Program and Abstracts of

SLaTE 2013

Interspeech 2013 Satellite workshop on

Speech and Language Technology in Education

August 30-31 & September 1st, 2013

Maison Jean Kuntzmann
Domaine Universitaire de Saint-Martin-d’Hères, France

Coorganized by
GIPSA-lab and LIDILEM
with the ISCA-SLaTE group
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table of Contents</td>
<td>3</td>
</tr>
<tr>
<td>Organization</td>
<td>4</td>
</tr>
<tr>
<td>Message from the Local Organizers</td>
<td>5</td>
</tr>
<tr>
<td>SLaTE 2013, the program at a glance</td>
<td>7</td>
</tr>
<tr>
<td>Friday 30(^{th}) August 2013</td>
<td>9</td>
</tr>
<tr>
<td>Opening Session</td>
<td>9</td>
</tr>
<tr>
<td>Keynote 1</td>
<td>9</td>
</tr>
<tr>
<td>Children's education / children ASR</td>
<td>9</td>
</tr>
<tr>
<td>CALL I</td>
<td>11</td>
</tr>
<tr>
<td>CALL II</td>
<td>12</td>
</tr>
<tr>
<td>Saturday 31(^{st}) August 2013</td>
<td>14</td>
</tr>
<tr>
<td>Keynote 2</td>
<td>14</td>
</tr>
<tr>
<td>Demonstration of applications and posters</td>
<td>14</td>
</tr>
<tr>
<td>Gradation / Evaluation I</td>
<td>19</td>
</tr>
<tr>
<td>Gradation / Evaluation II</td>
<td>20</td>
</tr>
<tr>
<td>SLaTE assembly</td>
<td>21</td>
</tr>
<tr>
<td>Banquet</td>
<td>21</td>
</tr>
<tr>
<td>Sunday 1(^{st}) September 2013</td>
<td>22</td>
</tr>
<tr>
<td>Keynote 3</td>
<td>22</td>
</tr>
<tr>
<td>Prosody</td>
<td>22</td>
</tr>
<tr>
<td>Phonetics / phonology</td>
<td>24</td>
</tr>
<tr>
<td>Authors list</td>
<td>27</td>
</tr>
<tr>
<td>Sponsors</td>
<td>28</td>
</tr>
</tbody>
</table>
SLaTE-2013 is being organized by GIPSA-lab and LIDILEM (Grenoble, France), within the framework of the ISCA-SIG SLaTE. The organizers would like to thank Martin Russell, Helmer Strik, and Maxine Eskenazi for giving us instructive advice to organize this workshop.

**Local Organizing Committee**

Pierre Badin, GIPSA-lab  
Thomas Hueber, GIPSA-lab  
Gérard Bailly, GIPSA-lab  
Didier Demolin, GIPSA-lab  
Françoise Raby, LIDILEM

**Local logistics**

Nadine Bioud, GIPSA-lab  
Lucia Bouffard-Tocat, GIPSA-lab  
Martine Giglio, CNRS  
Emilie Magnat, LIDILEM  
Cécilia Mendès, GIPSA-lab  
Akila Mokhtari, GIPSA-lab  
Jessica Réolon, GIPSA-lab

**ISCA – SLaTE International scientific committee**

Abeer Alwan, UCLA  
Jared Bernstein, Ordinate Corp.  
Rodolfo Delmonte, University of Venice  
Maxine Eskenazi, Carnegie Mellon  
Björn Granström, KTH  
Valerie Hazan, UCL  
Diane Litman, U. of Pittsburgh  
Dominic Massaro, USCS  
Nobuaki Minematsu, U. of Tokyo  
Patti Price, PPrice.com  
Martin Russell, Univ. of Birmingham  
Stephanie Seneff, MIT  
Helmer Strik, Radboud U., Nijmegen  
Catia Cucchiarini, Radboud U., Nijmegen

**Scientific Review Committee**

The organizers would like to thank the following individuals who took part in the review of papers submitted to SLaTE 2013.

Pierre Badin  
Gérard Bailly  
Anton Batliner  
Kay Berkling  
Jared Bernstein  
Lei Chen  
Jean-Pierre Chevrot  
Catia Cucchiarini  
Febe de Wet  
Rodolfo Delmonte  
Didier Demolin  
Ryan Downey  
Olov Engwall  
Donna Erickson  
Maxine Eskenazi  
Keelan Evanini  
Horacio Franco  
Björn Granström  
Nathalie Henrich  
Thomas Hueber  
Lewis Johnson  
Hiroaki Kato  
Mariko Kondo  
Jacques Koreman  
Diane Litman  
Joaquim Listerri  
Nuno Mamede  
Nobuaki Minematsu  
Hansjörg Mixdorff  
Jean-Paul Narcy-Combes  
Thomas Pellegrini  
Patti Price  
Françoise Raby  
Martin Russell  
Stephanie Seneff  
Theban Stanley  
Helmer Strik  
Joseph Tepperman  
Oscar Saz Torralba  
Isabel Trancoso  
Chiu-Yu Tseng  
Hugo Van Hamme  
Karl Weilhammer
Message from the Local Organizers

We are very happy to welcome you to SLaTE 2013 in Grenoble!

We are about 60 from all over the world to be gathered here to discuss the last advances in the domain of Speech and Language Technology in Education.

Following SLaTE’2007 in Farmington, USA, SLaTE’2009 in Birmingham, UK, L2WS in 2010 in Tokyo, Japan, and SLaTE’2011 in Venice, Italy, this workshop is the fifth ISCA-supported SLaTE workshop.

Fourty-six contributions have been submitted in all areas of SLaTE. All full or two-page contributions have been reviewed by three reviewers, and given a score. Contributions with an average score greater than +1 have been directly accepted. Those with a score lower than – 1 have been rejected. The remaining ones have been carefully discussed by the selection committee. Finally thirty-nine contributions have been selected for presentation at the workshop in either oral or demonstration and poster sessions.

