkm*Cevkqp$_6$+
Ej ctcevgtkuvkeu *Cevkqp$_7$+
Qvj gtu *Cevkqp$_8$+
Rgtur gevkxg"*Cevkqp$_9$+
Hqmy /wr *Cevkqp$_:$ +
Gpf "*Cevkqp$_;$ +

4

Three NTHU Corpora – Prof. LEE, CHI-CHUN

- **NNIME: The NTHU-NTUA Chinese Interactive Multimodal Emotion Corpus**
  - **Recording:** Dual-channel Audio, HD Video
  - **Size:** 6701 utterances, 11 hours, 102 dyadic sessions (3 min/session)
  - **Participants:** 44 (Man: 20; Female: 24)
- **THE NTHU-NTUH ADOS Audio-Video Database** (IRB# :10501HE002)
  - Real Autism Diagnosis Observation Schedule (ADOS) Diagnostic Interview
  - Synchronized dual-channels lapel microphones, Three HD camcorders
  - **Size & Participants:** ~80 hrs, 60 Clinical diagnoses (Classical Autism (28), Asperger (20), High-functioning Autism (12))
  - **Labels:** ADOS behavior codes (Module 3, and Module 4)
- **THE NTHU-CGMH Pain Audio-Video Database** (IRB#: 201700744B0)
  - **Scenario:** Real emergency triage session (prior and after treatment) at Chang Gung Memorial Hospital
  - **Size:** ~2600 sentences;
  - **Participants:** Non-major illness (246), Patients with pain (275)
  - **Labels:** Numerical Rating Scale Pain (NRS) : 0 – 10
  - **Recording:** Sony HDR handy cam (audio-visual data)

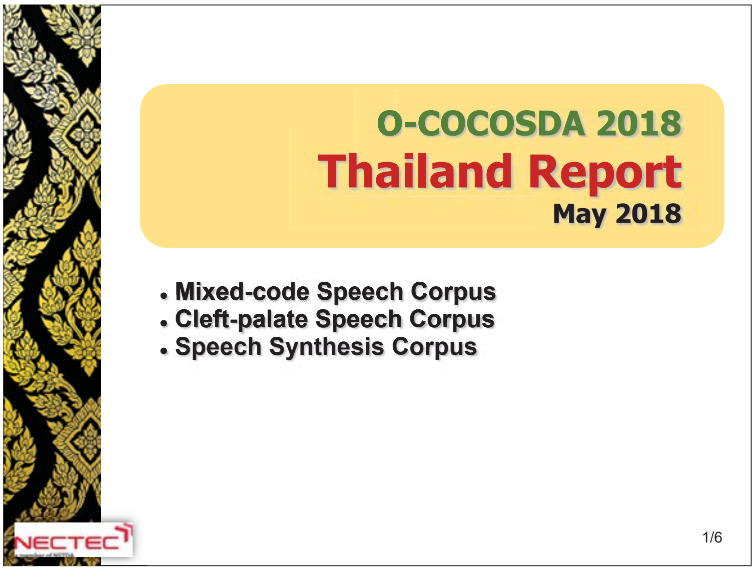
5

Taiwanese Speech in the Wild (TSW Corpora)

- **PI:** Prof. Yuan-Fu Liao, NTUT
- to support State-of-the-Art **Deep Learning-based ASR**
- Support MOST "Formosa Grand Challenge: Dialogue with AI"

	Sources	Hours	Quality	Remark
Public Television Service (PTS)	公共電視	2333	Station Copy	with some Subtitles, Code-Switching
National Education Radio (NER)	國立教育廣播電臺 National Education Radio	1280	Station Copy	no Subtitles, Dry Speech after 2018
National Chiao-Tung Univ. (NCTU)		200 courses	Recording	no Subtitles, many Proper Nouns
National ChengChi Univ. (NCCU)		400	Recording	no Subtitles, , many Proper Nouns
Junyi Academy	均一教育平台 Junyi Academy	500	Station Copy	no Subtitles, , many Proper Nouns

