

O-COCOSDA2010

Nov 24-25, 2010
Kathmandu, Nepal

O-COCOSDA 2010 Committee

Conference Chair

Professor Jai Raj Awasthi, Tribhuban University, Kathmandu

Conference Co-Chair

Dr. Chiu-yu Tseng, Academia Sinica, Taipei

Conference Secretary

Mr. Bhim Narayan Regmi, CECODES, Kathmandu

Organizing Committee

Mr. Krishna Prasad Parajuli, CECODES, Kathmandu

Mr. Sagun Dhakhwa, CECODES, Kathmandu

Mr. Gelu Sherpa, CECODES, Kathmandu

Mr. Kamal Poudel, CECODES, Kathmandu

Mr. Bhim Lal Gautam, CECODES, Kathmandu

Mr. Sandeep Khatri, CECODES, Kathmandu

Mr. Asoke Datta, CDAC, Kolkata

Dr. Om Vikas, CDAC, Noida

Pushpak Bhattacharya, IIT, Mumbai

Mrs. Swarn Lata, DIT, MCIT

Advisory Committee

Honorable Nilkantha Upreti, Executive Chief Commissioner, Election Commission, Nepal

Honorable Tirtha Raj Khaniya, Member, National Planning Commission, Nepal

Honorable Manohar Kumar Bhattarai, Vice Chairman, HLCIT, Nepal

Mr. Bairagi Kainla (Til Bikram Nembang), Chancellor, Nepal Academy

Prof. Dr. Surendra Raj Kafle, Vice Chancellor, Nepal Academy of Science and Technology

Professor Chura Mani Bandhu, Nepal

Professor Tej Ratna Kansakar, Nepal

Professor Novel Kishore Rai, CNAS, Tribhuvan University, Nepal

Professor Govinda Raj Bhattarai, Tribhuvan University, Nepal

B. Yegnanarayana, IIIT, Hyderabad, India

Hema Murthy, IIT, Madras, India

Professor Hiroya Fujisaki, Japan

D.D. Majumdar, ISI, Kolkata, India

P.V.S. Rao, Former-Head-CSC & Speech Tech. Group

Rajeev Sanjal, IIIT, Hyderabad, India

Dr. RMK Sinha, IIT, Kanpur, India

Professor Shuichi Itahashi, NII/AIST, Japan

President IETE

ED-CDAC, Noida, India

ED-CDAC, Kolkata, India

Director-IIIT, HYD, India

Technical Program Committee

Chairs

Dr. Shyam S. Agrawal (CDAC, Noida & KIIT, Gurgaon, India)

Professor Yoshinori Sagisaka (Waseda University, Japan)

Committee

Dr. Aijun Li (Chinese Academy of Sciences, Beijing, China)

Dr. B. Mallikarjun, CIIL, Mysore, India

Bimal Acharya, Nepal Telecom Company Limited, Nepal

Dr. Chai Wutiwivatchai (NECTEC, Thailand)

Dr. Chiu-yu Tseng (Academia Sinica, Taipei)

Dr. Haizhou Li (Institute of Infocom Research, Singapore)

Dr. Hammam Riza (BTTP, Indonesia)

Professor Hideaki Kikuchi (Waseda University, Japan)

Professor Hsiao-Chuan Wang, (National Tsing Hua University, Taiwan)

K K Arora, CDAC, Noida

Dr. K. Samudravijaya (Tata Institute of Fundamental Research, India)

Professor Kanika Kaur, KIIT, Gurgaon, India

Dr. Luong Chi Mai (The Vietnamese Academy of Sciences, Vietnam)

Professor Nick Campbell (Trinity College, Ireland)

Professor Rachael Roxas, (De la Salle University-Manila, Philippines)

Dr. Satoshi Nakamura (NICT/ATR, Japan)

Shyamal Das Mandal, CDAC, Kolkata

Swarn Lata, DIT, Govt. India

Professor Tan Lee (Chinese University of Hong Kong, Hong Kong)

Professor Thomas Fang Zheng (Tsing Hua University, Beijing)

Professor Yong Ju Lee (Wonkwang University, Korea)

Professor Zuraidah Mohd Don (University of Malaysia, Malaysia)

List of Reviewers

India

Mrs. Annu Khosla
 Mr. Ashok Dutta
 Dr. B. Mallikarjun
 Professor Girish Nath Jha
 Professor Hema Murthy
 Professor K. K. Arora
 Dr. K. Samudra Vijya
 Mrs. Kalika Bali
 Dr. M. P. Tripathi
 Dr. Madhusudan Singh
 Dr. Om Vikas
 Dr. P. K. Das
 Dr. P. K. Saxena
 Professor P. V. S. Rao
 Professor Pramod Pandey
 Dr. Preeti S Rao
 Dr. R. M. K Sinha
 Professor S. R. Savitri
 Dr. S. S. Agrawal
 Dr. Shyam Das Mandal
 Mrs. Swarn Lata

Professor Lin-shan Lee
 Professor Luong Chi Mai
 Professor Mirna Adriani
 Dr. Nattanun
 Professor Nick Campbell
 Professor P. C. Ching
 Professor S. Itahashi
 Professor Sakriani Sakti
 Professor Sarmad Hussain
 Dr. Shinsuke Sakai
 Professor Sin-Horng CHEN,
 Professor Tan Lee
 Professor Thomas Fang Zheng
 Professor Virach Sornlertlamvanich
 Professor Xia Wang
 Professor Xinhui Hu
 Professor Yih-Ru WANG
 Professor Yong Ju Lee
 Professor Yoshinori Sagisaka
 Professor Yuan-Fu LIAO
 Professor Yung-Hwan Oh

Japan, China, Taiwan, Malaysia, Honkong, Vietnam, Philippines, Korea, Pakistan

Professor Chai Wutiwiwatchai
 Dr. Chatchawarn Hansakunbuntheung
 Dr. Hammam Riza
 Professor Hideaki Kikuchi
 Dr. Hideharu Nakajima
 Professor Hiroya Fujisaki
 Professor Hsiao-Chuan WANG
 Professor Hsin-min WANG
 Professor Jesper Olson
 Dr. Jhing-Fa WANG
 Professor Kim-Teng Lua

Message from the Oriental-COCOSDA Convener

Welcome to the 13th Oriental-COCOSDA Conference at Katmandu, Nepal. The research community of the Oriental chapter of The International Coordinating Committee on Speech Databases and Speech I/O Systems and Assessment has been meeting annually since 1998, the meetings grown from workshops to conferences, and the Oriental-COCOSDA2010 is the first time for us to meet at the foot of the Himalayas. There will be activity reports from 14 countries/regions this year, namely, China, Hong Kong, India, Indonesia, Japan, Korea, Malaysia, Nepal, Pakistan, Philippine, Singapore, Taiwan, Thailand and Vietnam. We are further assured of how our community has been growing steadily.

The purpose of COCOSDA and O-COCOSDA meetings is mainly to bring together people working on speech databases and/or assessment methods of speech I/O of their own languages, to discuss the problems they are facing, to exchange experiences, and to suggest possible approaches, as well as to avoid possible overlapping of efforts, etc. Our common goals include collecting and preserve large amounts of speech data of various kinds using common platform, and providing unrestricted access of collected corpora towards research and development as well as performance assessment. Academically, the conference proceedings were indexed by IEEE Xplore, thanks to the organizers of Oriental-COCOSDA 2009. While this act proved to be positive for some researchers in the region to secure travel funds and receive due credits for sound scientific work, regrettably it cannot be carried out this year after many administrative and financial difficulties encountered by the local organizers. I would like to take this opportunity to remind the participants and supporters of Oriental-COCOSDA that in addition to the unique linguistic features we share as a region, our organization is also unique. It has functioned independently for over a decade, going stronger even, and without any financial support from the start. I believe maintaining the uniqueness and independence of Oriental-COCOSDA should take precedence so it continues to be a platform and venue for both speech scientists and linguists in this region to meet continuously for many annual conferences to come, while at the same time, we will try to have the

proceedings indexed whenever possible.

As your convener I also ask you all to help to promote membership in your respective region, seek funds for cross-country collaborative research, and share your resources with the Oriental-COCOSDA community.

Last but not least, I would like to thank the Nepali colleagues headed by Conference Chair Professor Jai Raj Awasthi of Tribhuban University, Kathmandu and Dr. Bhim Narayan Regmi of CECODES, Kathmandu; Technical Committee Chair Dr. Shaym S. Agrawal of CDAC, Noida & KIIT, Gurgaon, India, and Technical Committee Co-chair Professor Yoshinori Sagisaka of Waseda University, Japan for making the event possible.

Chiu-yu TSENG
Oriental COCOSDA Convener

Message from the Conference Chair

I am extremely delighted to welcome you all to the 13th Oriental-COCOSDA conference in Kathmandu, Nepal. The main goal of the conference is to bring the people who are involved in the spoken languages of this region. It will further provide an opportunity for a meeting of minds of the people enthusiastic to share their research findings and explore the possibilities of future collaborations.

Nepal, being a multi-lingual, multi-ethnic and multicultural country, houses more than 92 languages and many of them are not yet studied scientifically. Since this conference is the first of its kind in this country participated by the scholars of more than 15 countries, it will provide a rare opportunity to the local scholars to learn as to how to study the yet to be identified languages of this country. It will further teach us as to how to develop corpora of the languages of this country. More importantly, the conference will create awareness among the teachers, researchers and students towards the field of studies covered in the conference presentations and to provide the platform for the concerned Nepali institutions and people to collaborate with the international institutions and people.

I would like to thank so many people, and institutions who have enthusiastically supported us to materialize our long cherished desire to hold the present conference in Nepal. We are highly indebted to ICSCA for supporting the travel of Dr. Glass; KIIT groups of colleges for supporting the space, equipment, and manpower to assist the technical committee chair; HLCIT, Nepal Academy and NAST for financial assistance.

Similarly, we are grateful to Hetauda School of Management and Social Sciences (HMSS) and Speech Ocean (Beijing Haitian Ruiseng Science Technology Ltd.) and Outlines Research and Development for sponsoring the conference activities. Equally grateful are we to the media partner, Radio Sagarmatha for the publicity of the conference.

The conference could not have been convened without the initiation taken by Centre for Communication and Development Studies, the host institution- a non-profit, non-governmental organization established:

- to study and fill up the gaps on the specific theoretical, applied,

and interdisciplinary aspects of language, communication, and development.

- to empower and assist the ethnic communities, organizations, and related institutions to enhance their language and communication related complexities for resources mobilization and development.
- to share and transfer knowledge, skill, and technologies at national and international level through seminars, workshops, and conferences.
- to recommend the government decision making bodies at different levels for policy making in the relevant areas on the basis of research findings.

Most importantly, I would like to express our gratitude to Dr. Chiu-yu TSENG, Oriental-COCOSDA Convener, Technical committee Chair Dr. Shyam S. Agrwal of CDAC, NOIDA and KIIT, India and Technical Committee Co-Chair Professor Yoshinori Sagisaka of Waseda University, Japan for their untiring support to make the conference a success. Similarly, my colleagues here in Nepal, particularly Mr. Bhim Narayan Regmi who not only proposed Nepal to host the present conference but also tried day and night to have the dream come true, and Mr. Sagun Dhakhwa, Mr. Kamal Paudel and many more for their special contributions.

I wish the conference and grand success and assure all O- COCOSDA members present here for our continued support to this community in future as well.

Jai Raj Awasthi, Ph.D.
Conference Chair

Report of the Technical Committee

O-COCOSDA2010

I on my own behalf and on behalf of the Technical Programme committee welcome all the participants and authors of papers to the 13th Oriental COCOSDA meeting and conference being held in Kathmandu, Nepal for first time. The conference has attracted a large number of participants' researchers and scientists from all over the world. For the first time it has been supported by ISCA. During past several years, the importance of Oriental COCOSDA has been recognized by many Asian Countries and also by European countries and the United States of America. Interest in its participation is increasing year by year. This year papers have been submitted from 14 Countries and regions. The papers cover a wide range of topics, country wise submission & acceptance of papers is shown in Table 1 below.

S.No.	Country	Papers Received
1.	China	09
2.	India	26
3.	Ireland	01
4.	Japan	13
5.	Korea	01
6	Malaysia+ Singapore	03
7	Nepal	04
8.	Pakistan	02
9.	Philippines	01
10.	Taiwan	04
12	Thailand	02
13	Vietnam	01
Total		67

After a rigorous review process, out of 67 submitted papers, 37 papers were selected for general oral presentation and 26 papers for poster presentation. All the papers will be published in CD. The abstracts and the program will be published in an Abstract Book.

In addition to regular sessions, we have four distinguished speakers to deliver keynote addresses

Keynote 1: Dr. James Glass (Jim), USA

Keynote 2: Professor Keiichi TOKUDA, Japan

Keynote 3: Dr. Pramod Kumar Saxena, India

Keynote 4: Professor Madhav Prasad Pokharel, Nepal

In addition, there would be special session for presentation of the country reports by the respective distinguished country representatives. These are given in Table 2:

S.No.	Country	Country Representative
1	China	Prof. Thomas Fang Zheng, Prof. Aijun Li
2	Hong Kong	Prof. P. C. Ching, Prof. Tan Lee
3	India	Dr.S. S. Agrawal
4	Indonesia	Dr. Hamam Riza
5	Japan	Dr. S. Nakamura, Prof. S.Itahashi
6	Korea	Prof. Young Ju Lee, Prof. Sougil Ann
7	Malaysia	Dr. Zuraidah Mohd Don
8	Mongolia	Dr. Dawa Idomuco
9	Nepal	Mr.Bhim Narayan Regmi
10	Pakistan	Dr. Sarmad Hussain
11	Philippines	Prof. Rachael Roxas, Dr. Jocelyn Cu
12	Singapore	Prof. Haizhou Li, Prof. Kim-Teng Lua
13	Taiwan	Prof. Lin-shan Lee, Prof. H.C.Wang
14	Thailand	Dr. Chai Wutiwatchai, Dr. Virach Sornlertlamvanich, Dr. Thanaruk Theeramunkong
15	Vietnam	Prof. Luong Chi Mai

In addition, there would be a few other special presentations also

O-COCOSDA Book Presentation	: Prof. S. Itahashi
A-Star /U- Star Project	: Prof. S. Nakamura
AESOP Activities	: Prof. Y. Sagisaka
NCP...	: Prof. D. Datta Majumdar

The challenging task of reviewing the papers was handled by the members of the Technical Programme Committee and other experts working in these areas. We would like to thank them for their untiring efforts and timely support. We would like to thank, Prof. Chiu-yu-Tseng, Prof. S. Itahashi, Prof. S. Nakamura, Prof. H. Fujisaki for their continuous guidance, Prof. Jairaj Awasthi and Mr. Bhim Regmi for their cooperation, KIIT management Dr. Harsh Vardhan, Mrs. Neelima Karmah for their support and provision of facilities. Special thanks to Ms. Kanika Kaur for her untiring assistance and help during the entire technical work.