These contributions testify to very diverse approaches: children oriented research, corpus based studies of speech diversity, mobile or web-based applications, elaborate dialogue systems and serious games, phonetics / phonology studies on L2.

We have organised the workshop in the following way, with the aim to provide lively discussions.

We have planed a short round table at the end of each of the five oral sessions. We rely on chair persons to submit general or important questions about the field to the speakers or to the audience.

We will have a session for demonstration and posters with a dozen contributions. This session will start with a series of three minutes presentations of the contributions allowing authors to attract a larger audience.

Long lunches à la Française at the Canberra restaurant will provide time to discuss important matters … or to enjoy the university swimming pool.

In order to promote opening towards fields that are clearly related to SLaTE but finally not so present in this community, we have invited three keynote speakers known for their broad ideas and their vision of their domains.

*Diane Litman*, from the University of Pittsburgh, will tell us in a few moments her vision of how speech and language processing can be used to enhance dialogue applications for teaching STEM, *i.e.* Science, Technology, Engineering, and Mathematics.

Tomorrow morning, *Jozef Colpaert*, from the University of Antwerp, will share his experience about what is important in CALL systems, using the design concept behind a project on pronunciation training, and more generally about the new concept of Educational Engineering.

Finally, on Sunday morning, *Mary Beckman*, from the Ohio State University, will discuss how speech technology can be used to learn more about child speech and in particular to evaluate cross-language differences in the early emergence of contrasting vowel categories.
We are very lucky to be financially supported by a number of institutions. GIPSA-lab, the Engineering education Grenoble INP group, the Stendhal University that deals with Language and Literature, the Joseph Fourier University involved in Scientific and Medical research, and also from the Grenoble Cognition Pole which aims at initiating and coordinating all Grenoble research teams involved in cognition. We have also received a strong support from the Rhône-Alpes Region through its Academic Research Community on Information and Communication Technologies and Innovating Computer Usage. Finally, we could also mention the supports of the Grenoble – Alpes metropole consortium of cities and from the Grenoble city itself. We are grateful to all of them for supporting the workshop.

We would also like to thank all reviewers who accepted to read and assess the papers and provided relevant comments.

We hope that all of you will enjoy the workshop.

Pierre Badin, Thomas Hueber, Gérard Bailly, Didier Demolin & Françoise Raby
### SLaTE 2013, the program at a glance

<table>
<thead>
<tr>
<th>Time</th>
<th>Friday 30th August 2013</th>
<th>Saturday 31st August 2013</th>
<th>Sunday 1st September 2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30</td>
<td>Registration</td>
<td>Keynote 2 : Jozef Colpaert</td>
<td>Keynote 3 : Mary Beckman</td>
</tr>
<tr>
<td>09:00</td>
<td>Registration</td>
<td>Keynote 2 : Jozef Colpaert</td>
<td>Keynote 3 : Mary Beckman</td>
</tr>
<tr>
<td>09:40</td>
<td>Opening session</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:00</td>
<td>Coffee break</td>
<td>10:00 Coffee break</td>
<td>Coffee break</td>
</tr>
<tr>
<td>10:20</td>
<td>Keynote 1 : Diane Litman</td>
<td>10:20 Summary of demonstrations</td>
<td></td>
</tr>
<tr>
<td>10:40</td>
<td></td>
<td>10:40 Summary of demonstrations</td>
<td></td>
</tr>
<tr>
<td>11:00</td>
<td>Coffee break</td>
<td>11:00 Summary of demonstrations</td>
<td></td>
</tr>
<tr>
<td>11:20</td>
<td></td>
<td>11:20 Summary of demonstrations</td>
<td></td>
</tr>
<tr>
<td>11:40</td>
<td>Children's education / children ASR</td>
<td>11:40 Demonstrations and posters</td>
<td></td>
</tr>
<tr>
<td>12:00</td>
<td></td>
<td>11:40 Demonstrations and posters</td>
<td></td>
</tr>
<tr>
<td>12:20</td>
<td>Round Table</td>
<td>11:40 Demonstrations and posters</td>
<td></td>
</tr>
<tr>
<td>12:30</td>
<td>Lunch at the Canberra</td>
<td>12:30 Lunch at the Canberra</td>
<td>13:20 Phonology / phonetics</td>
</tr>
<tr>
<td>14:00</td>
<td>CALL I</td>
<td>Gradation / evaluation I</td>
<td>14:00</td>
</tr>
<tr>
<td>14:20</td>
<td>CALL I</td>
<td>Gradation / evaluation I</td>
<td>14:20 Coffee break</td>
</tr>
<tr>
<td>14:40</td>
<td>CALL I</td>
<td>Gradation / evaluation I</td>
<td>14:40</td>
</tr>
<tr>
<td>15:00</td>
<td>Coffee break</td>
<td>15:00 Coffee break</td>
<td>15:00 Phonology / phonetics</td>
</tr>
<tr>
<td>15:20</td>
<td>CALL II</td>
<td>Gradation / evaluation II</td>
<td>15:20 Round Table</td>
</tr>
<tr>
<td>15:40</td>
<td>CALL II</td>
<td>Gradation / evaluation II</td>
<td>15:40 Round Table</td>
</tr>
<tr>
<td>16:00</td>
<td>Round Table</td>
<td>Gradation / evaluation II</td>
<td>16:00 Closing session</td>
</tr>
<tr>
<td>16:20</td>
<td></td>
<td>Gradation / evaluation II</td>
<td>16:00 Closing session</td>
</tr>
<tr>
<td>16:40</td>
<td></td>
<td>Gradation / evaluation II</td>
<td>16:00 Closing session</td>
</tr>
<tr>
<td>17:00</td>
<td></td>
<td>Gradation / evaluation II</td>
<td>16:00 Closing session</td>
</tr>
<tr>
<td>19:00</td>
<td>Bastille cable car</td>
<td>19:00 Bastille cable car</td>
<td>23:00 Pèr' Gras restaurant</td>
</tr>
<tr>
<td>23:00</td>
<td></td>
<td>19:00 Bastille cable car</td>
<td>23:00 Pèr' Gras restaurant</td>
</tr>
</tbody>
</table>