Country Report - Thailand



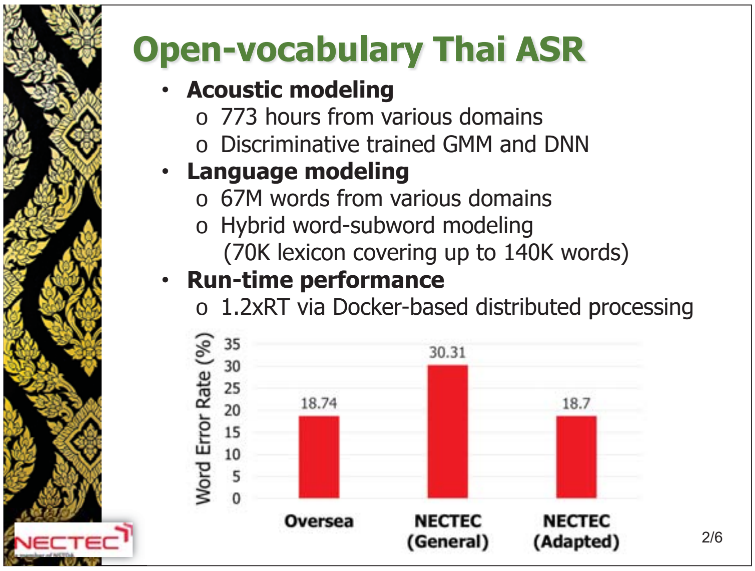
O-COCOSDA 2018

Thailand Report

May 2018

- Mixed-code Speech Corpus
- Cleft-palate Speech Corpus
- Speech Synthesis Corpus

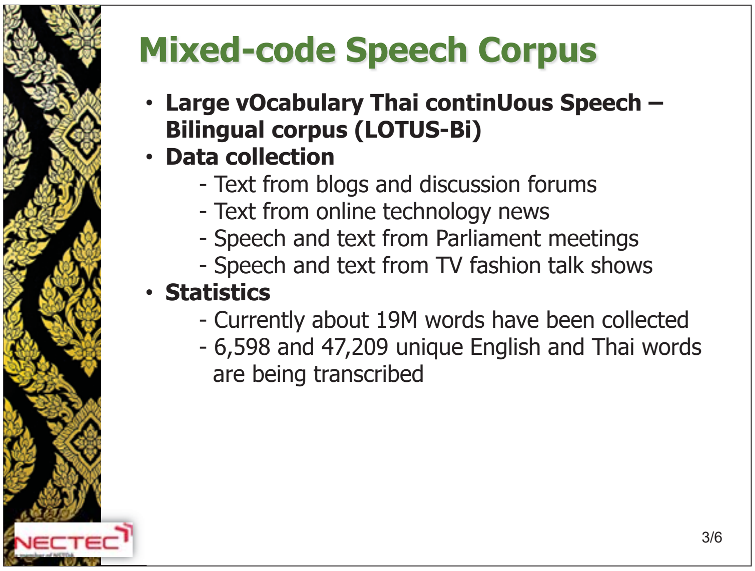
1/6



Open-vocabulary Thai ASR

- **Acoustic modeling**
  - o 773 hours from various domains
  - o Discriminative trained GMM and DNN
- **Language modeling**
  - o 67M words from various domains
  - o Hybrid word-subword modeling (70K lexicon covering up to 140K words)
- **Run-time performance**
  - o 1.2xRT via Docker-based distributed processing