Finally, we thank all the contributors of OC-2010, the Chairpersons and other functionaries for making it a great success.

Dr. Shyam S. Agrawal

Technical Chair

Prof. Yoshinori Sagisaka

Co-Chair

Technical Program of Oriental COCOSDA-2010

Day 1, 24th November 2010

S.N. Event/Timing

0. **Registration / Conference Material Collection**
8:00 am – 8:30 am
1. **Inauguration Ceremony**
8:30am -9:00am
2. **High Tea**
9:00am – 9:30am
3. **Key Note-1 by Dr. James Glass (Jim),USA**
9:30am – 10:15am
4. **Technical Session-1 (Oral presentation 1)**
10:15am -12:15pm
 - 1 **An Analysis of a Mandarin-English Code-switching Speech Corpus: SEAME**
Dau-Cheng Lyu, Tien-Ping Tan, Eng-Siong Chng, Haizhou Li
 - 2 **Applying Pitch Based Dynamic Pruning in Designing Real-Time Speaker Identification System**
Soma Khan, Joyanta Basu, Shyamal Kumar Das Mandal
 - 3 **Discourse Prosody Planning in L1 and L2 English**
Tanya Visceglia, Chiu-yu Tseng, Zhao-yu Su, Chi-Feng Huang
 - 4 **A Proposal for Standardizing Catalogue Specifications of Speech Corpora**
S. Itahashi, K. Yamakawa, T. Matsui, Y. Ishimoto
 - 5 **NICT Speech and Language Resources and Corpora**
Satoshi Nakamura, Kentaro Torisawa, Hisashi Kawai, Eiichiro Sumita
 - 6 **Spoken Disfluencies in Multilingual Spoken Corpora**
Samudravijaya K.
 - 7 **An ASR System for Spontaneous Urdu Speech**
Agha Ali Raza, Sarmad Hussain, Huda Sarfraz, Inam Ullah, Zahid Sarfraz
 - 8 **Speech Corpus Development for a Speaker Independent Spontaneous Urdu Speech Recognition System**
Huda Sarfraz, Sarmad Hussain, Riffat Bokhari, Agha Ali Raza, Inam Ullah, Zahid Sarfraz, Sophia Pervez, Asad Mustafa, Iqra Javed, Rahila Parveen

5. **Key Note -2 by Professor Keiichi TOKUDA, JAPAN**
12:15 pm – 1:00 pm
6. **Lunch Break**
1:00 pm- 2:00 pm
7. **Technical Session-2 (Oral presentation 1)**
2:00 pm – 4:00 pm
 - 9 **Phonation Type of Korean Stops - Research Based On Data Retrieved From Unified Acoustic Parameter Database**
Zhou Xuewen, Zheng Yuling, Chuai Zhenyu
 - 10 **An HMM-based Hakka Text-to-Speech System**
Yi-Ling Tsai, Hsiu-Min Yu, Yih-Ru Wang, Chen-Yu Chiang, Lieh-Shih Lo, Sin-Horng Chen
 - 11 **Speech Synthesis Using Epoch Synchronous Overlap Add (ESOLA)**
Ashoke Kr Datta, Arup Saha
 - 12 **Development of speech data base for various emotions and their recognition using Neural Network Classifier**
Jyoti Garg, Israr Khan , S.K.Gupta, S.S.Agrawal
 - 13 **Mental-State Analysis for Understanding Children's Behavior Based on Emotion-Label Sequences in Multimodal Speech-Behavior Corpus**
Shinya Kiriyama, Shogo Ishikawa, Shigeyoshi Kitazawa, Yoichi Takebayashi
 - 14 **Mora Pitch Level Recognition for the Development of a Japanese Pitch Accent Acquisition System**
Greg Short, Keikichi Hirose, Takeshi Yamada, Nobuaki Minematsu, Nobuhiko Kitawaki, Shoji Makino
 - 15 **Speech Technology and Empathy in Conversational Interaction**
Nick Campbell
 - 16 **Analysis and Synthesis of F_0 Contours for Bangla Readout Speech**
Shyamal Das Mandal, Anal Haque Warsi, Tulika Basu, Keikichi Hirose, Hiroya Fujisaki
8. **Tea Break**
4:00 pm – 4:15 pm
9. **Technical Session-3 (Poster presentation)**
4:15 pm – 5:45 pm
- 1 **Method for Collection of Diverse Speech for Emotion Research Database**
Takahiro Miyajima, Takeshi Fukuda, Hideaki Kikuchi, Katsuhiko Shirai
- 2 **Aerodynamic Study of Standard Mongolian Tense-lax Vowels**
Hu Axu, Gegen Tana, Yu Hongzhi
- 3 **Creation and Analysis of a Japanese Speaking Style Parallel Database for Expressive Speech Synthesis**
Hideharu Nakajima, Noboru Miyazaki, Akihiro Yoshida, Takashi Nakamura, Hideyuki Mizuno
- 4 **Garhwali Speech Database**
RK Upadhyay, MK Riyal
- 5 **Affect Recognition from Acted and Spontaneous Filipino Speech**
Jocelynn Cu
- 6 **Valency Analyzer of Verbs Arguments for Bangla**
Subhash Chandra, Pampa Bhattacharyya
- 7 **Performance evaluation of word boundary detection for Hindi speech database**
Anurag Jain, S.S. Agrawal, Nupur Prakash
- 8 **Sanskrit Verb Argument Valence: A Computational Analysis**
Subhash Chandra, Dr. Girish Nath Jha
- 9 **A Metric-based Phone Segmentation Method Using Wavelet Transform**
Ching-Feng Lu, Hsiao-Chuan Wang
- 10 **An Overview of Tibetan Corpus Construction**
Yu Hongzhi, Gao Lu, Guo Lei, Kou Jianqun
- 11 **Multi-channel Speech Data Collection on Mobiles**
Lin HE, Yufeng HAO, Ke LI, Xianfeng CHEN
- 12 **Introduction of Speech Ocean Multi-languages In-Car Project**
Lin HE, Yufeng HAO, Xianfeng CHENG, Ke LI
- 13 **An anti-noise MFCC extraction algorithm for speaker recognition**
Wang Hong Pan Jin'gui, Wang Hong
- 14 **Syntactic and Semantic analysis of Bangla language for developing grammar checker system**
Bibekananda Kundu

- 15 **IPTV / Navigation Environment Adaptation Speech DB and Usability Test of IPTV VOD Retrieval**
Dae-Lim Choi, Bong-Wan Kim, Yong-Ju Lee, Byung-Ok Kang, Eui-Sok Chung, Yun-Keun Lee, Gyu-Tae Baek, Ki-Hyung Hong
- 16 **An Approach to Mixed Language Automatic Speech Recognition**
Kiran Kumar Bhuvanagiri, Sunil Kumar Kopparapu
- 17 **Building a Cross Script Kashmiri Converter: Issues and Solutions**
Aadil Amin Kak, Nazima Mehdi, Aadil Ahmad Lawaye
- 18 **Present Scenario of Forensic Speaker Identification in India**
Shivani Sharma, S. K. Jain, R. M. Sharma, S.S Agrawal
- 19 **Phonetic Segmentation Based on HMM of Hindi Speech**
Archana Balyan, S.S. Agrawal, Amita Dev
- 20 **Recognition of Hindi Phoneme in Rhyming Words using Vector Quantization**
Shweta Sinha, Shalini Goyal, Mona Gaur, S.S Agrawal
- 21 **Intonation Patterns in Nepali Feedback Units**
Jens Allwood, Bhim Narayan Regmi
- 22 **Taiwan L2 German Database Design for Computer Assisted Language Learning**
Chia-yu Chiu, Yuan-fu Liao, Hansjörg Mixdorff, Hue-San Do, Shing-lung Chen
- 23 **Cases on Extension of Language Technology into the Related Fields and their Implications for Research and Development in University**
Satoru HAYAMIZU, Tadahiro MATSUMOTO, Satoshi TAMURA, Shinichi TAKEUCHI
- 24 **Application of Discriminative Training Technique for English Pronunciation Evaluation**
Jun Qi, Weiqian Liang, Runsheng Liu, Ruiying Wei
- 25 **Hindi ASR for Travel Domain**
Sunita Arora, Babita Saxena, Karunesh Arora, S S Agrawal
- 26 **Speaker Recognition Based on Multilingual Speech Features using Neural Network Model**
Sanjay Decate, Sanjay Kumar Singh, Anupam Shukla, Ritu Tiwari
10. **Dinner**
6:30 pm-7:45 pm

Day 2, 25th November 2010

S.N. Event/ Timing

1. **Key Note -3 by Prof. Pramod Kumar Saxena, INDIA**
8:30am - 9:15am
2. **Technical Session-4 (Oral presentation 1)**
9:15am – 10:30am
 - 17 **The Use of Indonesian Speech Corpora for Developing Filipino Continuous Speech Recognition System**
Sakriani Sakti, Ryosuke Isotani, Hisashi Kawai, Satoshi Nakamura
 - 18 **Utsunomiya University Spoken Dialogue Database for Paralinguistic Information Studies**
Hiroki Mori, Hideki Kasuya, Makoto Nakamura
 - 19 **Description of Puma, an endangered language from eastern Nepal**
Vishnu S Rai
 - 20 **Real world utterance collection using voice-enabled web system for child speaker identification**
Shoko MIYAMORI, Ryuichi NISIMURA, Lisa KURIHARA, Toshio IRINO, Hideki KAWAHARA
 - 21 **Building Unit Selection Speech Synthesis in Indian Languages: An Initiative by an Indian Consortium**
Hema A. Murthy, Ashwin Bellur, Vinodh Viswanath, Badri Narayanan, Anila Susan, G Kasthuri, K. Sreenivasa Rao, Sudhamay Maity, N. P. Narendra, Ramu Reddy, Krishnendu Ghosh, K. G. Sulochana, E. L. Abhilash, T. Sajini, M. Sasikumar, Bira Chandra Singh, Pranaw Kumar, P. Vijayaditya, E. Veera Raghavendra, Kishore Prahallad
3. **Tea Break**
10:30am - 10:45 am
4. **Technical Session-5 (Oral Presentation 2)**
10:45am -12:45 pm
 - 1 **An Independent Approach for Spoken Language Analysis**
Arup Saha, Tulika Basu, Ashoke Kr. Datta
 - 2 **Development of a Malay LVCSR System**
Xiong Xiao, Eng Siong Chng, Tien-Ping Tan, Haizhou Li
 - 3 **A Chunk Level Statistical Machine Translation (An Approach for English Language to Nepali Language Translation)**
Ashim Ghishing, Bikash Balami, Yoga Raj Joshi

- 4 **A Hybrid Speech Enhancement System Based On Wavelet Denoising**
Richa Tyagi, Sunita Maithani
- 5 **Dialogue-Act Analysis with a Conversational Telephone Corpus Recorded in Real Scenarios**
Keyan Zhou, Aijun Li, Chengqing Zong
- 6 **The Development of the Large Thai Telephone Speech Corpus: LOTUS-Cell 2.0**
Ananlada Chotimongkol, Nattanun Thatphithakkul, Sumonmas Purodakananda, Chai Wutiwiwatchai, Patcharika Chootrakool, Chatchawarn Hansakunbuntheung, Atiwong Suchato, Panuthat Boonpramuk
- 7 **Hindi Dialects Phonological Transfer Rules for Verb Root Cələ**
Diwakar Mishra, Kalika Bali
- 8 **Feasibility of the Characterisation Control by Text-based Speech Style**
Raymond SHEN, KIKUCHI Hideaki, OHTA Katsumi, MITAMURA Takeshi
- 9 **Age Group Estimation based on Acoustic Analysis of Speech**
Devendra Kumar Yadav, Kamini Malhotra, Anu Khosla
- 10 **Orthography development for the standardization of Bhujel: Issues and approaches**
Dan Raj Regmi
- 11 **Analysis of impression-prosody mapping in communicative speech consisting of multiple lexicons with different impressions**
Yoko Greenberg, Hiroaki Kato, Minoru Tsuzaki, Yoshinori Sagisaka
- 12 **THE DEVELOPMENT OF A SINGAPORE ENGLISH CALL RESOURCE**
Wenda Chen, Ying Ying Tan, Eng Siong Chng, Haizhou Li
5. **Lunch Break**
12:45 pm – 1:45pm
6. **Key Note -4 by Professor Madhav Prasad Pokharel, NEPAL**
1:45 pm-2:30 pm
7. **Technical Session-6 (Oral presentation 2)**
2:30 pm – 3:10 pm
- 13 **Creation of Time-Varying Voiceprint Database**
Linlin Wang, Thomas Fang Zheng
- 14 **Chinese Language Model Adaptation Using Semi-Supervised Approach**
Xinhui Hu, Ryosuke Isotani, Hisashi Kawai, Satoshi Nakamura
- 15 **Language Identification using Support Vector Machine**
Sanghamitra Mohanty, Basanta Kumar Swain
- 16 **The Development of a Thai Speech Set for Telephonometry**
Therdpong Daengsi, Apiruck Preechayasomboon, Saowanit Sukparungsee, Patcharika Chootrakool, Chai Wutiwiwatchi
8. **Tea Break**
3:10 pm – 3:25 pm
9. 1 **Country Presentation of 15 countries through their country representatives**
3:25 pm- 4:45 pm
- 2 **A-STAR/MASTAR Project Presentation by Dr. Satoshi Nakamura**
4:45pm – 4:55 pm
- 3 **Book Report Presentation by Prof. S.Itahashi**
4:55 pm- 5:05 pm
- 4 **AESOP workshop report presentation by Prof. Yoshinori Sagisaka**
5:05pm- 5:15pm
- 5 **Presentation by Prof. Datta Majumdar**
5:15pm - 5:25pm
- 6 **O-COCOSDA 2011 by Prof. Hsiao Chuan Wang**
5:25pm - 5:35pm
10. **Closing Ceremony**
5:35 – 6:00 pm