**Note:** The program includes sessions on topics such as children's education, children ASR, CALL, Gradation, evaluation, and more. Keynote speakers include Jozef Colpaert, Diane Litman, and Mary Beckman. The day concludes with a Bastille cable car and Pèr' Gras restaurant activities.
Friday 30th August 2013

Opening Session
• Pierre Badin, Thomas Hueber, Gérard Bailly, Didier Demolin, Françoise Raby

Keynote 1
Chair: Maxine Eskenazi
• Diane Litman

Enhancing the effectiveness of spoken dialogue for STEM education

Abstract: This talk will illustrate some of the opportunities and challenges in applying speech and language processing to two types of STEM (Science, Technology, Engineering, and Mathematics) dialogue applications: 1) one-on-one physics tutoring, where students engage in dialogues with either a computer or human tutor, and 2) engineering design, where students engage in multi-party dialogue to complete a group project. I will first present results illustrating that relationships exists between student learning and both student affect, as well as lexical/prosodic entrainment between conversational partners. I will then illustrate our use of such findings to build better educational dialogue systems.

Children's education / children ASR
Chair: Mary Beckman & Martin Russel
• Jared Bernstein, Ognjen Todic, Kayla Neumeyer, Katharyn Schultz & Liang Zhao

Young children’s performance on self-administered iPad language activities

Abstract: Twenty-nine different English language activities were implemented on a touch-tablet computer. Some activities were focused on a single skill (e.g. reading or speaking), while others involved several integrated skills (e.g. listening and writing). Materials were presented in several modalities, including speech only, speech with figure, silent video, text, speech with text, and speech with text and figure. Response modalities included speech, typing, touch, dragging screen objects, selecting and/or arranging words, and drawing figures. Various Test-like sequences of 24 to 45 activities were presented to 784 children, 53% from non-English speaking homes. Analysis of over 28,000 responses to self-administered activities indicates that most activities can be successfully modeled by a single short video example. By age 7 or 8 years, nearly all children will respond meaningfully to about 95% of these specific activities. Examples of these activities and child responses are presented.
• Felix Claus, Hamurabi Gamboa Rosales, Rico Petrick, Horst-Udo Hain & Rüdiger Hoffmann

A survey about ASR for children

Abstract: This paper is intended to survey the state of the art of automatic speech recognition (ASR) for children's speech. Investigating ASR for children is a current trend in research. Therefore databases of children's speech are needed for training and testing of ASR systems. In the first part of this paper the most relevant databases of children's speech are described. There are less speech data of children available than of adults and speech of preschool children is even more rarely available. In the second part of this paper the common techniques for recognizing children's speech are summarized. Most investigations about children's ASR focus on the acoustic model. The common methods are described and approaches regarding the lexical and speech model are mentioned subsequently. In an extensive literature research we collected papers investigating ASR for children. Several studies have been carried out investigating children's ASR. Due to the lack of data from preschool children only a few investigations for this age group have been accomplished. This is illustrated by presenting a statistic on the age of the children in past studies.

• Annika Hämäläinen, Fernando Miguel Pinto, Silvia Rodrigues, Ana Júdice, Sandra Morgado Silva, António Calado & Miguel Sales Dias

A multimodal educational game for 3-10-year-old children: Collecting and automatically recognising European Portuguese children’s speech

Abstract: Speech interfaces have tremendous potential in education. In this paper, we present our work in the Contents for Next Generation Networks project, an ongoing Portuguese industry-academia collaboration developing a multimodal educational game aimed at improving the physical coordination and the basic mathematical and musical skills of 3-10-year-old children. We focus on our work in the area of children’s speech recognition: designing, collecting, transcribing and annotating a 21-hour corpus of prompted European Portuguese children’s speech, as well as our first experiments with different acoustic modelling approaches. Our speech recognition results suggest that training children’s speech models from scratch is a more promising approach than retraining adult speech models using children’s speech when a sufficient amount of training data is available from the targeted age group. This finding also holds for adult female speech models retrained using children’s speech. As compared with a baseline recogniser comprising gender-dependent adult speech models, the best-performing children’s speech models that we have trained so far – gender-independent cross-word triphones trained with 17.5 hours of speech from 3-10-year-old children – resulted in a 45-percent (relative) decrease in word error rate in a task expecting isolated cardinal numbers, sequences of cardinal numbers or musical notes as speech input.
CALL I

Chairs: Jared Bernstein & Jozef Colpaert

• Pei-Hao Su, Tien-Han Yu, Ya-Yunn Su & Lin-Shan Lee

A cloud-based personalized recursive dialogue game system for computer-assisted language learning

Abstract: In this paper we present the design and experimental results of a cloud-based personalized recursive dialogue game system for computer-assisted language learning. A number of tree-structured sub-dialogues are used sequentially and recursively as the script for the game. The dialogue policy at each dialogue turn is optimized to offer the most appropriate training sentence for every individual learner considering the learning status, such that the learner can have the scores for all selected pronunciation units exceeding a pre-defined threshold in minimum number of turns. The policy is modeled as a Markov Decision Process (MDP) with high-dimensional continuous state space and trained with a huge number of simulated learners generated from a corpus of real learner data. A real cloud-based system is implemented and the experimental results demonstrate promising outcomes.

• Elizabeth Davis, Oscar Saz & Maxine Eskenazi

POLLI: a handheld-based aid for non-native student presentations

Abstract: Language learning is finally leaving the classroom for the real world. This usually takes the form of a dedicated implementation on a handheld device. In this paper we describe POLLI, an app that helps non-native students prepare scientific presentations.