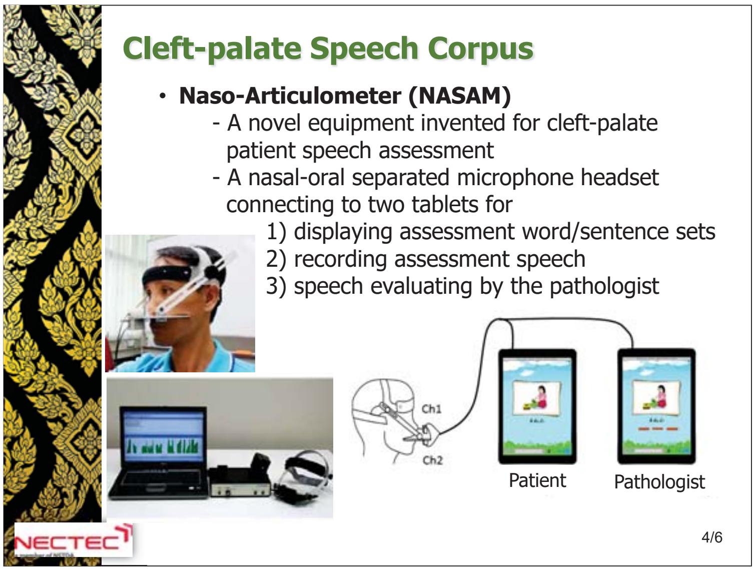
2/6



Mixed-code Speech Corpus

- **Large vOcabulary Thai continUous Speech – Bilingual corpus (LOTUS-Bi)**
- **Data collection**
  - Text from blogs and discussion forums
  - Text from online technology news
  - Speech and text from Parliament meetings
  - Speech and text from TV fashion talk shows
- **Statistics**
  - Currently about 19M words have been collected
  - 6,598 and 47,209 unique English and Thai words are being transcribed

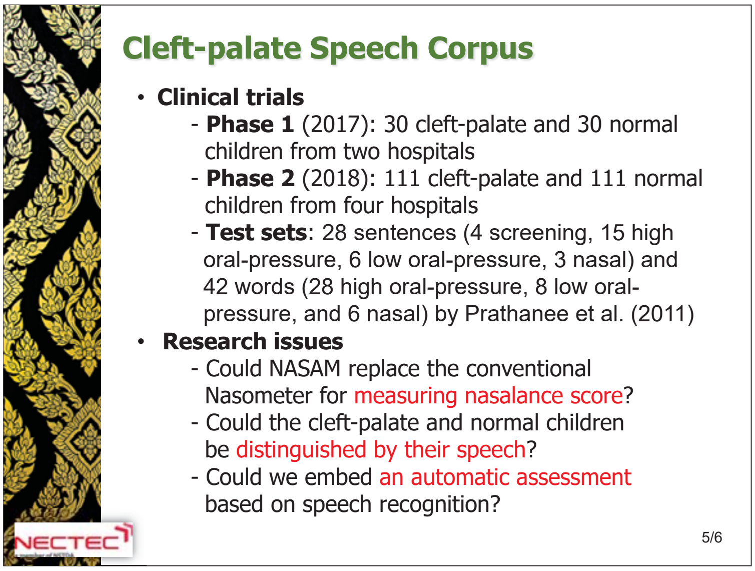
3/6



Cleft-palate Speech Corpus

- **Naso-Articulometer (NASAM)**
  - A novel equipment invented for cleft-palate patient speech assessment
  - A nasal-oral separated microphone headset connecting to two tablets for
    1) displaying assessment word/sentence sets
    2) recording assessment speech
    3) speech evaluating by the pathologist