Keynotes

Crowds and Clouds for Spoken Language Processing

Jim Glass

Speech Synthesis as A Machine Learning Problem

Keiichi Tokuda

Speech Signal Processing for Secure Communication

P. K. Saxena

Phonetic Typology of Nepalese languages

Madhav P Pokharel

Crowds and Clouds for Spoken Language Processing

Jim Glass

MIT Computer Science and Artificial Intelligence Laboratory, USA

In this talk I discuss how crowdsourcing techniques and cloud-based services can lead to a paradigm shift in spoken language processing development. Through these new opportunities, distributed and unstructured knowledge sources can be leveraged by exploiting weakly-supervised learning in place of expert supervision. I illustrate several different ways in which crowdsourcing can be used to accelerate spoken language technology development, including data collection and annotation, as well as system assessment. I also present recent work that highlights how crowdsourcing can be used for learning new word pronunciations, and how unstructured opinions can be summarized to enhance spoken information access. These capabilities will enable a new breed of spoken language technology that can operate semi-autonomously in a cloud-based environment and use crowdsourcing techniques to expand their capabilities. The talk will include several demonstrations of prototype web-based spoken dialogue systems that leverage an open-source framework we have developed to deploy spoken language technology on a variety of web-enabled platforms.

Speech Synthesis as A Machine Learning Problem

Keiichi Tokuda

Nagoya Institute of Technology, Nagoya, Japan

The problem of speech synthesis can be stated as follows: We have a speech database, i.e., a set of speech waveforms and corresponding texts. Given a text to be synthesized, what is the speech waveform corresponding to the text?

This talk will formulate the basic problem of speech synthesis in a statistical machine learning framework and discuss how we can decompose it into feasible subproblems. One of the subproblems would be the statistical parametric speech synthesis, which is called “HMM-based speech synthesis” when we use hidden Markov models (HMMs) as statistical models. The talk will also discuss future challenges and the direction in speech synthesis research.

Speech Signal Processing for Secure Communication

P. K. Saxena

Scientific Analysis Group, DRDO, Metcalfe House, Delhi-110054, India

Communication is one of the important areas of Electronics and Speech is the most natural means of Communication. The speech is produced through our articulatory organs (tongue, lips, jaw, vocal chords etc) as a result of resonance in the vocal tract when the air is pushed through larynx (voice box) from the lungs. The physiology of individuals not only makes the speech a fingerprint of the speaker but also facilitates embedding behavioral information such as emotions, mood, health etc. Humans have unique capability in producing speech with immense amount of variability on one side and to listen and understand such sounds even with distortions and noise. Our cognitive capabilities enable us even to fill in the sounds which are not present otherwise while we speak. The high redundancy present in speech is exploited in speech modeling as well as speech synthesis and many time the intelligibility is the focus rather than the quality. So is the case when one likes to protect the speech communication going on channel. Speech redundancy helps us in designing mechanism to modify the signal in such a way that it is unintelligible while travels but can be recovered by the authorized receiver at the other end.

The production of speech is modeled based on resonance cavities and speech is approximated by combinations of sign waves of different frequencies present in speech (infinite in number but selected few in practice for study). Energy is another parameter which is prominent in characterizing speech. One can play around time, frequencies, amplitudes, energies and transform them to distort the signal for achieving secrecy. Yet another way to achieve is to digitize the speech, using different types of coding techniques and use technology available to encrypt digital signal.

In this talk such aspects of speech signal processing and means of protecting speech communication will be discussed

Phonetic Typology of Nepalese languages

Madhav P Pokharel

Central Department of Linguistics, Tribhuvan University, Nepal

Out of the three airstream mechanisms, in Nepal the velaric is missing; the glottalic is used by four of the geographically contiguous Kiranti languages and the pulmonic is default.

Nepal uses both the directions of the air. A few ingressive sounds are found in the above mentioned four of the Kiranti languages. Out of the two types of sounds found in the glottalic airstream Nepal is missing ejectives.

The ingressive lateral sound found in Umbule (or Wambule) Rai stands out in the phonetic literature. This sound cannot be fully described by the existing grids provided by the IPA which provides templates to classify only for the pulmonic and velaric sounds, but the ingressive lateral in Umbule is the glottalic ingressive sound.

The languages of Nepal do not use creaky sounds; therefore the state of the glottis appropriate for the articulation of creaky sounds is missing from Nepal. Rests of the phonation types provided in the phonetic literature are found in Nepal. In most of the Bodish languages and in a few Himalayish languages the low tone is marked by breathiness. Tonal languages in Nepal have register tones.

Nasalization is a characteristic in almost one-third of the Nepalese languages. Nepalese languages use all the stricture types or manners of articulation provided by the IPA; therefore they have plosives, nasals, a trill, flaps/taps, both lateral and median fricatives and approximants.

Nepalese languages use all the ten places of articulation.

Vowels in Nepalese languages contrast for height, backness, lip rounding, nasalization, tones (laryngeal height, breathiness, devoicing and aspiration) and length/duration. Only Sunuwar has phonemic syllable stress.

Syllabic nasals are found in Yakkha, Bantawa and Dhimal.

Branching in the coda consonants is in a very few Nepalese languages. In the branching of syllable peripheral sounds, there is a dominant role of liquids and glides to respect sonority hierarchy.

Papers

Contents

SN	Title	Page
1	An ASR System for Spontaneous Urdu Speech Agha Ali Raza, Sarmad Hussain, Huda Sarfraz, Inam Ullah, Zahid Sarfraz	35
2	An anti-noise MFCC extraction algorithm for speaker recognition Wang Hong Pan Jin'gui, Wang Hong	35
3	Syntactic and Semantic analysis of Bangla language for developing grammar checker system Bibekananda Kundu	36
4	Speech Corpus Development for a Speaker Independent Spontaneous Urdu Speech Recognition System Huda Sarfraz, Sarmad Hussain, Riffat Bokhari, Agha Ali Raza, Inam Ullah, Zahid Sarfraz, Sophia Pervez, Asad Mustafa, Iqra Javed, Rahila Parveen	36
5	Valency Analyzer of Verb Arguments for Bangla Subhash Chandra, Pampa Bhattacharyya	37
6	Phonation Type of Korean Stops - Research Based On Data Retrieved From Unified Acoustic Parameter Database Zhou Xuewen, Zheng Yuling, Chuai Zhenyu	37
7	Performance evaluation of word boundary detection for Hindi speech database Anurag Jain, S. S. Agrawal, Nupur Prakash	38
8	An Analysis of a Mandarin-English Code-switching Speech Corpus: SEAME Dau-Cheng Lyu, Tien-Ping Tan, Eng-Siong Chng, and Haizhou Li	39
9	Sanskrit Verb Argument Valence: A Computational Analysis Subhash Chandra, Girish Nath Jha	39
10	A Metric-based Phone Segmentation Method using Wavelet Transform Ching-Feng Lu, Hsiao-Chuan Wang	40
11	Development of a Malay LVCSR System Xiong Xiao, Eng Siong Chng, Tien-Ping Tan, Haizhou Li	40

12	IPTV / Navigation Environmental Speech DB and Usability Test of IPTV VOD Retrieval	41	22	Mental-State Analysis for Understanding Children's Behavior Based on Emotion-Label Sequences in Multimodal Speech-Behavior Corpus	47
	Dae-Lim Choi, Bong-Wan Kim, Yong-Ju Lee, Byung-Ok Kang, Eui-Sok Chung, Yun-Keun Lee, Gyu-Tae Baek, Ki-Hyung Hong			Shinya Kiriya, Shogo Ishikawa, Shigeyoshi Kitazawa, and Yoichi Takebayashi	
13	A Chunk Level Statistical Machine Translation (An Approach for English Language to Nepali Language Translation)	41	23	A Hybrid Speech Enhancement System Based On Wavelet Denoising	47
	Ashim Ghishing, Bikash Balami, Yoga Raj Joshi			Richa Tyagi, Sunita Maithani	
14	Applying Pitch Based Dynamic Pruning in Designing Real-Time Speaker Identification System	42	24	Age Group Estimation based on Acoustic Analysis of Speech	48
	Soma Khan, Joyanta Basu, Shyamal Kumar Das Mandal			Devendra Kumar Yadav, Kamini Malhotra, Anu Khosla	
15	Discourse Prosody Planning in L1 and L2 English	43	25	Aerodynamic Study of Standard Mongolian Tense-lax Vowels	48
	Tanya Visceglia, Chiu-yu Tseng, Zhao-yu Su, Chi-Feng Huang			Hu Axu, Gegentana, Yu Hongzhi	
16	A Proposal for Standardizing Catalogue Specifications of Speech Corpora	43	26	Creation of Time-Varying Voiceprint Database	49
	S. Itahashi, K. Yamakawa, T. Matsui, Y. Ishimoto			Linlin Wang, Thomas Fang Zheng	
17	Method for Collection of Diverse Speech for Emotion Research Database	44	27	Mora Pitch Level Recognition for the Development of a Japanese Pitch Accent Acquisition System	49
	Takahiro Miyajima, Takeshi Fukuda, Hideaki Kikuchi, Katsuhiko Shirai			Greg Short, Keikichi Hirose, Takeshi Yamada, Nobuaki Minematsu, Nobuhiko Kitawaki, Shoji Makino	
18	Feasibility of the Characterisation Control by Text-based Speech Style	44	28	Speech Technology and Empathy in Conversational Interaction	50
	Raymond SHEN, KIKUCHI Hideaki, OHTA Katsumi, MITAMURA Takeshi			Nick Campbell	
19	An Approach to Mixed Language Automatic Speech Recognition	45	29	Analysis and Synthesis of F₀ Contours for Bangla Readout Speech	50
	Kiran Kumar Bhuvanagiri, Sunil Kumar Kopparapu			Shyamal Das Mandal, Anal Haque Warsi, Tulika Basu, Keikichi Hirose, Hiroya Fujisaki	
20	Cases on Extension of Language Technology into the Related Fields and their Implications for Research and Development in University	46	30	Creation and Analysis of a Japanese Speaking Style Parallel Database for Expressive Speech Synthesis	51
	Satoru HAYAMIZU, Tadahiro MATSUMOTO, Satoshi TAMURA, Shinichi TAKEUCHI			Hideharu Nakajima, Noboru Miyazaki, Akihiro Yoshida, Takashi Nakamura, Hideyuki Mizuno	
21	The Development of a Singapore English CALL Resource	46	31	Application of Discriminative Training Technique for English Pronunciation Evaluation	51
	Wenda Chen, Ying Ying Tan, Eng Siong Chng, Haizhou Li			Jun Qi, Weiqian Liang, Runsheng Liu, Ruiying Wei	
			32	An Overview of Tibetan Corpus Construction	52
				Yu Hongzhi, Gao Lu, Guo Lei, Kou Jianqun	
			33	Dialogue-Act Analysis with a Conversational Telephone Speech Corpus Recorded in Real Scenarios	52
				Keyan Zhou, Aijun Li, Chengqing Zong	

34	Chinese Language Model Adaptation Using Semi-Supervised Approach	53	45	An HMM-based Hakka Text-to-Speech System	59
	Xinhui Hu, Ryosuke Isotani, Hisashi Kawai, and Satoshi Nakamura			Yi-Ling Tsai, Hsiu-Min Yu, Yih-Ru Wang, Chen-Yu Chiang, Lieh-Shih Lo, Sin-Horng Chen	
35	Multi-channel Speech Data Collection on Mobiles	53	46	Real world utterance collection using voice-enabled web system for child speaker identification	59
	Lin HE, Yufeng HAO, Ke LI, Xianfeng CHENG			Shoko MIYAMORI, Ryuichi NISIMURA, Lisa KURIHARA, Toshio IRINO, Hideki KAWAHARA	
36	The Development of a Large Thai Telephone Speech Corpus: LOTUS-Cell 2.0	54	47	Speech Synthesis Using Epoch Synchronous Overlap Add (ESOLA)	60
	Ananlada Chotimongkol, Nattanun Thatphithakkul, Sumonmas Purodakananda, Chai Wutiwiwatchai, Patcharika Chootrakool, Chatchawarn Hansakunbuntheung, Atiwong Suchato, Panuthat Boonpramuk			Ashoke Kr Datta, Arup Saha	
37	Introduction of SpeechOcean Multi-languages In-Car Project	55	48	Garhwali Speech Database	61
	Lin HE, Yufeng HAO, Xianfeng CHENG, Ke LI			RK Upadhyay, MK Riyal	
38	Building a Cross Script Kashmiri Converter: Issues and Solutions	55	49	Hindi ASR for Travel Domain	61
	Aadil Amin Kak, Nazima Mehdi and Aadil Ahmad Lawaye			Sunita Arora, Babita Saxena, Karunesh Arora, S S Agrawal	
39	The Use of Indonesian Speech Corpora for Developing a Filipino Continuous Speech Recognition System	56	50	Present Scenario of Forensic Speaker Identification in India	62
	Sakriani Sakti, Ryosuke Isotani, Hisashi Kawai, Satoshi Nakamura			Shivani Sharma, S. K. Jain, R. M. Sharma, S.S Agrawal	
40	Utsunomiya University Spoken Dialogue Database for Paralinguistic Information Studies	56	51	Phonetic Segmentation Based on HMM of Hindi Speech	62
	Hiroki MORI, Hideki KASUYA, Makoto NAKAMURA			Archana Balyan, S.S. Agrawal, Amita Dev	
41	An Independent Approach for Spoken Language Analysis	57	52	Development of speech data base for various emotions and their recognition using Neural Network Classifier	63
	Arup Saha, Tulika Basu, Ashoke Kr. Datta			Jyoti Garg, Israr Khan, S.K.Gupta, S.S.Agrawal	
42	Hindi Dialects Phonological Transfer Rules for Verb Root Cələ	57	53	The Development of a Thai Speech Set for Telephony	64
	Diwakar Mishra, Kalika Bali			Therdpong Daengsi, Apiruck Preechayasomboon, Saowanit Sukparungsee, Patcharika Chootrakool, Chai Wutiwiwatchai	
43	Language Identification using Support Vector Machine	58	54	Speaker Recognition Based on Multilingual Speech Features using Neural Network Models	64
	Sanghamitra Mohanty, Basanta Kumar Swain			Sanjay Decate, Anupam Shukla, Sanjay Kumar Singh, Ritu Tiwari	
44	Description of Puma, an endangered language from eastern Nepal	58	55	Orthography development for the standardization of Bhujel: Issues and approaches	65
	Vishnu S Rai			Dan Raj Regmi	