Index Terms: handheld applications, pronunciation skills

• Helmer Strik, Polina Drozdova & Catia Cucchiarini

GOBL: Games Online for Basic Language Learning

Abstract: In the GOBL project we develop and test small web-based mini games for low-educated and disadvantaged beginning learners of Dutch, English, and French. An innovative aspect of this project is that we incorporate language and speech technology to practice speaking skills. The present paper explains the notion of mini games employed in the project and the advantages of their use with respect to the target group. We then present the first results of the project concerning pedagogical and user requirements, based on the literature and user-based research. We then introduce our plans for the immediate future.

• Carrie Cai, Robert Miller & Stephanie Seneff

Enhancing speech recognition in fast-paced educational games using contextual cues

Abstract: Arcade-style games like Tetris and Pacman are often difficult to adapt for educational purposes because their fast-paced intensity and keystroke-heavy nature leave little room for simultaneous practice of other skills. Incorporating spoken language technology could make it possible for players to learn as they play, keeping up with game speed through multimodal interaction. To date, however, it remains exceedingly difficult to augment fast-paced games with speech interaction because the frustrating effect of recognition errors highly compromises entertainment. In this paper, we design a modified version of Tetris with speech
recognition to help students practice and remember word-picture mappings. Using utterances collected from learners interacting with the speech-enabled Tetris game, we present and evaluate several techniques for leveraging contextual cues to increase recognition accuracy in fast-paced game environments.

CALL II

**Bart Penning de Vries, Stephen Bodnar, Catia Cucchiari, Helmer Strik & Roeland van Hout**

*Spoken grammar practice in an ASR-based CALL system*

**Abstract**: In this paper we present a computer assisted language learning (CALL) system that is developed to practice grammar in spoken language. To enable this, the system uses Automatic Speech Recognition (ASR) to process the L2 learner’s responses. We investigate the possibility of providing corrective feedback (CF) on learner errors, and compare that with self-monitored language learning through output practice. In this paper we present the comparison of the two conditions 1) one group of learners received oral practice and CF on spoken performance, and 2) the other group received oral practice and no CF on spoken performance. We found that our system is successful for L2 speaking practice. The main finding is that both groups show learning after treatment. Between the groups, we did not find a learning difference, but the groups’ sessions proceeded differently. Additionally we found that the CF group was more positive about the system than the NO CF group.

**Stephen Bodnar, Bart Penning de Vries, Catia Cucchiari, Helmer Strik & Roeland van Hout**

*Learners’ situated motivation in oral grammar practice with an ASR-enabled CALL system*

**Abstract**: Advances in speech recognition (ASR) technology have resulted in computer applications that provide compelling forms of speaking practice to learners of a second language (L2). Evaluation of such applications typically does not include an analysis of how learners' situated motivational states fluctuate during language practice. In connection with recent developments in L2 motivation theory, this paper investigates situated learner motivation in practice with an ASR-enabled system. An experiment was conducted in which our system provided oral grammar practice for Dutch L2 under two learning conditions: with and without corrective feedback. We report on learners' motivational experiences by triangulating an analysis of 1) motivational trajectories from a periodic motivation questionnaire, 2) post-practice reflective questionnaires, and 3) behavioural log data recorded by the system during practice. Our analysis shows that learners maintained positive attitudes towards the system throughout practice and became increasingly confident over time.
Kyusong Lee, Soo-Ok Kweon, Hae-Ri Kim & Gary Geunbae Lee

Filtering-based automatic cloze test generation

Abstract: We propose a method to generate high-quality cloze test questions using a computational approach. Previous methods for automatic cloze test generation have contained some problems; specifically, there can be multiple correct answers. We found that approximately 50% of the generated answers have such errors with previous methods, which requires human post-editing was necessary in previous research. We propose an N-gram filtering method that can detect the answer to a given question. We compare the errors of the generated questions before and after applying the filtering methods. We found that our filtering method can select quality distractors by reducing errors in the generated questions. Moreover, when we generate cloze tests using semantic similarity, non-native speakers are very hard to answer the questions.

Manny Rayner & Nikos Tsourakis

Methodological issues in evaluating a spoken CALL game: can crowdsourcing help us perform controlled experiments?

Abstract: We summarise a series of experiments we have carried out over the last three years on CALL-SLT, a speech-enabled web-based CALL game for learning and improving fluency in domain language, focussing on the methodological aspects. In particular, we argue that our previous evaluations have been systematically flawed due to the lack of a control group. We present a detailed description of our most recent evaluation, where 130 subjects, recruited using crowdsourcing methods, followed a short course in basic French over a period of one week, with 24 subjects completing the course. About a third of the subjects (half of the ones that finished) were assigned to a control group who used a version of the system with speech recognition feedback disabled; subjects in both groups demonstrated significant improvements in language skills over the duration of the experiment, but the improvements were significantly larger for the non-control subjects. We argue in conclusion that this type of experiment opens up interesting new ways to attack the difficult problem of performing controlled experiments with CALL applications.
Saturday 31st August 2013

Keynote 2
Chair: Helen Meng

• Jozef Colpaert
The role and shape of speech technologies in well-designed language learning environments

Abstract: No technology carries an inherent, direct, measurable and generalizable effect on learning. Nor does speech technology. Its assumed added value is not a starting point for design, but a hypothesis resulting from design. Role and shape of speech technologies are simply logical and natural consequences of the specification of an optimal language learning environment for a specific context. This ecological paradigm shift will be illustrated by a discussion of the design concept behind a project on pronunciation training. Challenges for CALL research on this topic will be briefly discussed. This presentation will conclude with some considerations on Educational Engineering and how this approach can be recognized as a novel research method.