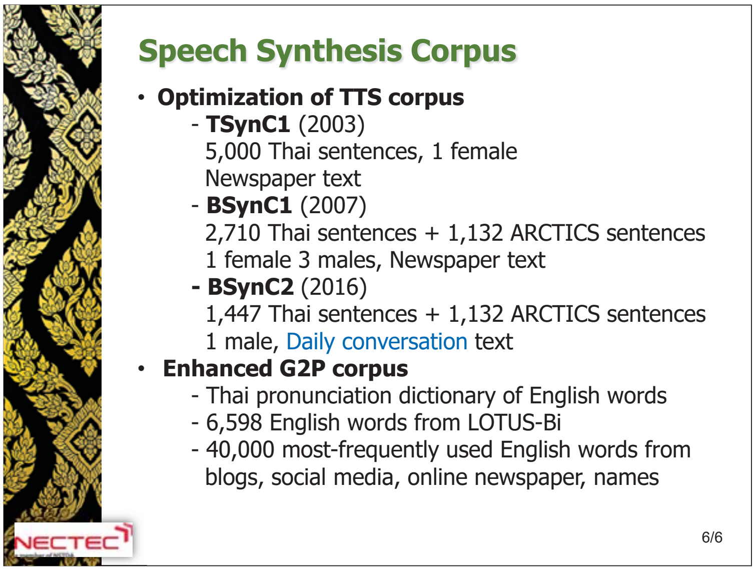
4/6



Cleft-palate Speech Corpus

- **Clinical trials**
  - **Phase 1** (2017): 30 cleft-palate and 30 normal children from two hospitals
  - **Phase 2** (2018): 111 cleft-palate and 111 normal children from four hospitals
  - **Test sets**: 28 sentences (4 screening, 15 high oral-pressure, 6 low oral-pressure, 3 nasal) and 42 words (28 high oral-pressure, 8 low oral-pressure, and 6 nasal) by Prathanee et al. (2011)
- **Research issues**
  - Could NASAM replace the conventional Nasometer for measuring nasalance score?
  - Could the cleft-palate and normal children be distinguished by their speech?
  - Could we embed an automatic assessment based on speech recognition?

5/6



Speech Synthesis Corpus

- **Optimization of TTS corpus**
  - **TSynC1** (2003)
    5,000 Thai sentences, 1 female
    Newspaper text
  - **BSynC1** (2007)
    2,710 Thai sentences + 1,132 ARCTICS sentences
    1 female 3 males, Newspaper text
  - **BSynC2** (2016)
    1,447 Thai sentences + 1,132 ARCTICS sentences
    1 male, Daily conversation text
- **Enhanced G2P corpus**
  - Thai pronunciation dictionary of English words
  - 6,598 English words from LOTUS-Bi
  - 40,000 most-frequently used English words from blogs, social media, online newspaper, names

6/6

Country Report - Timor Leste

2018 Country Report Timor Leste

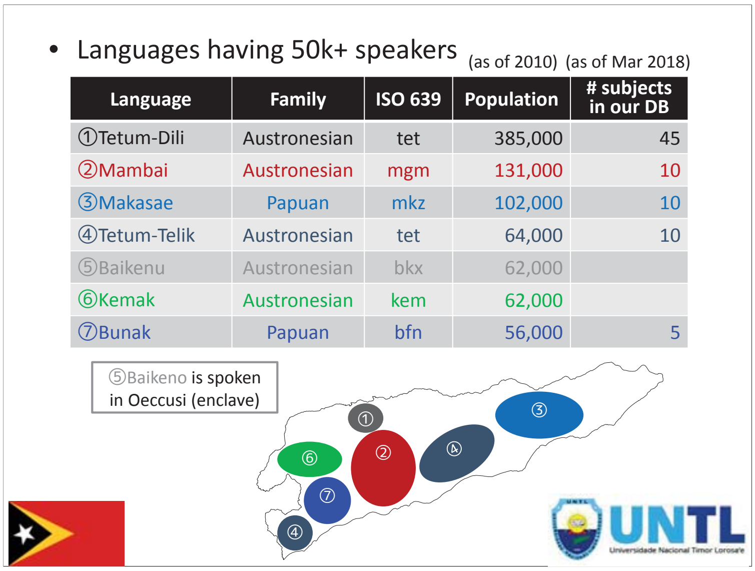
Borja L. C. Patrocinio Antonino
Informatics Departments
Faculty of Engineering, Science and Technology
Universidade Nacional de Timor Leste

Introduction

- Timor Leste, also known as East Timor, is one of Southeastern Asian countries, which is the first new sovereign state of the 21st century.
- Currently, Timor Leste has about 1.2 million people.
- Portuguese and Tetum are the official languages.
- There are about 30 indigenous languages including Tetum, widely spread over Timor Leste.