- 56 **Recognition of Hindi Phoneme in Rhyming Words using Vector Quantization** 65
Shweta Sinha, Shalini Goyal, Mona Gaur, S.S Agrawal
- 57 **Building Unit Selection Speech Synthesis in Indian Languages: An Initiative by an Indian Consortium** 66
Hema A. Murthy, Ashwin Bellur, Vinodh Viswanath, Badri Narayanan, Anila Susan, G Kasthuri, K. Sreenivasa Rao, Sudhamay Maity, N. P. Narendra, Ramu Reddy, Krishnendu Ghosh, K. G. Sulochana, E. L. Abhilash, T. Sajini, M. Sasikumar, Bira Chandra Singh, Pranaw Kumar, P. Vijayaditya, E. Veera Raghavendra, Kishore Prahallad
- 58 **NICT Speech and Language Resources and Corpora** 67
Satoshi Nakamura, Kentaro Torisawa, Hisashi Kawai, Eiichiro Sumita
- 59 **Analysis of impression-prosody mapping in communicative speech consisting of multiple lexicons with different impressions** 67
Yoko Greenberg, Hiroaki Kato, Minoru Tsuzaki, Yoshinori Sagisaka
- 60 **Affect Recognition from Acted and Spontaneous Filipino Speech** 68
Jocelynn Cu
- 61 **Intonation Patterns in Nepali Feedback Units** 68
Jens Allwood, Bhim Narayan Regmi
- 62 **Spoken Disfluencies in Multilingual Spoken Corpora** 69
Samudravijaya K
- 63 **Taiwan L2 German Database Design for Computer Assisted Language Learning** 69
Chia-yu Chiu, Yuan-fu Liao, Hansjörg Mixdorff, Hue-San Do, Shing-lung Chen

An ASR System for Spontaneous Urdu Speech

Agha Ali Raza, Sarmad Hussain, Huda Sarfraz, Inam Ullah, Zahid Sarfraz

National University of Computer and Emerging Sciences, Pakistan

{ali.raza, sarmad.hussain, huda.sarfraz, inam.ullah, zahid.sarfraz}@nu.edu.pk

One of the major hurdles in the development of an Automatic Spontaneous Speech Recognition System is the unavailability of large amounts of transcribed spontaneous speech data for training the system. On the other hand transcribed read speech data is available comparatively easily. This paper explores the possibilities of training a spontaneous speech recognition system by using a mixture of read and spontaneous speech data. A single speaker, medium vocabulary spontaneous speech recognition system for Urdu has been developed.

An anti-noise MFCC extraction algorithm for speaker recognition

Wang Hong Pan Jin'gui¹, Wang Hong²

¹*State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China*

²*Institute of Computer Application and Research, Changji University, Changji, China*

whlogs@gmail.com

In order to acquire satisfactory performance of speaker recognition system under noisy environment, an anti-noise Mel-scale frequency cepstrum coefficients (MFCC) extraction algorithm based on the general noisy speech model is proposed. The algorithm uses spectrum mean normalization (SMN) to suppress the additive noise, and uses cepstral mean normalization (CMN) to remove the effect of convolutional noise. Theoretical analyses show that the combination of SMN and CMN can inhibit additive and convolutional noise at the same time. To verify the performance of the new algorithm, we have conducted some speaker recognition tests by using this algorithm and the conventional MFCC approach, respectively. The additive white noise experiments and the additive factory noise experiments with the same convolutional noise component show that the proposed algorithm provides 10.5% and 9.6% relative improvement than the conventional MFCC approach, respectively.

Syntactic and Semantic analysis of Bangla language for developing grammar checker system

Bibekananda Kundu

CDAC Kolkata

bibekananda.kundu@cdackolkata.in

This paper describe about types of error may occur in the time of writing a text and methodology of development of grammar checker to rectify such error and alert the user of every error it detects. In this paper the grammatical and ungrammatical error are described on the basis of Bangla Language.

Speech Corpus Development for a Speaker Independent Spontaneous Urdu Speech Recognition System

Huda Sarfraz*, Sarmad Hussain*, Riffat Bokhari**, Agha Ali Raza**, Inam Ullah*, Zahid Sarfraz**, Sophia Pervez**, Asad Mustafa*, Iqra Javed*, Rahila Parveen*

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

**firstname.lastname@kics.edu.pk.*

***Center for Research in Urdu Language Processing, National University of Computer and Emerging Sciences, Lahore, Pakistan. **firstname.lastname@nu.edu.pk*

This paper reports the design and development of an 82 speaker Urdu speech corpus for speaker independent spontaneous speech recognition using the CMU Sphinx Open Source Toolkit for Speech Recognition. The corpus consists of 45 hours of spontaneous and read speech data from 82 speakers (42 male and 40 female), recorded over a microphone and a telephone line. The speech was collected from speakers ranging from 20 to 55 years of age. Recording sessions were conducted in office and home environments.

Valency Analyzer of Verb Arguments for Bangla

Subhash Chandra, Pampa Bhattacharyya

NLP Group, Centre for Development of Advanced Computing (CDAC), Kolkata

subhash.chandra@cdackolkata.in

pampa.bhattacharyya@cdackolkata.in

This paper presents a methodology for analysis of the requirements of Bangla verbs for its arguments. Literary it can be called the searching for expectation of Bangla verbs and compatibility of Bangla nouns. There are two phases in the total study. First one is to build up a method for generating the expectation of verbs and compatibility of nouns for Bangla. Bangla ontological tree and verb expectation system are the outcome of this method. Second is to establish the relationship between verb and noun. As a result an analyzer will be generated for valency checking. After completion of this work system can distinguish between different types of noun entities. The work is under development.

Phonation Type of Korean Stops -Research Based On Data Retrieved From Unified Acoustic Parameter Database

Zhou Xuewen, Zheng Yuling, Chuai Zhenyu

*Institute of Ethnology & Anthropology, Chinese Academy of Social Sciences
zhouxw@cass.org.cn, zhyl-cass@163.com, chuaizy@cass.org.cn*

Based on parameters retrieved from *Unified Acoustic Parameter Database Platform* (developed by The Phonetic Lab, Institute of Ethnology & Anthropology, Chinese Academy of Social Sciences), this paper examined Korean (spoken in China) three-way contrast stops, called lenis, aspirated and fortis stops. Through research on acoustic parameters of following vowel including intensity of harmonics, voice quality, pitch onset & intensity as well as acoustic characteristics in sound-wave and spectrogram and referring to other scholars' research results of Korean (spoken in Korea) three-way contrast stops, we conclude that Korean three-way contrast stops differ in phonation types and the articulation mechanism works in whole syllable. Lenis stop belongs to slack voice and fortis stop belongs to modal voice.

Performance evaluation of word boundary detection for Hindi speech database

Anurag Jain¹, S.S. Agrawal², Nupur Prakash¹

1 GGSIP University, Delhi India, 2 CDAC Noida India

anuragjain76@gmail.com, ss_agrawal@hotmail.com,

nupurprakash@rediffmail.com

Word boundary detection (WBD) is very common and important problem in the field of speech synthesis and recognition. Several studies have made relation between syntactic structure and prosodic features. Several researches are open on this field, since there is no sign of start of the word, end of the word and number of words in the spoken utterance of any natural language, so study of intonation pattern for a particular language database is essential. The pitch contour of a sentence indicates the sentence type and its expressive style. Similarly pitch variation also indicates the boundaries between major syntactic units in a sentence. In this paper a word boundary detection algorithm is proposed for speech signal for Hindi speech database. A careful study of the intonation pattern of Hindi language has been done. Based on the study it is observed that, there are several suprasegmental parameters of speech signal such as pitch, fundamental frequency, duration, intensity, and pause, which may play an important role in finding some cues to detect the start and the end of the word from the spoken utterance of Hindi Language. Several studies have been done to detect word boundary along with the syllable boundary, as there are sufficient information related to specific language, which makes a syllable as a word based on pause and duration. For example studies say that for Hindi very few words end in a short vowel and vowel occurs twice as often as consonants before a word boundary, But the proposed algorithm tries to find the word boundary based on merely two prosodic parameters, pitch and intensity, which makes it a simplest and efficient algorithm to find the word boundary as compared with existing WBD techniques. It can also be claimed that the proposed algorithm can be implemented for any language as it is not based on syntactic structure of any language. This algorithm also provides the word duration that makes a signification role in emotion transformation. To perform the experiment 15 speakers are selected and each speaker has been given 25 sentences. These sentences are recorded in different emotions by native Hindi speakers under the normal lab condition with 44.1 KHz sampling rate and 16 bit precision with mono channel and stored in a Intonation rich speech database.

It is found that proposed word boundary detection algorithm has recognition accuracy is of 90.8% and 84.4% for actual word boundary and actual no-word boundary respectively. It is worthwhile to mention here that the proposed algorithm detects very few syllable boundaries as compared to other techniques and gives good idea about the number of words present in an utterance.

An Analysis of a Mandarin-English Code-switching Speech Corpus: SEAME

Dau-Cheng Lyu^{1,4}, Tien-Ping Tan², Eng-Siong Chng^{1,4}, and Haizhou Li^{1,3,4}

1 School of Computer Engineering, Nanyang Technological University, Singapore 639798

2 School of Computer Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia

3 Institute for Infocomm Research, 1 Fusionopolis Way, Singapore 138632

*4 Temasek Laboratories, Nanyang Technological University, Singapore 639798
dclyu@ntu.edu.sg, tienping@cs.usm.my, aseschng@ntu.edu.sg, hli@i2r.a-star.edu.sg*

SEAME (South East Asia Mandarin-English) is a 30 hours spontaneous Mandarin-English code-switching speech corpus recorded from Singapore and Malaysia speakers. In this paper, we report a series of analyses on the recording, processing time and voice activity rate (VAR) of the speech recording, transcription, validation and language boundaries labeling processes. In addition, the duration of the monolingual segment in the code-switching utterance and the analysis of the speakers' behavior in language switching during conversation are also described. The results of the analysis show that 80% and 72% monolingual segments of English and Mandarin in the code-switching utterance are shorter than one second. In over 80% of the cases, speakers directly switch language without any short pause and discourse particle between two adjacent different languages.

Sanskrit Verb Argument Valence: A Computational Analysis

Subhash Chandra, Dr. Girish Nath Jha

Special Centre for Sanskrit Studies, Jawaharlal Nehru University, New Delhi

subhash.jnu@gmail.com, girishj@mail.jnu.ac.in

In this paper authors present a methodology to develop a "Sanskrit Verb Argument Analysis System (SVAAS)" for ascertaining verb's arguments and their semantic compatibility with the verb. The work involves two major goals - to develop a knowledgebase for verb-

expectancy and to map semantic compatibility (or logic of it in the real world) of the arguments with the verbs. The object language taken for this work is Sanskrit. The data entry is being done with the help of an interface consisting of Sanskrit ontological tree. After the completion of this work, the system will be able to distinguish between different types of entities like human, animal, animate and inanimate etc.

A Metric-based Phone Segmentation Method using Wavelet Transform

Ching-Feng Lu, Hsiao-Chuan Wang

Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan 300-13

g9761589@oz.nthu.edu.tw ; hcwang@ee.nthu.edu.tw

This paper proposes a metric-based algorithm for the text-independent phone segmentation without prior knowledge about the phonetic information. It detects phone boundaries in a sequential manner. By setting a starting point, the discrete wavelet transform is applied to variable length frames in order to search for a candidate phone boundary. Once a candidate phone boundary is detected, the verification process follows to confirm the existence of a phone boundary. This verification process is accomplished by detecting the spectral change based on a model selection criterion and the normalized spectral variation function (SVF). If the phone boundary is confirmed, this location is considered the new starting point for finding next phone boundary. The proposed method was evaluated on TIMIT corpus, and the performance was measured in F1-score and R-score. The average F1-score and Rscore of 640 test utterances are 72.4% and 75.1% respectively with 20 ms tolerance.

Development of a Malay LVCSR System

Xiong Xiao¹, Eng Siong Chng¹, Tien-Ping Tan², Haizhou Li^{3,1}

¹*School of Computer Engineering, Nanyang Technological University, Singapore 639798*

²*School of Computer Science, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia*

³*Institute for Infocomm Research, Singapore 138632*
{xiaoxiong,aseschnj}@ntu.edu.sg, tienping@cs.usm.my, hli@i2r.a-star.edu.sg

Recently, we have collected a Malay read speech corpus and news text corpus. In this paper, we develop a large vocabulary continuous speech recognition (LVCSR) system based on this corpus. To the best of our knowledge, there is very few study on Malay LVCSR.