Demonstration of applications and posters
Chairs: Thomas Hueber & Diane Litman

• Jeesoo Bang, Sechun Kang & Gary Geunbae Lee
An automatic feedback system for English speaking integrating pronunciation and prosody assessments – D1

Abstract: We have proposed a computer-assisted language learning (CALL) system, called Postech English Speaking Assessment and Assistant (PESAA) for non-native English learners, especially Koreans, to improve their overall language skills. PESAA is an automatic feedback system for speaking English that integrates both pronunciation and prosody assessments. The system has three error-feedback modules: the pronunciation, rhythm, and phrase break error-feedback modules. PESAA generates scores on each of three modules as well as a combined score. The pronunciation assessment gives feedback that is based on comparing canonical and actual phoneme alignment results. The rhythm and phrase break assessments give feedback that is based on comparing predictions with detected results. English learners can use PESAA to practice pronunciations, rhythm and phrase breaks by themselves. We evaluated PESAA in three different ways: accuracy, correlation of the system’s assessment with human assessments, and user satisfaction from the expected learning effectiveness and user interface. The evaluation showed that PESAA could work as a CALL system well, with good accuracy and positive learning results for the users.
• **Haruko Miyakoda**

*Visual approach to speech sounds – D2*

**Abstract:** Many people often struggle to master the pronunciation of foreign languages without much success. One of the reasons why L2 learners are not successful is because teaching pronunciation in the classroom is usually marginalized. With the advent of computers, this problem may partially have been overcome, due to the fact that many different types of systems and software for autonomous learning have been developed, allowing learners to improve their pronunciation skills outside the classroom. However, there are few, if any, systems and software that can present a form of visual feedback that allows learners to actually understand what their problems are.

In this paper, we present the auditory-visual pronunciation system that we have developed. One of the key features of this system is that it employs easy-to-understand visuals of the speech organ that can be seen from different angles. In addition, the internal organs can also be presented by changing the mode to transparent. Furthermore, movement of the speech organs can freely be adjusted by the instructors so that the learner’s movements (especially the deviant) can be highlighted by comparing them with those of the model samples.

• **Hiroko Hirano, Ibuki Nakamura, Nobuaki Minematsu, Masayuki Suzuki, Chieko Nakagawa, Noriko Nakamura, Yukinori Tagawa, Keichi Hirose & Hiroya Hashimoto**

*OJAD: a free online accent and intonation dictionary for teachers and learners of Japanese – D3*

**Abstract:** We developed the very first online and free framework for teaching and learning Japanese prosody including word accent and phrase intonation. This framework is called OJAD (Online Japanese Accent Dictionary), which provides three functions. Subjective assessment by teachers shows very high pedagogical effectiveness of the framework.

• **Rodolfo Delmonte & Ciprian Bacalu**

*SPARSAR: a System for Poetry Automatic Rhythm and Style AnalyzeR – D4*

**Abstract:** Any poem can be characterized by its rhythm which is also revealing of the poet's peculiar style. In turn, the poem's rhythm is based mainly on two elements: meter, that is distribution of stressed and unstressed syllables in the verse, presence of rhyming and other poetic devices like alliteration, assonance, consonance, enjambements, etc. which contribute to poetic form at stanza level. Traditionally, poetic meter is visualized by a sequence of signs, typically a straight line is used to indicate vowels of stressed syllables and a half circle is positioned on vowels of unstressed ones. We also agree with this view, however, we would like to be more specific on the notion of rhythm that we intend to purport. In our view, a prosodic acoustic view needs to be implemented as well, if any precise definition of rhythm and style is the goal. Syllables are not just any combination of sounds, and their internal structure is fundamental to the nature of the poetic rhythm that will ensue. This is partly amenable to the use and exploitation of poetic devices, which we also intend to highlight in our system. But what is paramount in our description of rhythm, is the use of the acoustic parameter of duration. In our demo we will show how poems can be characterized by the use of rhythmic and stylistic features in a very revelatory manner, by comparing metrically similar poems of the same poet.
• **Catia Cucchiarini, Ineke van de Craats, Jan Deutekom & Helmer Strik**  
*The digital instructor for literacy learning – D5*

**Abstract:** The DigLIn project aims at providing concrete solutions for adult literacy students by developing and testing L2 literacy acquisition material in four different languages and by employing Automatic Speech Recognition (ASR) to analyse the learner’s read speech output and provide feedback. We develop the technology and design sample exercises for different languages (Dutch, English, German and Finnish) and test them in literacy classes in adult education centres with adult L2 learners. Existing language learning material for low-educated second language learners developed at Friesland College (the digital sources of the FC Sprint2) is augmented with an ASR module capable of recognizing what the learners say, of diagnosing possible errors in reading aloud or pronunciation and of providing practice and feedback in learning to read aloud in the L2.

• **Nic J. De Vries & Febe De Wet**  
*Off-line mobile-assisted vocabulary training for the developing world – D6*

**Abstract:** Mobile-assisted language learning applications (MALL) has significant potential in the developing world where access to teachers and classrooms are real barriers to learning. This demonstration of an Android-based mobile language learning application is used to teach vocabulary and employs off-line Automatic Speech Recognition (ASR) and Text-to-speech (TTS) technologies with custom-built language and acoustic models, incorporating the key design criteria outlined in the article.

• **Pei-Hao Su, Tien-Han Yu, Ya-Yunn Su & Lin-Shan Lee**  
*NTU Chinese 2.0: A personalized recursive dialogue game for computer-assisted learning of Mandarin Chinese – D7*

**Abstract:** We present and demonstrate a cloud-based personalized dialogue game for computer-assisted learning of Mandarin Chinese. A sequence of tree-structured sub-dialogues in restaurant scenario are linked recursively and used as the script for the game. Based on NTU Chinese, a Mandarin Chinese pronunciation evaluation software (http://chinese.ntu.edu.tw/), the user can get immediate evaluation on pronunciation, pitch, timing and emphasis and corresponding corrective feedback on each syllable as well as on sentence level for each utterance produced. The system policy is optimized to offer personalized dialogue path planning for each individual learner such that more practice opportunities are given along the dialogue path to poorly produced pronunciation units. When using the system, the learner can practice the sub-dialogues in either sequential or random order; at each dialogue turn, the learner also can choose to pronounce an arbitrary candidate sentence or following the recommended sentences by the system policy. Following the system recommendation along the sub-dialogues sequentially offers the fastest learning though. The above evaluation and learning records are displayed and stored in personal profile.
The system framework is modeled as a Markov Decision Process (MDP) with high-dimensional continuous state space considering the learning status of the learner. The dialogue policy is trained using a huge number of simulated learners generated from a corpus recorded by 278 real Mandarin Chinese learners from 36 countries with various mother tongues. The detailed principles of this system are presented in a companion paper also submitted to SLaTe 2013 [1]. This is a joint work with the International Chinese Language Program of National Taiwan University.