Ongoing Work

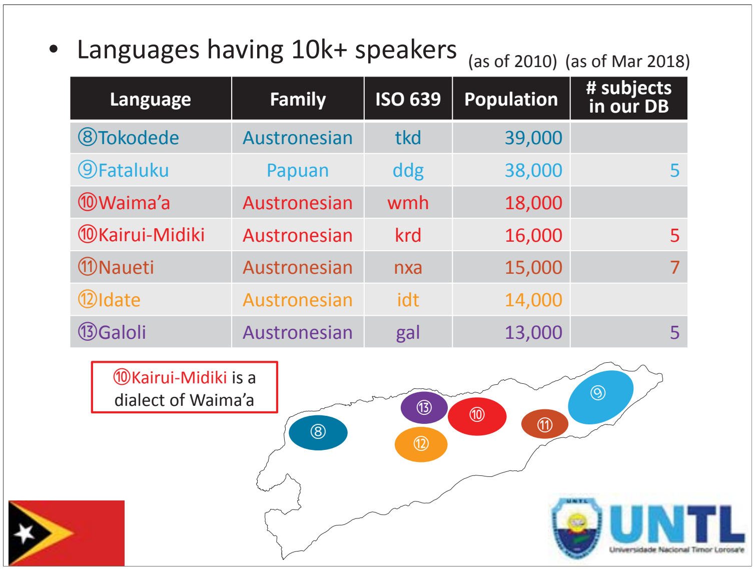
- Supported by JICA (Japan International Cooperation Agency) and Gifu University (Japan), we started to build speech corpora for several languages.
- Tasks
  - Greeting
  - Words
  - Connected digits



- Languages having 50k+ speakers (as of 2010) (as of Mar 2018)

Language	Family	ISO 639	Population	# subjects in our DB
①Tetum-Dili	Austronesian	tet	385,000	45
②Mambai	Austronesian	mgm	131,000	10
③Makasae	Papuan	mkz	102,000	10
④Tetum-Telik	Austronesian	tet	64,000	10
⑤Baikenu	Austronesian	bkx	62,000	
⑥Kemak	Austronesian	kem	62,000	
⑦Bunak	Papuan	bfn	56,000	5

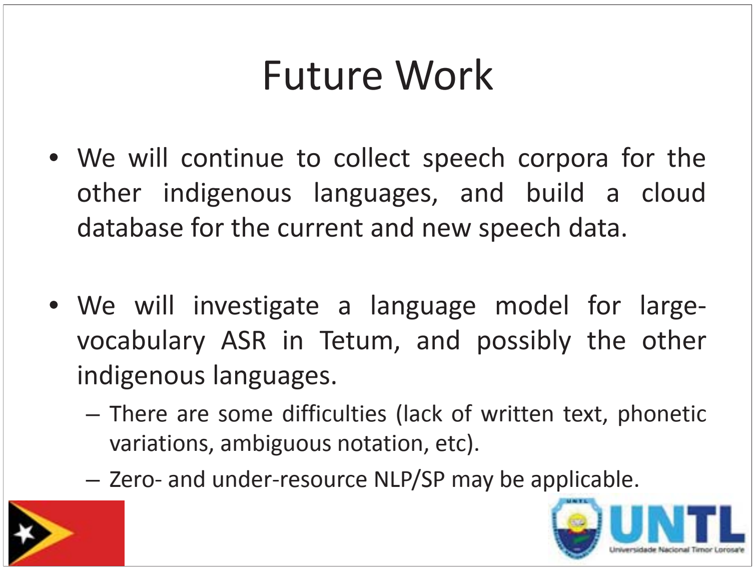
⑤Baikeno is spoken in Oecussi (enclave)



- Languages having 10k+ speakers (as of 2010) (as of Mar 2018)

Language	Family	ISO 639	Population	# subjects in our DB
⑧Tokodede	Austronesian	tkd	39,000	
⑨Fataluku	Papuan	ddg	38,000	5
⑩Waima'a	Austronesian	wmh	18,000	
⑩Kairui-Midiki	Austronesian	krd	16,000	5
⑪Naueti	Austronesian	nxa	15,000	7
⑫Idate	Austronesian	idt	14,000	
⑬Galoli	Austronesian	gal	13,000	5