In this paper, we discuss one aspect of a Malay LVCSR system, i.e. the amount of training data required to train robust acoustic and language models for the Malay language. This results may be of interest to reader for future study in Malay LVCSR system development.

IPTV / Navigation Environmental Speech DB and Usability Test of IPTV VOD Retrieval

Dae-Lim Choi¹, Bong-Wan Kim¹, Yong-Ju Lee², Byung-Ok Kang³, Eui-Sok Chung³, Yun-Keun Lee³, Gyu-Tae Baek⁴, Ki-Hyung Hong⁵

¹*Speech Information Technology & Industry Promotion Center, Wonkwang Univ.*

²*Department of Computer Engineering, Wonkwang Univ.*

³*Electronics and Telecommunications Research Institute*

⁴*Future Technology Laboratory, KT*

⁵*School of Media and Information, Sungshin W. Univ.*

{dlchoi, tacanemo, yjlee}@wku.ac.kr, {bokang, eschung, yklee@etri.re.kr}, baegt@kt.com, khbhong@sungshin.ac.kr

With the expansion of digital TV, IPTV (Internet Protocol Television) and cable networks which are becoming increasingly popular in everyday life, the number of channels, programs and VOD (Video On Demand) services is increasing geometrically. The number of POIs (Points Of Interest) for navigation system may amount to millions. It is very inconvenient to use touch-screen control or typing to search from this huge amount of data, so it is attempted to use voice to search for needed information with more ease. In this paper, we introduce the speech DBs for voice search technology in IPTV STB (Set Top Box) and in navigation system. In addition, we describe the test results for speech recognition rate and usability of speech-based multimodal user interface for IPTV VOD retrieval.

A Chunk Level Statistical Machine Translation

(An Approach for English Language to Nepali Language Translation)

Ashim Ghishing, Bikash Balami, Yoga Raj Joshi

CDCSIT, TU

ashimji@gmail.com, bikuji@gmail.com, joshi.yogaraj@gmail.com

Machine Translation (MT) is a task of translating from one language to another by the use of computer. The peculiarities and morphological structures' differences among languages create

ambiguity and make MT more challenging. This paper is mainly concentrated on Chunk Level Statistical Machine Translation (SMT) rather than the traditional rule-based translation. SMT acquires knowledge that is required for the statistical translation by training. This training is conducted over the bilingual corpus. The knowledge, which is typically in the form of probabilities of various language features, is used to guide the translation process. The paper overviews an SMT technique which is implemented for English to Nepali translation and discusses some issues related with the translation ambiguities such as **gender ambiguities, dropping words, unknown words** etc.

Applying Pitch Based Dynamic Pruning in Designing Real-Time Speaker Identification System

Centre for Development of Advanced Computing

Soma Khan, Joyanta Basu, Shyamal Kumar Das Mandal

soma.khan@cdackolkata.in, joyanta.basu@cdackolkata.in, shyamal.dasmandal@cdackolkata.in

In real-time Speaker Identification (SI) systems, matching or calculating likelihood distance, demands efficient and fast computation techniques. A huge amount of identification time and complexity can be reduced by pruning some of the unlikely speakers before matching. Present paper introduces Pitch Based Dynamic Pruning (PBDP) technique regarding optimization of Vector Quantization (VQ) based Speaker Identification process. The system is being trained and tested by the voice samples of 50 speakers across different age groups. After detecting the Frequent Voicing Activity zones (FVAZ) in the test speech, pruning criteria is set by calculating the Voicing Occurrence Frequency Rate (VOFR) and based on these values, some of the speakers are selected for matching. Matching scores are calculated only for those survived speakers. Overall System performance is shown over varying codebook sizes, with and without using PBDP. PBDP result shows a reduction of 42.56% in total identification time with the accuracy of 97.17%.

Discourse Prosody Planning in L1 and L2 English

Tanya Visceglia¹, Chiu-yu Tseng², Zhao-yu Su² and Chi-Feng Huang²

¹ *Department of Applied English, Ming Chuan University, Taipei, Taiwan*

² *Phonetics Lab, Institute of Linguistics, Academia Sinica, Taipei, Taiwan
orlandotaipei@hotmail.com, cytling,morison,chifeng@sinica.edu.tw,*

L1 English and L1 Taiwan Mandarin discourse length English speech data extracted from the TWNAESOP corpus was analyzed using a perceptually-based hierarchy of prosodic phrase group (HPG) framework in order to investigate similarities and differences in the organization of discourse-level speech planning in English across L1 (native) and L2 (non-native) speaker groups. While both groups appear to produce similar configurations of acoustic contrasts to signal discourse units and boundaries, L1 speakers were found to produce these cues more robustly. Between-group differences in discourse units were also found through the distribution of prosodic break levels and break locations. These findings can be attributed to the size and scope of speech planning and chunking, whereby L2 speakers, possibly due to on-line processing limitations in L2, use more intermediate chunking units and fewer larger-scale planning units in prosodic discourse organization. Future cross-L1 comparisons will investigate whether these differences represent L2-universal processing limitations and strategies.

A Proposal for Standardizing Catalogue Specifications of Speech Corpora

S. Itahashi^{*}, K. Yamakawa^{**}, T. Matsui^{***}, Y. Ishimoto^{*}

^{*} *National Institute of Informatics, Tokyo, Japan*

^{**} *Aichi Shukutoku University, Nagoya, Japan*

^{***} *The Institute of Statistical Mathematics, Tokyo, Japan*

^{*} *National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan*

itabashi@mii.ac.jp

Speech corpora are indispensable to speech research. There are several data centers in the world that serve as repositories for various speech corpora. However, they use different specification items for their corpora, and so it is difficult to compare their corpora. It would be more convenient for corpus users if the data centers were to use a common set of specification items for describing speech corpora. We propose a tentative list for standardizing such specification items. As an example of utilizing them, we describe a visualization system that shows similarities among various speech corpora by incorporating corpus specification items based on the multidimensional scaling (MDS) method.

Method for Collection of Diverse Speech for Emotion

Research Database

Takahiro Miyajima¹, Takeshi Fukuda², Hideaki Kikuchi², Katsuhiko Shirai³

¹*Information Technology Research Organization, Waseda University, Tokyo, Japan,*

²*Faculty of Human Sciences, Waseda University, Saitama, Japan*

³*Faculty of Science and Engineering, Waseda University, Tokyo, Japan*
 miyajima@toki.waseda.jp, tksh_fkd@moegi.waseda.jp, kikuchi@waseda.jp,
 shirai@waseda.jp

Techniques for the processing of information regarding human emotions have been improved by the steady development of computer science and the efforts of specialists. Problems with the results achieved, however, are often pointed out. For example, only quite limited material (such as several types of emotions) has been verified in recent studies. In the research area of speech and emotion, an ideal approach to this problem would be to prepare spontaneous and diverse data collected from a large sampling of actual speech. Such an approach, however, requires an inordinate amount of time, and detailed control of the obtained data is rarely possible. In our approach, which focuses on the efficiency and facility of control, we address the possibility of constructing a corpus of data using a professional voice actor and an acting script extracted from TV programs to produce a corpus with diverse acoustical/psychological features.

Feasibility of the Characterisation Control by Text-based Speech Style

Raymond SHEN¹, KIKUCHI Hideaki¹, OHTA Katsumi², MITAMURA Takeshi²

¹ *Faculty of Human Sciences, Waseda University*

² *Mobility Services Laboratory, Nissan Research Center, Nissan Motor Co. Ltd.*

In this study, we discussed the feasibility of representation and control of characterisation by speech style (narrative, rhetoric, pragmatics, etc.) in text levels. Through three steps of experiment and work, we also attempted to confirm the effect of text-based speaker characterisation.

First, we obtained some impression ratings of spontaneous speech and selected some highly-rated speech transcripts. Then, from these transcripts we extracted and categorized several patterns of expressions and rhetoric, which were found to be effective in impression formation. Finally, by applying these patterns into plain texts, we generated several texts and carried out an evaluation experiment to verify the effectiveness of the patterns on impression formation.

As a result, we found that the extracted linguistic patterns are very effective in impression formation, which may also prove the feasibility of our approach. In the conclusion part, based on the review of the whole process, we also pointed out what shall be done for the next step of this study.

An Approach to Mixed Language Automatic Speech Recognition

Kiran Kumar Bhuvanagiri, Sunil Kumar Kopparapu

TCS Innovation Labs – Mumbai, Yantra Park, Pokharan Road 2, Thane(West), Maharashtra, INDIA

{KiranKumar.Bhuvanagiri,SunilKumar.Kopparapu}@TCS.Com

Use of mixed language in day to day spoken speech is becoming common and is being accepted as being syntactically correct. However recognition of mixed language spoken speech is a challenge to a speech recognition engine. Though sparse, there have been studies on how to enable recognition of mixed language spoken speech. At one extreme is to use acoustic models of the complete phone set of the mixed language to enable recognition while on the other extreme is to use a language identification module followed by a language dependent speech recognition engine to recognize mixed language. Each of this has its own implications. In this paper, we approach the problem of mixed language recognition by constraining ourselves to use readily available resources and show that by (a) suitably modifying the language model to use mixed language and (b) by constructing a pronunciation dictionary, one can achieve a good recognition of mixed language spoken speech.

Cases on Extension of Language Technology into the Related Fields and their Implications for Research and Development in University

Satoru HAYAMIZU, Tadahiro MATSUMOTO, Satoshi TAMURA, Shinichi TAKEUCHI

Gifu University

{hayamizu@, tad@info., tamura@info., takeuchi@asr.info.} gifu-u.ac.jp

This paper presents some cases on extension of language technology into the related application fields. First, we summarize the efforts related with our university. We describe five activities; (1) caption-generation from speech; (2) audio-visual speech recognition and related works; (3) machine translation system for Asian Languages; (4) Ibuki-ten; (5) Hand sign language description system. Based on our experience of research and development on related fields with language technology, we discuss roles of university for extension of the basic core technology into related applications. Those roles could be source of future diffusion of language technology to more languages with different modalities.

The Development of a Singapore English CALL Resource

Wenda Chen¹, Ying Ying Tan², Eng Siong Chng¹, Haizhou Li^{3,1}

¹*School of Computer Engineering, Nanyang Technological University, Singapore 639798*

²*School of Humanities and Social Sciences, Nanyang Technological University, Singapore 639798*

³*Institute for Infocomm Research, Singapore 119613*
wdchen@ntu.edu.sg, yytan@ntu.edu.sg, aseschn@ntu.edu.sg, hli@i2r.a-star.edu.sg

This paper describes the development of an English Computer Assisted Language Learning (CALL) resource for the Singapore environment in Nanyang Technological University, Singapore. Specifically, this paper describes the development of a Singapore English-based pronunciation lexicon, the collection of a 44-speaker and a 39-speaker audio corpus for pronunciation scoring and acoustic model training respectively, and the development of a software for manual prosodic scoring. These resources can be used for the development of a pronunciation scoring system for Singapore users.

Mental-State Analysis for Understanding Children's Behavior Based on Emotion-Label Sequences in Multimodal Speech-Behavior Corpus

Shinya Kiriyaama, Shogo Ishikawa, Shigeyoshi Kitazawa, and Yoichi Takebayashi

Faculty of Informatics, Shizuoka University, Japan

Graduate School of Science and Technology, Shizuoka University, Japan
kiriyaama@inf.shizuoka.ac.jp

This paper describes a methodology for understanding the internal mental state of children by using a multimodal speech-behavior corpus. Emotion is a mental state that determines how problems are solved, and speech is the most important clue for observing the emotional state. The objective of the study is to construct common-sense thinking models for developing advanced speech-understanding frameworks. The target domain for constructing the corpus is the children's learning environment, because children are more naive than adults in terms of exhibiting their internal mental state in their behavior. A multimodal behavior-recording environment consisting of multi-angle cameras and wearable microphones produced video data that captured the children's behavior in detail, as well as high-quality speech data. This paper also presents a description method and an annotation tool for emotional behavior analysis. The results of mental-state analysis based on an emotion-label sequence proved that the changes in emotion labels were valuable in considering children's problem-solving processes, and that the corpus was a rich source of scenes that are useful for observing children's communicative skills and those developmental processes that deal with fundamental common-sense thinking in human social interaction. The results also indicated that the examples extracted from the analysis were effective in creating valuable content for child-care support.

A Hybrid Speech Enhancement System Based On Wavelet Denoising

Richa Tyagi and Sunita Maithani

Scientific Analysis Group, Defense Research & Development Organization
Metcalfe House, Delhi-110054 India
richa.drdo@yahoo.in, ysmaithani58@yahoo.com

The paper presents a generic speech enhancement system which is mainly combination of a variant of spectral subtraction and wavelet

denoising techniques, optimized to deal with any kind of real world noisy speech. The enhancement system is preceded by a high pass filter to remove hum noise and followed by a noise gate for noise masking at non speech regions. The developed hybrid technique is tested for performance evaluation in terms of improvement in estimated SNR (Signal to Noise Ratio) and other objective quality measures for enhanced speech on different types of noises and noise levels in noisy speech. The technique performs efficiently in stationary & non-stationary kind of noises as well as their combinations. The developed system is also tested for quality improvement on real HF/VHF channel field kind of noisy speech with very low SNR and highly non stationary noises.

Age Group Estimation based on Acoustic Analysis of Speech

Devendra Kumar Yadav, Kamini Malhotra, Anu Khosla

Scientific Analysis Group, DRDO, Metcalfe House, Delhi – 110054
kum_dev@yahoo.co.uk, kaminiimal@yahoo.com, khoslaanu@yahoo.co.in

In this paper an approach for identification of the age group of a person is presented. The recordings were taken from members of two different age groups – below 30 years and above 50 years. Two approaches based on Gaussian Mixture Models (GMM) and Probabilistic Neural Networks (PNN) were tried. Only spectral features have been incorporated in the form of Mel Frequency Cepstral Coefficients (MFCC). The results indicate that the PNN based technique is able to capture the age related differences in the spectrum in a better fashion as compared to GMM's. Though the younger age group has higher correct identification score (96.5%) with GMM's but the misclassifications are very high for the elder age group. PNN gives a better overall classification with the elder age group giving 95% correct classification and younger age group being classified 76% correctly.