Karin Harbusch, Johannes Härtel & Christel-Joy Cameran
COMPASS III: Teaching L2 grammar graphically on a tablet computer – D8

Abstract: We demonstrate a prototype of the tablet-based L2 grammar teaching system COMPASS III. COMPASS stands for COMbina-torial and Paraphrastic Assembly of Sentence Structure; for a de-scription of the underlying computational-linguistic software, see [1]. COMPASS invites the student to construct sentences by composing syntactic trees out of lexically anchored “treelets” via the graphical drag & drop user interface provided by tablet and touchscreen. After each move (i.e. each attempt to combine two treelets, or to reorder a branch), the system’s natural-language generator computes all possible grammatically well-formed sentences entailed by the attempted tree. COMPASS provides positive feedback if the student-composed tree belongs to the well-formed set, and negative feedback otherwise. In the latter case, COMPASS may propose alternatives based on a comparison between the student-composed tree and its own well-formed trees (informative feedback on demand). As system feedback may explicitly refer to grammar rules, the learner needs to have elementary syntactic knowledge. COMPASS III targets L2 learners of German with high-school level understanding of word classes and grammatical functions. The user interface allows the student to select words and to move (parts of) trees around through finger or stylus gestures. No typing is required. COMPASS III focuses on word order and case morphology—difficult topics in L2 German. The grammar formalism underlying COMPASS is Performance Grammar [2], which assumes separate rules for the hierarchical structures of a sentence and the linear order of its constituents. This split allows the student to break sentence construction exercises into relatively small parts. For instance, the learner can select a word and inflect it according to the intended grammatical function without having to worry about the linear position of the constituent in the sentence under construction. At any time during this “scaffolded” sentence construction process, the tree built so far remains visible on the screen, ready to be expanded by attaching additional words/treelets; any earlier decision can be undone and corrected.

Imran Ahmed, Meghna Pandharipande & Sunil Kopparapu
A suite of mobile applications to assist speaking at right speed – D9

Abstract: One of the prominent reason for ineffective communication in call center telephone conversations is primarily due to the manner in which the voice agents speak and not necessarily due to what they speak. Speaking rate, a non-linguistic aspect of speech, is a critical factor affecting intelligibility and comprehension of speech in general and specifically in call center telephone conversations. There have been attempts to monitor speaking rate of agents in a call center setup. However, there has been a paradigm shift to the conventional call centers with companies opting for home-agents. In this model the agents, unlike in the current call center setup, can operate virtually from anywhere using their mobile phones. In this
paper, we present SpeakRite - a suite of mobile applications that can assist home-agents. While one of the components of SpeakRite analyzes the speaking rate during an ongoing telephone conversation and provides a real time feedback to assist the speaker modify his speaking rate, there are several other components that allow the home-agent to assess their speaking rate and learn to speak right themselves.

• **Teeraphon Pongkittiphan, Nobuaki Minematsu, Takehiko Makino & Keikichi Hirose**

*Automatic detection of the words that will become unintelligible through Japanese accented pronunciation of English – D10*

**Abstract:** This study examines automatic detection of the words that will be unintelligible if they are spoken by Japanese speakers of English. In our previous study, 800 English utterances spoken by Japanese speakers, which contained 6,063 words, were presented to 173 American listeners and correct perception rate was obtained for each spoken word. By using the results, in this study, we define the words that will be very unintelligible through Japanese accented English pronunciation and also define the words that will be rather unintelligible. Then, by using Classification And Regression Tree (CART) with linguistic features and lexical features only, we examine automatic detection of these words. After that, we introduce an additional feature derived by considering phonological and phonotactic differences between Japanese and English. This additional feature is found to be very effective and our proposed method can detect very unintelligible words and rather unintelligible words automatically with F1-scores of 65.44 and 70.45 [%], respectively.

• **Morten Højfeldt Rasmussen & Zheng-Hua Tan**

*Fusing eye-gaze and speech recognition for tracking in an automatic reading tutor – A step in the right direction? – D11*

**Abstract:** In this paper we present a novel approach for automatically tracking the reading progress using a combination of eye-gaze tracking and speech recognition. The two are fused by first generating word probabilities based on eye-gaze information and then using these probabilities to augment the language model probabilities during speech recognition. Experimental results on a small dataset show that the tracking error rate of the system using only speech recognition is 37.9% whereas the tracking error rate for the system that incorporates eye-gaze tracking into the speech recognizer is 35.8%.

• **Manny Rayner & Nikos Tsourakis**

*A speech-based internet game for beginner students of English – D12*

**Abstract:** We demonstrate a web-enabled serious game, based on the CALL-SLT platform, which is intended to help German-speaking beginner students of English improve their generative and auditory competence. Empirically, English is a language which German-speakers find easy to learn; the main challenge for the student is not so much grammar or phonology, but rather acquiring the experience needed to build up confidence. The game, designed to address this need, offers a short course of eight interactive lessons using a combined vocabulary of about 450 words. It was developed in collaboration with a secondary school teacher, with the content taken from a commonly used textbook. The system is gamified to increase student motivation, using common game elements such as badges.
The game is structured as a series of short dialogues between the student and the machine, where the student is encouraged to use simple language in practical contexts like booking a hotel room or ordering a meal in a restaurant. At each turn, the system starts by playing a short video file in English (“How may I help you?”), and simultaneously displays a piece of text in German indicating to the student how they should reply (frag: Zimmer für 2 Nächte). The student gives a spoken response; they can usually respond in several different ways to each prompt (“A room for 2 nights please”/ “Could I have a room for 2 nights” / “I would like a room for 2 nights” …). If the student is uncertain how to respond, they can ask for a help example, which is presented in both written and spoken form. After the student answers, the system performs speech recognition, machine translation and matching, and either accepts or rejects. Depending on the result and the defined lesson structure, it can either go to a new step or repeat the current one; there are multiple dialogue paths. The student gains badges by completing lessons and achieving a minimum score dependent on the difficulty of the badge.