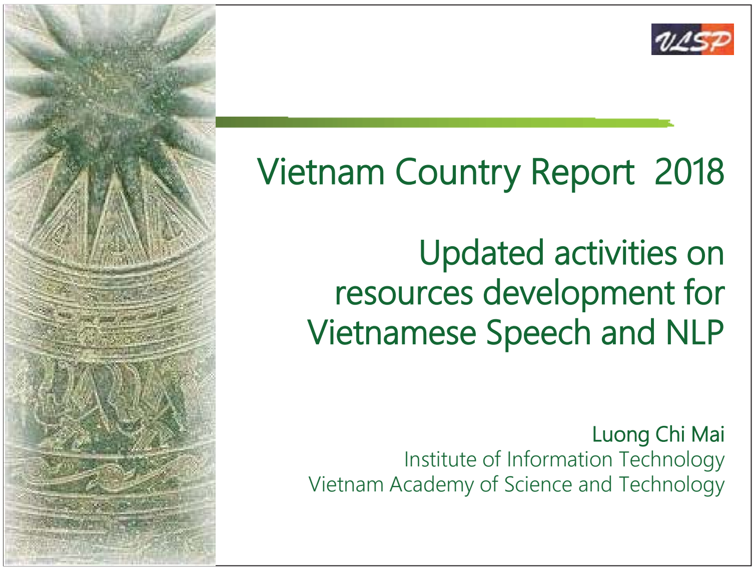
⑩Kairui-Midiki is a dialect of Waima'a



Future Work

- We will continue to collect speech corpora for the other indigenous languages, and build a cloud database for the current and new speech data.
- We will investigate a language model for large-vocabulary ASR in Tetum, and possibly the other indigenous languages.
  - There are some difficulties (lack of written text, phonetic variations, ambiguous notation, etc).
  - Zero- and under-resource NLP/SP may be applicable.