Aerodynamic Study of Standard Mongolian Tense-lax Vowels

Hu Axu, Gegentana & Yu Hongzhi

Key Lab of Ethnic Languages and Information Technology of China
Northwest University for Nationalities, Lanzhou, China
huaxu84@foxmail.com

Based on experimental methods, this article described the aerodynamic characteristics of the standard Mongolian tense-lax

vowels, and discovered that the tense-lax change of vowels and the tongue movement have an effect on 6 aerodynamic parameters such as the airflow rate, glottal resistance, phonation efficiency etc. by analyzing the relationship between tense-lax shift and dorsal movement, the article proved that the standard Mongolian tense-lax characteristics is mutually to the variation of dorsal position.

Creation of Time-Varying Voiceprint Database

Linlin Wang and Thomas Fang Zheng

Center for Speech and Language Technologies, Division of Technical Innovation and Development, Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing, 100084, China
wangll07@mails.tsinghua.edu.cn, fzheng@tsinghua.edu.cn

Performance degradation with time varying in speaker recognition is a generally acknowledged phenomenon and it is widely assumed that speaker models should be updated from time to time to maintain representativeness. However, lack of a longitudinal voiceprint database which specially focuses on the time-varying effect has prevented researchers from finding out reasons behind this phenomenon. In this paper, an on-going voiceprint database creation project is presented, aiming to examine solely the time-varying impact on speaker recognition and explore hidden factors causing possible performance degradation. In this speech database, speakers are requested to utter in a reading way with fixed prompt texts instead of free-style conversations throughout 16 sessions in a period of about three years. Sessions are of gradient time intervals where initial ones are of shorter time intervals and following ones of longer and longer time intervals. Initial experimental results on the partially completed database are also presented, which demonstrated time-varying effect evidently.

Mora Pitch Level Recognition for the Development of a Japanese Pitch Accent Acquisition System

Greg Short¹, Keikichi Hirose¹, Takeshi Yamada², Nobuaki Minematsu¹, Nobuhiko Kitawaki², Shoji Makino²

¹University of Tokyo, ²University of Tsukuba
[short,hirose,mine]@gavo.t.u-tokyo.ac.jp, [takeshi,kitawaki]@cs.tsukuba.ac.jp, maki@tara.tsukuba.ac.jp

A language learning system built for guiding a student on how to pronounce words in a second language must provide meaningful

feedback while locating the learner's errors with a high accuracy. We propose a method to detect pitch accent errors in the speech of learners of Japanese. In previous methods, Tokyo accent type recognition of the learner utterance was focused on. However, the learner may produce pitch level patterns that do not exist in these accent types. This paper covers a proposed technique for identifying mora pitch level to recognize all patterns. Employing a 2 mora model trained on the continuous speech corpus JNAS, this method identifies pitch level for contiguous two mora unit sets. Then it determines the most likely combination of these units to find the pitch level for each mora of the word. Through this method, we achieved an 80.5% correct mora pitch level identification rate.

Speech Technology and Empathy in Conversational Interaction

Nick Campbell

Centre for Language & Communication Studies, Trinity College Dublin, Ireland
nick@tcd.ie

This paper describes recent work on the development of technology for processing spoken interaction. It extends speech recognition and synthesis by incorporating a multimodal element which serves as the eyes and ears of the speaker to facilitate more interactive speech and to better model the two-way process of normal conversational interaction. A prototype robot is described that incorporates a set of sensors for conversation tracking and that thereby facilitates the testing of robotic participation in a human conversation using empathic sensing.

Analysis and Synthesis of F_0 Contours for Bangla Readout Speech

Shyamal Das Mandal¹, Anal Haque Warsi¹, Tulika Basu¹, Keikichi Hirose², Hiroya Fujisaki²

¹CDAC, Kolkata, ²University of Tokyo
{shyamal.dasmandal, anal.warsi, tulika.basu}@cdackolkata.in, hirose@gavo.t.u-tokyo.ac.jp, fujisaki@alum.mit.edu

It is well known that the F_0 contour plays an important role in conveying prosodic information, but the process of synthesizing the F_0 contour from the underlying linguistic information has not been elucidated for Bangla. This paper defines the prosodic units

of Bangla on the basis of F_0 contour analysis using the command-response model by Fujisaki et al. For the study, 200 Bangla declarative sentences spoken in the readout mode are analyzed. Based on the analysis, rules are constructed for predicting both phrase and accent command parameters of the model for generating the F_0 contours of Bangla readout speech. A perceptual evaluation test for the naturalness of prosody shows that there is no significant difference between synthetic speech with model-generated F_0 contours and the original natural speech.

Creation and Analysis of a Japanese Speaking Style Parallel Database for Expressive Speech Synthesis

Hideharu Nakajima, Noboru Miyazaki, Akihiro Yoshida, Takashi Nakamura, Hideyuki Mizuno

NTT Cyber Space Laboratories, NTT Corporation

This paper describes a newly developed database for expressive speech synthesis. This speech database is characterized by two features: i) the sentences are taken from real domains such as sales talk, storytelling, and telephone conversation, where speech is uttered in expressive (or conversational) style, hence, sentences are domain-dependent, ii) each sentence is uttered in both reading style and expressive style, hence this database stores parallel speaking style speech. This database is designed to capture the acoustic and prosodic differences between parallel styles and to elucidate the domain-dependent linguistic characteristics that cause those differences. This paper describes both the concept of this speech database and the issues raised by its implementation. We detail the basic characteristics and preliminary results of two style comparisons to elucidate the linguistic characteristics that contribute to establishing expressive speech synthesis.

Application of Discriminative Training Technique for English Pronunciation Evaluation

Jun Qi², Weiqian Liang¹, Runsheng Liu¹, Ruiying Wei²

¹ Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

² Institute of Microelectronics of Tsinghua University, Beijing 100084, China

¹ *lwq@tsinghua.edu.cn* ² *qij08@mails.tsinghua.edu.cn*

This paper introduces our project of English pronunciation system with acoustic model trained by discriminative training techniques. In the project, a tester's evaluation result is given based on objective machine score of

Goodness of Pronunciation (GOP). The core of getting a more reasonable GOP score depends on better training acoustic model which is viewed as a kind of reference of standard pronunciation. As proven in the field of speech recognition, although the acoustic model trained in ML-sense can best align all the training data, the general ML-sense acoustic model cannot settle discriminative problem among several phones such that the evaluation system cannot give enough discriminative and correct scores for those phones as well. However, the subjective expert tends to be easy to find the difference. Since the correlation between objective machine score and subjective expert grade can be used as a kind of approach to access whether the machine score is correct enough, mismatch between testers' spoken speech and evaluation acoustic model can be seen as the most significant problem. In this paper, the technique of discriminative training can be introduced to eliminate the existing problem. The corresponding experiments show that the state of art discriminative training methods MPE and BMMI can bring about absolute at least 2% rise of relativity in any case.

An Overview of Tibetan Corpus Construction

Yu Hongzhi, Gao Lu, Guo Lei, Kou Jianqun

Key Lab of China's National Linguistic Information Technology, Lanzhou, China

lzguolei@126.com, lyhweiwei@126.com

The Tibetan resource construction is the foundation for the development of Tibetan information technology. After 10 years accumulation, China National Information Technology Research Institute has made remarkable achievements in the construction of Tibetan corpus, including the construction of speech corpus, the text corpus and the video library three parts, the establishment of these corpus are the foundation for the research of the Tibetan information, machine translation, speech synthesis, Tibetan intelligent information retrieval and other aspects.

Dialogue-Act Analysis with a Conversational Telephone Speech Corpus Recorded in Real Scenarios

Keyan Zhou¹, Aijun Li², Chengqing Zong¹

¹ *NLPR, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190*

² *Institute of Linguistics, Chinese Academy of Social Sciences, Beijing, 100732*

¹ *{kyzhou, cqzong}@nlpr.ia.ac.cn, ²liaj@cass.org.cn*

CASIA-CASSIL is a large-scale corpus of Chinese spontaneous telephone conversations in tourism domain underdevelopment. This

paper gives some statistics of linguistic characteristics based on the Dialogue-Act (DA) annotation in CASIA-CASSIL. Distributions of DA are presented and compared in different domains. And also, we describe and discuss two kinds of Question sentences in detail, which are Yes-or-No question and Wh-Question. In Yes-or-No question sentences, a large part of them can be called intonational question realized by intonation cues rather than any question markers. We believe the details on linguistic and paralinguistic information will help to study the prosodic analysis pertinent to DAs.

Chinese Language Model Adaptation Using Semi-Supervised Approach

Xinhui Hu, Ryosuke Isotani, Hisashi Kawai, and Satoshi Nakamura

National Institute of Information and Communications Technology, Kyoto, Japan

{xinhui.hu, ryosuke.isotani, hisashi.kawai, satoshi.nakamura}@nict.go.jp

To adapt a Chinese language model for automatic speech recognition (ASR) within a speech-to-speech translation (S2ST) system in a fixed location, we use the semi-supervised approach in which the ASR's output is used as adaptation data. The semi-supervised approach is realized by correcting the content words of the recognized results, and building an adaptation model using the corrected transcripts. Taking a test set that is open to the adaptation data but originates from the same field experiment, we evaluated two adaptation models – a linear interpolation of n-grams and a unigram marginal adaptation model, and verified the effective method for our target. By correcting about half of the misrecognized words, the performance (Chinese character recognition accuracy) is improved 7.43% when compared with the nonadaptation language model and becomes close to the supervised method. It is shown that by using very few real data and having partial corrections on the adaptation data and using simple adaptation method, the recognition performance can be improved at a large scale.

Multi-channel Speech Data Collection on Mobiles

Lin HE¹, Yufeng HAO², Ke LI², Xianfeng CHENG²

¹ *Institute of Acoustic, China Academy of Science, Beijing, China, 100080*

² *Beijing HaiTian RuiSheng Technology & Science Co., Ltd, Beijing, China 100080*

linhe@vip.sina.com, {haoyufeng, chengxianfeng, like}@speechocean.com

In recent years, along with the increasing popularity of embedded

electronic equipment, speech recognition technology also get more and more widely employed, especially the command & control speech recognition and voice phone search on mobile. Consequently, the data collection on embedded devices is becoming more and more urgent for speech technical R&D. In this paper, we build an embedded audio recording system (EARec) for mobile phones under WinMo6.1 platform. This software (EARec) collects speech data through internal mobile microphone and Bluetooth channel simultaneously, records the environment easily, supports a variety of ways to recording style, and obtains the high quality speech data with the voice quality control module. Till now, EARec has been employed in 4-language 2-channel speech data collection project on mobile and achieve excellent performance.

The Development of a Large Thai Telephone Speech Corpus: LOTUS-Cell 2.0

Ananlada Chotimongkol¹, Nattanun Thatphithakkul¹, Sumonmas Purodakananda¹, Chai Wutiwiwatchai¹, Patcharika Chootrakool¹, Chatchawarn Hansakunbuntheung¹, Atiwong Suchato², Panuthat Boonpramuk³

¹*National Electronics and Computer Technology Center (NECTEC)
112 Phahonyothin Road, Klong Nueng, Klong Luang, Pathumthani, 12120, Thailand*

²*Department of Computer Engineering, Faculty of Engineering,
Chulalongkorn University, 254 Phayathai Road, Pathumwan, Bangkok,
10330, Thailand*

³*Department of Control System and Instrumentation Engineering, King
Mongkut's University of Technology Thonburi, 126 Prachautid Road,
Bangmod, Thrungkru, Bangkok, 10140, Thailand
{ananlada.cho, nattanun.tha, sumonmas.pur, chai.wut, patcharika.cho,
chatchawarn.han}@nectec.or.th, atiwong@cp.eng.chula.ac.th, panuthat.boo@
kmutt.ac.th*

This paper describes the design and construction of the LOTUS-Cell 2.0 corpus, a large Thai telephone speech corpus. The corpus contains 3 parts of speech data, answers to closed-ended questions, answers to open-ended questions and dialog speech. The questions and discussion topics were designed to elicit speech data which have their contents conform to the domains of potential automatic speech recognition (ASR) systems over a telephone channel. This corpus also includes detailed annotations of interesting spoken

language characteristics e.g. pronunciation variations, incorrect pronunciations and false starts. Up until now, we have recorded 90 hours of speech from 212 speakers with 12 hours fully transcribed and annotated. From the analysis of the transcribed speech, we found that telephone speech contains both formal and informal speaking styles and quite a number of incorrect pronunciations. These variations should be taken into account when developing ASR systems for telephone speech.

Introduction of SpeechOcean Multi-languages In-Car Project

Lin HE¹, Yufeng HAO², Xianfeng CHENG², Ke LI²

¹*Institute of Acoustic, China Academy of Science, Beijing, China, 100080*

²*Beijing HaiTian RuiSheng Technology & Science Co., Ltd, Beijing, China
100080*

linhe@vip.sina.com, {haoyufeng, chengxianfeng, like}@speechocean.com

In order to better serve the rapid development of speech recognition market, we give a detail introduction of SpeechOcean Multi-language In-Car Speech Data Collection Project in this paper. To make better performance and more fit the reality, we construct multi-channel recording hardware / software system with friendly UI, design prompts which fitting application tightly, find speakers with good distribution in gender, age, dialect and noise environment. All speech sentences are carefully transcribed and annotated under the rules. Till now, there are 4 languages in-car speech data available with minimal 300 speakers for each corpus.

Building a Cross Script Kashmiri Converter: Issues and Solutions

Aadil Amin Kak, Nazima Mehdi and Aadil Ahmad Lawaye

University of Kashmir

aadilaminakak@yahoo.com; nazimamehdi@yahoo.com; aadillawaye@yahoo.com

Kashmiri is a new entrant in the realm of Natural Language Processing. Efforts in this direction are only now taking place by developing different NLP tools. The paper in question talks about the development of a Persio-Arabic Devanagari converter. Here the main focus is on handling some issues which were faced while developing the converter.