• Thomas Hueber, Gérard Bailly, Pierre Badin & Frédéric Elisei

Vizart3D – Real-time system of visual articulatory feedback – D13

Abstract: We present recent developments on our system of visual articulatory feedback, named Vizart3D, which is based on the 3D orofacial clone developed at GIPSALab. This system aims at providing any speaker with a feedback on his/her own articulation, especially by displaying the tongue movements. Application areas are computer-assisted pronunciation training (phonetic correction) for second-language learning and speech rehabilitation. In our system, the orofacial clone is animated in real-time from the speech audio signal, using a statistical mapping method based on multiple Gaussian mixture regressions (GMR). This approach combines in the same framework a spectral mapping step with an acoustic-articulatory inversion step. All the modules of the system are implemented in the real-time environment Max/MSP (acoustic analysis, mapping and 3D rendering). More information is available at http://www.gipsa-lab.fr/projet/vizart3D/.

Gradation / Evaluation I

Chairs: Pierre Badin & Catia Cucchiarini

• Vaishali Patil & Preeti Rao

Automatic pronunciation feedback for phonemic aspiration

Abstract: The computer-assisted learning of spoken language is closely tied to automatic speech recognition (ASR) technology which, as is well known, is challenging with non-native speech. By focusing on specific phonological differences between the target and source languages of non-native speakers, pronunciation assessment can be made more reliable. Aspiration, an important phonemic attribute in plosives of Indo-Aryan languages such as Hindi, Marathi and Gujarati, is rarely found in the world’s languages. The improper production of the aspiration contrast is thus often the most important cue to non-native accents of spoken Hindi. A system for the detection of phonemic aspiration in unvoiced and voiced stops based on discriminative acoustic features is shown to be effective for rating non-native accents and providing reliable phoneme-level feedback.
• Ann Lee & James Glass

**Pronunciation assessment via a comparison-based system**

**Abstract:** In this paper, we present preliminary results on applying a comparison-based framework to the task of pronunciation scoring. The comparison-based system works by aligning a student’s utterance with a teacher’s utterance via dynamic time warping (DTW). Features that describe the degree of mis-alignment are extracted from the aligned path and the distance matrix. We focus on a dataset in Levantine Arabic, a low-resource language for which there is not enough automatic speech recognition (ASR) capability available. Three different speech representations are investigated: MFCCs, Gaussian posteriorgrams, and English phoneme state posteriorgrams decoded on Levantine data. Experimental results show that the system can improve both correlation and mean squared error between machine predicted scores and human ratings compared to a template-based system.

Gradation / Evaluation II

• Hao Wang, Xiaojun Qian & Helen Meng

**Predicting gradation of L2 English mispronunciations using crowd sourced ratings and phonological rules**

**Abstract:** Pedagogically, CAPT systems can be improved by giving effective feedback based on the severity of pronunciation errors. We obtained perceptual gradation of L2 English mispronunciations through crowdsourcing, and conducted quality control utilizing the WorkerRank algorithm to refine the collected results and reach a reliable consensus on the ratings of word mispronunciations. This paper presents our work on modeling the relationship between the phonetic mispronunciations and the actual word ratings. Based on phonological rules representing phonetic mispronunciation productions, we propose two approaches to predict the gradation of word mispronunciations. Reasonable correlation and agreement are found between the human-labeled and machine-predicted gradations for both approaches, which imply that the use of phonological rules in word-level mispronunciation gradation prediction is promising.

• Jeesoo Bang & Gary Geunbae Lee

**Determining sentence pronunciation difficulty for non-native speakers**

**Abstract:** This paper investigates the features that determine the sentence pronunciation difficulty for non-native speakers. We selected three types of features: length, word frequency, and phonemes that Korean speakers generally replace with other phonemes. Support vector machines and a multiple linear regression model were used to determine the pronunciation difficulty of given sentences, and the results were measured with a five-fold cross validation. We demonstrated that these features could determine sentence pronunciation difficulty with an accuracy and a correlation coefficient sufficient for computer-assisted pronunciation training (CAPT) systems. The combination of all three feature types had the highest accuracy and correlation coefficient in determining sentence pronunciation difficulty. For single features, the length-based feature type was the most accurate in determining sentence pronunciation difficulty. The phoneme-specific feature type also had high accuracy. Length, phoneme, and word features can be used to guide the automatic choice of sentences for CAPT systems that depend on users’ proficiency levels.
• Wenting Xiong, Keelan Evanini, Klaus Zechner & Lei Chen

Automated content scoring of spoken responses containing multiple parts with factual information

Abstract: This paper presents approaches to automated content scoring of spoken language test responses from non-native speakers of English which contain multiple parts addressing factual information that the test taker has previously heard via auditory stimulus materials. While previous work relating to content scoring of spontaneous, unpredictable speech has focused only on entire responses and on general topic matching approaches, such as content vector analysis, the specific nature of spoken responses in our data requires response segmentation and extraction of features that indicate the relevance and correctness of the facts contained in the different parts of the response. Our best content features, based on similarity with key facts and concepts, achieve correlations of $r = 0.615$ (for speech recognition output) and $r = 0.637$ (using human transcriptions) with expert human rater scores. Furthermore, we show that these content features outperform traditional vector space based features. Finally, we demonstrate that the performance of a scoring model based on a combination of features developed previously and some of the newly designed content features improves significantly from $r = 0.624$ to $r = 0.664$ on an unseen evaluation set when using speech recognition output.