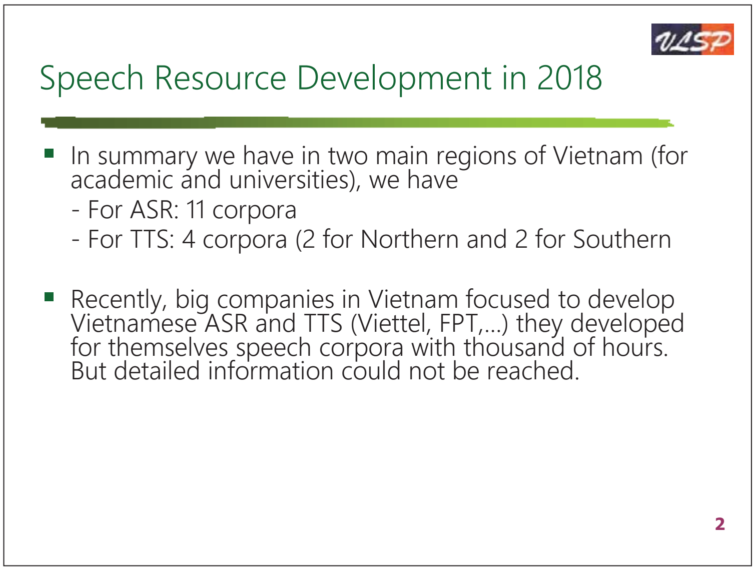
Country Report - Vietnam



Vietnam Country Report 2018

Updated activities on resources development for Vietnamese Speech and NLP

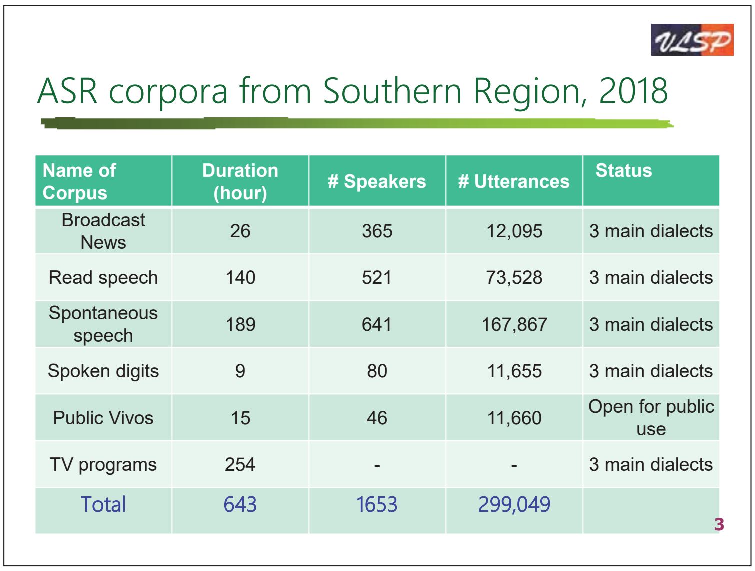
Luong Chi Mai
Institute of Information Technology
Vietnam Academy of Science and Technology



Speech Resource Development in 2018

- In summary we have in two main regions of Vietnam (for academic and universities), we have
  - For ASR: 11 corpora
  - For TTS: 4 corpora (2 for Northern and 2 for Southern
- Recently, big companies in Vietnam focused to develop Vietnamese ASR and TTS (Viettel, FPT,...) they developed for themselves speech corpora with thousand of hours. But detailed information could not be reached.

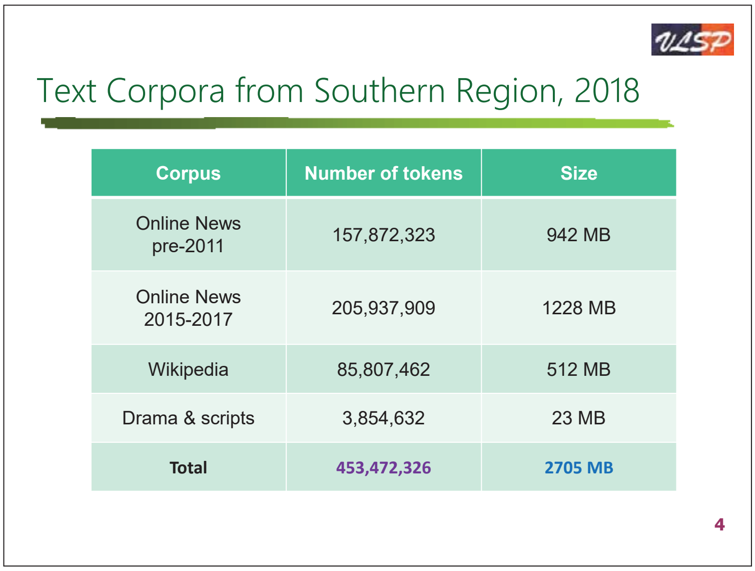
2



ASR corpora from Southern Region, 2018

Name of Corpus	Duration (hour)	# Speakers	# Utterances	Status
Broadcast News	26	365	12,095	3 main dialects
Read speech	140	521	73,528	3 main dialects
Spontaneous speech	189	641	167,867	3 main dialects
Spoken digits	9	80	11,655	3 main dialects
Public Vivos	15	46	11,660	Open for public use
TV programs	254	-	-	3 main dialects
Total	643	1653	299,049	