The Use of Indonesian Speech Corpora for Developing a Filipino Continuous Speech Recognition System

Sakriani Sakti, Ryosuke Isotani, Hisashi Kawai, Satoshi Nakamura

NICT Spoken Language Communication Research Group

3-5 Hikaridai, "Keihanna Science City", Kyoto 619-0289, Japan

{sakriani.sakti, ryosuke.isotani, hisashi.kawai, satoshi.nakamura}@nict.go.jp

The development of an automatic speech recognition system for a new language requires collection of a huge amount of speech data, as well as manual annotation and transcription. That is why the feasibility of cross-language transfer of speech technology has become a matter of increasing concern as the demand for recognition systems in multiple languages grows. This paper shows the possibility of developing a Filipino continuous speech recognition system by using Indonesian speech data. It is based on the cross-language approach with following procedure: (1) Normalize the phoneme sets of Indonesian and Filipino, in order to have the same phonetic transcription convention across Indonesian and Filipino language. (2) Train the Indonesian speech corpora with normalized phoneme set and apply it as an initial acoustic model of the Filipino language. (3) Use the initial Filipino acoustic model to segment the limited utterances of Filipino training data by the Viterbi alignment algorithm. (4) Retrain and adapt the parameters of the initial acoustic model using the Filipino training data. Experimental results reveal that even with the initial acoustic model (trained on pure Indonesian speech data), the system could recognize Filipino continuous speech up to 79.50% word accuracy.

Utsunomiya University Spoken Dialogue Database for Paralinguistic Information Studies

Hiroki MORI¹, Hideki KASUYA¹, Makoto NAKAMURA²

¹ Graduate School of Engineering Utsunomiya University, 7-1-2 Yoto, Utsunomiya, 321-8585 Japan

² Faculty of International Studies, Utsunomiya University, 350 Mine-machi, Utsunomiya, 321-8505 Japan
uudb@speech-lab.org

The Utsunomiya University (UU) Spoken Dialogue Database for Paralinguistic Information Studies is a public corpus which is especially intended for use in understanding the usage, structure and effect of paralinguistic information in expressive Japanese conversational speech. This paper gives brief overview of the UU

Database. It also focuses on the technical details of the structure of the UU Database from the viewpoint of computer processing.

An Independent Approach for Spoken Language Analysis

Arup Saha¹, Tulika Basu², Ashoke Kr. Datta³

¹ Department of Computer Science, Jadavpur University, Kolkata, India

² Department of Linguistics, Calcutta University, Kolkata, India

³ Society for Natural Language, Technology and Research, Kolkata, India
arupmtech@gmail.com, tbasu123@gmail.com, dattashoke@yahoo.com

The paper presents an approach to spoken language analysis which is independent of the textual language processing. It uses some sort of manner based representation of a normal spoken word which gives rise to sets of words, named here as 'legeics'. The legeic of most frequently spoken 40,000 Bangla words constitutes a legeicon for Bangla. The legeics show some functionality in the sense that they could be mostly classed into two groups' verbs and non-verbs. Spoken sentences are labeled into the selected manner based representation with adequate robustness using phase space analysis. The analysis process is described. These labeled sentence signals are then segmented into legeics using the legeicon. As the legeic represents words the process can be thought of as word boundary detection. The results of operation on about 100 sentences spoken by 3 male and 3 female native informants are also presented.

Hindi Dialects Phonological Transfer Rules for Verb Root Cələ

Diwakar Mishra¹, Kalika Bali²

¹ Special Centre for Sanskrit Studies, Jawaharlal Nehru University, New Delhi

² Microsoft Research Lab India, Bangalore
diwakarmishra@gmail.com, kalikab@microsoft.com

Most Natural Language Processing (NLP) applications need to account for synchronic variations in a language as represented by its major dialects. However, most corpora available for the training and development of such systems tend to be dialect neutral. A framework that models synchronic variation can make NLP and Speech technology systems more robust to dialect variations. In this paper we present basic phonological transfer rules from standard Hindi to a number of its prominent dialects. We believe that this can be the first step towards a

more general model for dialect variation in Hindi. The rules here describe morphophonemic change in simple verb forms between dialects taking the example of verb root $c\partial l\partial$.

Language Identification using Support Vector Machine

Sanghamitra Mohanty, Basanta Kumar Swain

Department of Computer Science and Application, Utkal

University, Bhubaneswar, Orissa, India - 751004

sangham1@rediffmail.com, techmobks@yahoo.com

This paper deals with the results of our research work on language identification performed on the five languages namely Oriya, Hindi, Indian English, Sanskrit and Nepali. Language identification from the mentioned category is carried out using Support Vector Machine with radial basis function (RBF) kernel as pattern recognition classifier. The corpus used in this research work is gathered from 35 speakers for five language category. Jitter and shimmer are used as acoustic feature parameters during training and testing stage. Language identification is performed on vowel, syllable and word domain. Shimmer performs better in comparison to jitter when language identification is carried out individually. The accuracy level is further improved when both features used jointly. Our studies indicate that the average accuracy of language identification in level of vowel domain is 81.4%, syllable domain average accuracy is 76% and word domain accuracy is 78.4 %.

Description of Puma, an endangered language from eastern Nepal

Vishnu S Rai

Associate Professor

Faculty of Education, Tribhuvan University, Kathmandu, Nepal

vpsrai@yahoo.com

The present paper describes the major findings of this documentation project, which include phonology, morphology and syntax of the language. It is not at all possible to mention every outcome of this research in such a short presentation. The paper, therefore, briefly focuses on more interesting aspects of the language, such as nominalization processes, spatial deixis, two ways of suspending object agreement one following the general Kiranti model, and one involving a dedicated prefix kha-(which is not found in other Kiranti

languages), and how the language is fighting a desperate fight to preserve itself from being extinct (the language has processes which it uses to digest a loan word so that the word becomes a Puma word).

An HMM-based Hakka Text-to-Speech System

Yi-Ling Tsai, *Hsiu-Min Yu, Yih-Ru Wang, Chen-Yu Chiang,

*Lieh-Shih Lo, and Sin-Horng Chen

Institute of Communication Engineering, NCTU, Hsinchu, Taiwan

**College of Hakka, NCTU, Hsinchu, Taiwan*

**Language Center, Chung Hua University, Hsinchu, Taiwan*

yiling750124@yahoo.com.tw, kuo@chu.edu.tw, yrwang@cc.nctu.edu.tw, gene.cm91g@nctu.edu.tw,

asiisl@mail.nctu.edu.tw, schen@mail.nctu.edu.tw

In this paper, a Hakka Text-to-Speech (TTS) system is implemented. It is an HMM-based speech synthesis system. The work focuses on the realization of a Hakka parser to tag the input text into word and POS sequences. The difficulty lies in the lack of a large text corpus to train a robust Hakka parser. Motivated by the fact that Hakka is a dialect of Mandarin Chinese so as to share many linguistic properties of Chinese, we adopt a new approach to constructing a Hakka parser via extending an existing CRF-based Chinese parser to attach a Hakka dictionary and incorporate some Hakka word construction rules. Besides, a pause predictor is designed to estimate the inter-syllable locations to insert pauses for improving the fluency of the synthesized speech. A subjective quality test confirms that the Hakka TTS system is a promising one.

Real world utterance collection using voice-enabled web system for child speaker identification

Shoko MIYAMORI, Ryuichi NISIMURA, Lisa KURIHARA,

Toshio IRINO, Hideki KAWAHARA

Graduate School of Systems Engineering, Wakayama University, Japan

{s115058, nisimura, s125021, irino, kawahara}@sys.wakayama-u.ac.jp

We have been developing a method of identifying child speakers by using voices as the information related to human behavior. To obtain an adequate quantity of child voice samples to test our approach, a collection of 3,053 short sentences uttered by 1,050 users has been built via a voice-enabled web system. We confirmed the captured voices manually since these voices could include invalid recording

data. As results, the number of child voices became 1,533. We have considered the linguistic features which have different tendencies that help distinguish between child and adult utterances. The results proved that children prefer to utter simple sentences of "Single word", whereas adults answer with sentences composed of "Four kanji character compounds" and "Proverbs". In the experiments, an automatic acoustical classification and a human subjective evaluation were compared. The classifier based on combining 24-class HMMs (Hidden Markov Models) and SVMs (Support Vector Machines) derived 72.7% correct rate.

Speech Synthesis Using Epoch Synchronous Overlap Add (ESOLA)

Ashoke Kr Datta¹, Arup Saha²

¹*Society for Natural Language, Processing(SNLTR)*

²*Department of Computer Science, Jadavpur University*
dattashoke@yahoo.com, arupmtech@gmail.com

This paper presents details of a Text-To-Speech synthesis procedure using Epoch Synchronous Overlap Add (ESOLA). The synthesis model consists of 3 units: Intonation control, Duration control and Amplitude control. The intonation unit generates the F0 contour fluctuations: overshoot, vibrato and fine fluctuation. This change of the F0 contour is done by ESOLA technique. As the perceptual phonetic load is significantly borne by only in the small segment (about 1.5 ms) of the pitch-period measured from a particular point called epoch, the epoch synchrony in concatenative synthesis allows large manipulation to the extent even two octaves of the voiced regions without any significant distortion of the phonetic quality. It will be observed that changing of pitch by this method not only keep the formant structures of the speech signal constant but also introduction of spectral noise is insignificant. The use of partemes as the elements of the signal dictionary is discussed in detail.

Garhwali Speech Database

RK Upadhyay, MK Riyal

Department of physics, Govt P.G. College, Rishikesh
rku8@rediffmail.com, manoj.riyal@gmail.com

We present a progress report of the creation of speech database of Garhwali language that will be used for development of automatic speech recognition system. The speech corpus will consist of spontaneous speech as well as phonetically rich sentences of Garhwali, spoken by a variety of speakers in front of computers by using PRAAT speech software. An account of the design of phonetically rich sentences, speech data acquisition and data validation are given. A statistical analysis of the phonetic richness of isolated tokens is presented.

Garhwali Hindi is a regional dialect of Garhwal region of Uttarakhand state of India. To make the database as phonetically balanced database, special list of words and sentences are being created using a statistical package which fulfils the criteria such as the sentences used for the database (a) preferably contain all the phonemes, (b) incorporating dialectal variations from different regions and (c) rich in phonetic context.

Hindi ASR for Travel Domain

Sunita Arora, Babita Saxena, Karunesh Arora, S S Agrawal

Centre for Development of Advanced computing, Noida, India
{sunitaarora, babita, karunesharora, ssagrawal}@cdacnoida.in

This paper presents our experiments for a baseline speech recognizer for Hindi language. The recognizer is developed using Julius Speech recognition engine. Julius is a high performance; two pass large vocabulary continuous speech recognizer (LVCSR) which performs recognition taking an acoustic model and a language model as input. HTK is used for building acoustic model and SRILM toolkit is used for building language model. This system recognizes spoken sentences in the travel domain. The acoustic model is trained on 26 hours of audio data of 30 speakers. The Language Model is trained with 167057 words. The vocabulary size of the recognizer is 9103 words. The system is tested on 20 speakers and the performance of the system is reported.

Present Scenario of Forensic Speaker Identification in India

Shivani Sharma¹, S. K. Jain², R. M. Sharma³, S.S Agrawal⁴

¹Senior Research Fellow, Central Forensic Science Lab., MHA, Chandigarh, India

²Deputy Director, Central Forensic Science Lab., MHA, Chandigarh, India

³Professor, Deptt. of Forensic Science, Punjabi University, Patiala, Punjab, India

⁴Advisor, Centre for Development of Advance Computing (CDAC), Delhi, India

shivanisharma.cfsi@gmail.com

This paper presents the various problems and challenges in the field of forensic speaker identification and the available possible technologies and methods to analyze the speech sample especially in Indian scenario. The communication facilities are now not only limited to landline telephones, mobile phones but running over internet and so is crime execution. Criminals use telephones, mobile phone, satellite and wireless phone, internet phone and tape recorders for communication to establish their networks worldwide and maintain their anonymity in commission of crime such as kidnapping, threatening and hoax calls, blackmail, trafficking of illegal drugs, match fixing. Criminals use different methods of disguise and different text and languages to conceal their identity and mislead investigating agencies. In all these contexts, analysis of voice is valid and reliable source to associate the malefactors to a particular crime, if recorded. The present study also focuses on the effect of compression and decompression on voice sample when processed through various codecs.

Phonetic Segmentation Based on HMM of Hindi Speech

Archana Balyan¹, S.S. Agrawal², Amita Dev³

¹Department of Electronics and Communication Engineering, MSIT, New Delhi, India

²Advisor CDAC & Director KIIT, Gurgaon, India

³Department of Computer Science, AIT, Delhi, India
archanabalyan@rediff.com, ss_agrawal@hotmail.com

In this paper, we study the performance of baseline Hidden Markov Model (HMM) for segmentation of speech signal. Here, it is applied on single speaker segmentation task, using Hindi speech database. The automatic phoneme segmentation framework evolved imitates the human phoneme

segmentation process. A set of 44 Hindi phonemes were chosen for the segmentation experiment, wherein, we used Continuous Density Hidden Markov Model (CDHMM) with a mixture of Gaussian distribution. The left-to-right topology with no skip states has been selected as it is effective in speech recognition due to its consistency with the natural way of articulating the spoken words. This system accepts speech utterances along with their orthographic “transcriptions” and generates segmentation information of the speech. This corpus was used to develop context-independent Hidden Markov Models (HMMs) for each of the Hindi phonemes. The system was also validated against a few manually segmented speech utterances. The evaluation of the experiments shows that the best performance is obtained by using a combination of 2 Gaussians with 5 HMM states. The modeling of HMMs has been implemented using Microsoft Visual Studio 2005 (C++) and the system is designed to work on Windows operating system.