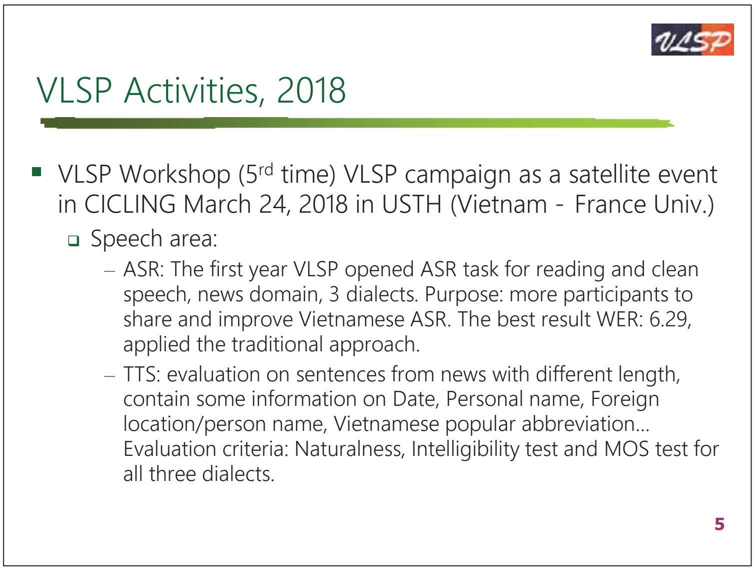
3



Text Corpora from Southern Region, 2018

Corpus	Number of tokens	Size
Online News pre-2011	157,872,323	942 MB
Online News 2015-2017	205,937,909	1228 MB
Wikipedia	85,807,462	512 MB
Drama & scripts	3,854,632	23 MB
Total	**453,472,326**	**2705 MB**

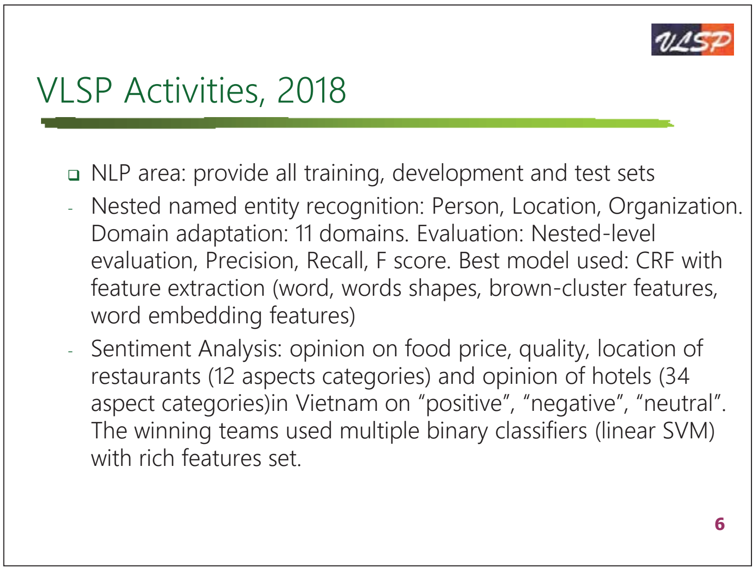
4



VLSP Activities, 2018

- VLSP Workshop (5rd time) VLSP campaign as a satellite event in CICLING March 24, 2018 in USTH (Vietnam - France Univ.)
  - ❑ Speech area:
    - ASR: The first year VLSP opened ASR task for reading and clean speech, news domain, 3 dialects. Purpose: more participants to share and improve Vietnamese ASR. The best result WER: 6.29, applied the traditional approach.
    - TTS: evaluation on sentences from news with different length, contain some information on Date, Personal name, Foreign location/person name, Vietnamese popular abbreviation... Evaluation criteria: Naturalness, Intelligibility test and MOS test for all three dialects.

5



VLSP Activities, 2018

- ❑ NLP area: provide all training, development and test sets
  - Nested named entity recognition: Person, Location, Organization. Domain adaptation: 11 domains. Evaluation: Nested-level evaluation, Precision, Recall, F score. Best model used: CRF with feature extraction (word, words shapes, brown-cluster features, word embedding features)
  - Sentiment Analysis: opinion on food price, quality, location of restaurants (12 aspects categories) and opinion of hotels (34 aspect categories)in Vietnam on "positive", "negative", "neutral". The winning teams used multiple binary classifiers (linear SVM) with rich features set.

6

Sponsored by

Arcadia
Computing Innovation

Supported by

Nara Institute of Science and Technology
NAIST ®

O-COCOSDA 2018
MIYAZAKI