Development of speech data base for various emotions and their recognition using Neural Network Classifier

Jyoti Garg¹, Israr Khan¹, S.K.Gupta¹, S.S.Agrawal²

¹Department of Physics, (A.M.U), Aligarh 202 002, India

²KIIT college of Engineering, Gurgaon, India

jyotigarg29@gmail.com, israrkhan62@yahoo.co.in,

skgupta48@rediffmail.com, ss_agrawal@hotmail.com

This paper presents and discusses development of emotion specific Hindi speech data base and their recognition using Neural Network Classifier. The emotions are classified into four categories i.e anger, fear, happiness and sadness in addition to neutral category. Six male students of drama club uttered natural Hindi sentences in emotion categories of neutral, happiness, anger, sadness and fear. Prosody related features and spectral features were analyzed for the evaluation of emotion recognition using artificial neural network classifier. 15 mel frequency cepstral coefficients (MFCCs) were studied as spectral features whereas prosody related features consisted of mean value of pitch (F0), duration, rms value of sound pressure, and speech power. The human capability to recognize the emotion from speech was also studied and compared with classifiers. It was found that performance of neural network classifier was better than the human.

The Development of a Thai Speech Set for Telephonometry

Therdpong Daengsi¹, Apiruck Preechayasomboon², Saowanit Sukparungsee³, Patcharika Chootrakool⁴, Chai Wutiwiwatchai⁵

¹*Faculty of Information Technology, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand*

²*International Business Development Department, TOT Public Company Limited, Bangkok, Thailand*

³*Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology, North Bangkok, Thailand*

^{4,5}*Human Language Technology Laboratory, National Electronics and computer Technology Center, Pathumthani Thailand
therdpong1@yahoo.com, apiruck@tot.co.th, saowanits@hotmail.com,
patcharika.chootrakool@nectec.or.th, chai.wutiwiwatchai@nectec.or.th*

This paper presents the development for the Thai Speech Set for Telephonometry (TSST), which is mainly required for voice quality measurement for telecommunications. TSST was designed and developed by following the International Telecommunication Union – Telecommunication Standardization Sector (ITU-T) recommendation. Tasks were divided into three parts. The first part was about survey to find frequently used sentences (or phrases). The second part was to investigate for fifty frequently used sentences and to create the representative sentences of those, called Thai Text Set for Telephonometry (TTST). Finally, the last part was about speech recording in a high standard studio for TSST. The output from this work will be useful for telecommunication research and related research areas in Thai environments.

Speaker Recognition Based on Multilingual Speech

Features using Neural Network Models

Sanjay Decate, Anupam Shukla, Sanjay Kumar Singh, Ritu Tiwari

ABV-IIITM, Gwalior, India

sanjaydekate@gmail.com, dranupamshukla@gmail.com, sksinghiitm@gmail.com, tiwariritu2@gmail.com

In the present paper an attempt is made to develop multilingual speaker recognition system which is used to identify the identity of an unknown speaker among several speakers of known speech characteristics, from a sample of his or her utterances.

Every speaker has different individual characteristics embedded in

his/her speech utterances. To evaluate speech characteristics from utterances they are stored in digitized form. Speech features namely LPC, RC, APSD, Number of zero crossing and Formant frequencies are extracted from speech signal and formed speech feature vectors. The database used for this system consists of 30 speakers including both male and female from different parts of India and languages are Hindi, English and Sanskrit containing total 14 words.

The average identification rate was 87.82% with BPA and improved by 2.87% with RBF and further by 3.57% with LVQ neural network models.

Orthography development for the standardization of Bhujel: Issues and approaches

Dr. Dan Raj Regmi

*Central Department of Linguistics, Tribhuvan University, Nepal
danraj_regmi@hotmail.com*

This paper attempts to examine the issues and approaches to orthography development for the standardization of Bhujel, a preliterate and endangered Tibeto-Burman language with different dialects. It is spoken in different villages of Gorkha, Tanahun, Nawalparasi and Chitwan districts of Nepal. Developing orthography is indispensable for literacy and multilingual education. It is predominantly central for the standardization of the language itself. Writing system is a linguistic as well as a social reality (Robinson, 2003). There are three scripts choices for Bhujel: Roman, Tibetan and Devanagari. However, it is hard to find common voice regarding script choice.

Recognition of Hindi Phoneme in Rhyming Words using Vector Quantization

Shweta Sinha¹, Shalini Goyal², Mona Gaur², S.S Agrawal²

¹*KIIT College Of IT and Management, Gurgaon*

²*KIIT College Of Engineering, Gurgaon*

*meshweta_7@rediffmail.com, shalinigoyal1980@gmail.com
monagaur1311@gmail.com, ss_agrawal@hotmail.com*

This paper presents and discusses the recognition of Phoneme in Rhyming word environment using Vector Quantization. Recognition

has been carried out on rhyming words extracted from Hindi speech database. Our objective of study is to evaluate and compare the performance of Phoneme recognition when rhyming pairs are taken as input to the recognizer. The word recognition system requires a reference database of the words spoken by different speakers which is used to create templates to be used in recognition process. A decision algorithm has been proposed based upon arithmetic mean of the VQ and is used for acceptance of the matched word on the basis of defined threshold value. In this paper 39 feature vector including MFCC, delta MFCC and delta-delta MFCC of each words has been obtained and then VQ is used to create code book which works as word template in reference database for the recognition system. Samples of 22 rhyming words in Hindi were obtained from 20 speakers of different age groups from different regions. Training set is created with the data of 15 speakers and recording from 5 speakers are used for test data. It has been observed that the recognition rate for few words varied significantly when number of feature vectors being used for template creation was made to vary. The observation show the recognition rate equals 81% when code book based on 39 features are used, it came out to be 73% when word templates based on 26 features are used and is 67% when only 12 MFCC features were used in code book creation for recognition. This paper also contains a comparative study to test the recognition performance of machines and human for the same phonemes. It was observed that human perception was better than VQ based recognition for rhyming phonemes with 89% success.

Building Unit Selection Speech Synthesis in Indian Languages: An Initiative by an Indian Consortium

Hema A. Murthy¹, Ashwin Bellur¹, Vinodh Viswanath¹, Badri Narayanan¹, Anila Susan¹, G. Kasthuri¹, Raghav Krishnan¹, K. Sreenivasa Rao², Sudhamay Maity², N. P. Narendra², Ramu Reddy², Krishnendu Ghosh², K. G. Sulochana³, E. L. Abhilash³, T. Sajini³, M. Sasikumar⁴, Bira Chandra Singh⁴, Pranaw Kumar⁴, P. Vijayaditya⁵, E. Veera Raghavendra⁵, Kishore Prahallad⁵

¹IIT-Madras, ²IIT-Kharagpur, ³CDAC-Trivandrum, ⁴CDAC-Mumbai, ⁵IIT-Hyderabad.

This paper describes the efforts of a consortium consisting of five institutions in India to build TTS systems in six Indian languages –

Hindi, Telugu, Tamil, Bengali, Marathi and Malayalam. This paper discusses the research issues that are addressed in building these TTS systems and states the expected deliverables and the current state of the project. Index Terms: speech synthesis, unit selection, Indian languages.

NICT Speech and Language Resources and Corpora

Satoshi Nakamura, Kentaro Torisawa, Hisashi Kawai, Eiichiro Sumita

National Institute of Information and Communications Technology (NICT)
{satoshi.nakamura|torisawa|hisashi.kawai|eiichiro.sumita}@nict.go.jp

A universal communication technology is the one that enables human beings to **communicate with machines, breaking through** any kinds of digital divide. **One of the research goals of the universal communication is to achieve language-based communication** regardless of who or where speakers are, when or how they use it, or in which language they are communicating. Towards this **goal**, we had launched **Multi-lingual Advanced Speech and Text Research Project, MASTAR**, project in 2008 and **Advanced Language Information Forum, ALAGIN**, in 2009. This paper introduces current activities of these project and forum.

Analysis of impression-prosody mapping in Communicative speech consisting of multiple lexicons with different impressions

Yoko Greenberg¹, Hiroaki Kato², Minoru Tsuzaki³, Yoshinori Sagisaka¹

¹ *GITII/Language and Speech Science Res. Lab Waseda University, Japan*

² *NICT/ATR Media Information Science Lab, Japan, 3 Kyoko City University of Arts, Japan*

yoko.kokenawa@toki.waseda.jp, kato.hiroaki@nict.go.jp, minoru.tsuzaki@kuca.ac.jp, ysagisaka@gmail.com

Aiming at prosody control for communicative speech synthesis, we analyzed communicative prosody of phrases consisting of adverbs, adjectives and particles showing multiple impression combinations in six directions of three-dimensional impression space (confident-doubtful allowable-unacceptable and positive-negative) derived by dimension reduction using Multi Dimensional Scaling. Through

this analysis, we tried to generalize our impression-prosody mapping scheme previously proposed for phrases with single impression. Analysis showed that changes in average F0 height, F0 dynamics and utterance duration were systematically explained as a sum of single control characteristics assigned by each impression using the impression-prosody mapping scheme. These results suggest the applicability of our impression-prosody mapping scheme to more general inputs consisting of multiple lexicons with different word impressions.

Affect Recognition from Acted and Spontaneous Filipino Speech

Jocelynn Cu

Center for Empathic Human-Computer Interactions, College of Computer Studies, De La Salle University – Manila, Philippines
jji.cu@delasalle.ph

Affect recognition in speech enhances the interaction between human and machine communication. It adds to the appeal of computing systems by contributing to the user's perception of the machine's intelligence and adaptability. In this paper, we compared two types of affective speech models with a common set of acoustic features. One is trained using an acted speech dataset and the other is trained using a spontaneous speech dataset. The models classify affective speech into five classes: happy, sad, angry, fear, and neutral. Preliminary results showed a 74.8% and 68% correct classification for acted and spontaneous affective speech, respectively.

Intonation Patterns in Nepali Feedback Units

Jens Allwood¹, Bhim Narayan Regmi²

¹*SSKKII, Gothenburg University, Gothenburg, Sweden*

²*CECODES, Lalitpur, Nepal and CDL, TU, Kathmandu, Nepal*

jens@ling.gu.se, bhim_regmi@yahoo.com

This paper analyzes the intonation patterns of Nepali Feedback units. Both eliciting and giving feedback units with basic, acceptance and additional level function have been collected from a spoken language corpus. The excerpts have been taken from transcriptions of 35 recordings, along with their audio files from 11 different social activities with both male and female participants, a total of 95 participants. Xaira and Adobe Audition were used to collect, convert and extract the excerpts and PRAAT was used to get the intonation

pattern and annotate the sample examples. The study has found that rising intonation is used in all feedback eliciting units and in units where feedback has a giving non-accept function. All the other feedback giving units in Nepali have a falling intonation pattern. However, when these units function as turn holders they have rising intonation.

Spoken Disfluencies in Multilingual Spoken Corpora

Samudravijaya K

Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai, 400005, India.

samudravijaya@gmail.com

This paper is a report of the progress of an effort to create spoken corpora in 3 languages: Marathi, Hindi and Indian accented English. Multilingual subjects not only read phonetically rich sentences, but also interacted with a simulated speech interface to railway reservation information system via impromptu speech. The word level transcriptions of utterances include tags for speech disfluencies as well as unexpected sounds. A preliminary report of an analysis of speech disfluency markers is given and observations are presented and discussed.

Taiwan L2 German Database Design for Computer Assisted Language Learning

Chia-yu Chiu¹, Yuan-fu Liao², Hansjörg Mixdorff³, Hue-San Do⁴, Shing-lung Chen⁵

^{1,5}*National Kaohsiung First University of Science and Technology, Kaohsiung, Taiwan,* ²*National Taipei University of Technology, Taipei, Taiwan*

^{3,4}*Beuth University of Applied Sciences, Berlin, Germany*

chiuchiayu@hotmail.com, yfliao@ntut.edu.tw

mixdorff@beuth-hochschule.de, hsd@beuth-hochschule.de

chenst@ccms.nkfust.edu.tw

This paper reports on the progress of our joint Germany-Taiwan computer assisted language learning (CALL) project. In this year, our major goal is to collect a German speech corpus of L2 learners in Taiwan. A set of prompt sheets is thus designed to embed potential pronunciation and prosody errors in both segmental and supra-segmental level. The resulting database collection will be completed in two years.

SpeechOcean, whose legal Chinese name is Beijing Haitian Ruisheng Science Technology Ltd., is a professional Language Resources and Data-related Services supplier supporting the research and development of Technologies of many fields such as Speech (TTS, ASR), Natural language, Machine Translation, Web search, Pattern recognition, etc.

At present time it is capable of providing language resources and data-related services in 40+ languages and accents across Asia, North Africa, Europe, and both North and South America and established ten strategic workshops in countries of UK, Germany, Canada, Russia, etc. The main scopes of services Speechocean can provide are as follows:

1. **Language Resources Licensing**
 - a. Free language resources for researching.
 - b. Huge and Efficient databases to industrial developing.
2. **Data Collecting and Processing Services**
 - a. Huge volume in-country data collecting of Speech, Text, Videos, Images, Handwritings, etc.
 - b. High quality processing services in the speech transcribing and labeling, text annotating, image labeling, webpage labeling, etc.

Based on its efficient services at low cost and high quality, Speechocean has established a long-term relationship with many World Famous Customers. Now, “Kingline” has become a famous brand of Speechocean’s databases in the world.

For detailed information, please visit our website:

www.speechocean.com

or contact us by Tel: +86(10) 5873 2559

or Email us through chengxianfeng@speechocean.com

Hetauda School Of Management
and Social Sciences

is proud to be a part of **O-COCOSDA 2010**

“
IT FOR YOUNG MINDS
”

Bachelor Programs in

Information Management
Business Studies , Education
Humanities (*English, Sociology, Population*)

Masters Program in

Business Studies , Education
Humanities (*English, Sociology, Population*)



For Further information

Hetauda-4 Makawanpur , Nepal

Phone : 057-524701 ,524711

Fax : 057-524711

info@hsmonline.edu.np