SLaTE assembly
• Martin Russel, president

SLaTE discussions and votes

Banquet
• Bastille cable car

Transportation to the restaurant

• Chez le Pèr' Gras

Restaurant

• Charlatan Transfer

Vocal jazz
Keynote 3  
Chair: Gérard Bailly  

- Mary Beckman  

**Enriched technology-enabled annotation and analyses of child speech**  

**Abstract**: This talk will review a range of studies illustrating the ways in which speech technology has enabled richer analyses of corpora of children's speech and non-speech vocalizations. One set of studies involves VLAM, an articulatory synthesis system that models the transfer functions of infant- and child-proportioned vocal tracts. VLAM has been used to evaluate cross-language differences in the early emergence of contrasting vowel categories. Another set of studies involves the paidologos corpus of utterances elicited in a picture-prompted word-repetition task from 2- through 5-year-old child speakers of a variety of languages. Tools for tagging the corpus on multiple tiers allow researchers to extract target sounds for spectral analysis as well as for calculating transcribed accuracy rates. The tag sets also can be used to extract stimuli for perception experiments, yielding naive-listener responses that can then become another layer of tags.

Prosody  

Chair: Gérard Bailly & Chiu-Yu Tseng  

- Rongna A, Ryoko Hayashi & Tatsuya Kitamura  

**Naturalness on Japanese pronunciation before and after shadowing training and prosody modified stimuli**  

**Abstract**: This study attempts to investigate the change of naturalness impression for the Japanese utterance by Japanese as foreign language learners (JFL) before and after shadowing and repeating trainings and to discuss the crucial prosodic cue for the naturalness judgment. The speech of 8 JFL learners before and after pronunciation training was used, and their durational pattern, pitch pattern, or both of them were replaced with those of the model speech. 52 Japanese native speakers (JNS) assessed the naturalness of these stimuli. The results showed JNS judge the duration AND F0 modified stimuli most natural. In addition, the shadowing trained group tended to get higher scores than the repeating trained group after training. The acoustical analysis of speech material showed a difference of moraic structure and pitch accent between the shadowing and repeating group.
Hansjörg Mixdorff & Murray Munro
Quantifying and evaluating the impact of prosodic differences of foreign-accented English

Abstract: The identification and correction of prosodic deviations still poses a significant challenge for computer-aided language learning. With this ultimate goal in mind, the current study compares utterances by Cantonese speakers of Canadian English with those of native English subjects through both acoustic analysis and perceptual evaluation. We aim to find measurable prosodic differences accounting for the perceptual results. Our outcomes indicate, inter alia, that unstressed syllables are relatively longer compared to stressed ones in the Cantonese corpus than in the Canadian English corpus. Furthermore, the correlations of syllabic durations in utterances of one and the same sentence are much higher for Canadian English subjects than for Cantonese speakers. The latter use a similar range of F0, but produce more and longer pitch-accents than Canadian English speakers. In a perception study we found that applying native durations together with F0 contours to the foreign-accented speech led to significantly improved listener judgments of prosodic goodness. Adjustments to duration alone also tended to yield better ratings, though the effect was not statistically significant. When durations of native English utterances were adjusted to those of Cantonese speakers, significant decrements in ratings were observed.

Hansjörg Mixdorff & Hamurabi Gamboa Rosales
Prosodic chunking of German as a foreign language

Abstract: This study concerns the perception of boundaries and accented syllables by native German subjects as compared to foreign non-speakers and learners of the language at different proficiency levels. To this effect six-syllable sequences excised from a context of three poly-syllabic words of German were presented to participants who had to select the syllables they perceived as accented, as well as the locations of word boundaries. Results show that German native subjects perform well at the word boundary task, but mark correctly less than two thirds of accented syllables. Chinese and Mexican non-learners still detect a considerable number of word boundaries and accented syllables. Learners of German show improvement at the task with growing experience though they often pick legal subword units that do not necessarily form a plausible sequence. Correlation analysis of factors for syllable and boundary selection performed for non-learners and German subjects – as expected – shows considerably different behaviours. Whereas the boundary location does not influence the Germans’ decision on the accent location, Chinese and Mexican non-learners show a preference to mark an accent when the syllable is followed by a word boundary. We also found that the acoustic properties of the syllables had a larger impact on the non-learners’ decisions since they could not operate on linguistic knowledge of German.
• Catherine Lai, Keelan Evanini & Klaus Zechner

Applying rhythm metrics to non-native spontaneous speech

Abstract: This study investigates a variety of rhythm metrics on two corpora of non-native spontaneous speech and compares the non-native distributions to values from a corpus of native speech. Several of the metrics are shown to differentiate well between native and non-native speakers and to also have moderate correlations with English proficiency scores that were assigned to the non-native speech. The metric that had the highest correlation with English proficiency scores (apart from speaking rate) was rPVI{syl} (the raw Pairwise Variability Index for syllables), with $r = -0.43$.

Phonetics / phonology

Chairs: Rodolfo Delmonte & Jacques Koreman

• Chiu-Yu Tseng, Chao-Yu Su & Tanya Visceglia

Underdifferentiation of English lexical stress contrasts by L2 Taiwan speaker

Abstract: Learning the stress patterns of English words presents a challenge for L1 speakers from syllable-timed and/or tone languages. Realization of stress contrasts in previous studies has been measured in a variety of ways. This study adapts and extends PVI, a method generally used to measure duration as a property of speech rhythm, to compare F0 and amplitude contrasts across L1 and L2 production of stressed and unstressed syllables in English multisyllabic words. L1 North American English and L1 Taiwan-Mandarin English speech data were extracted from the AESOP-ILAS corpus. Results of acoustic analysis show that overall, stress contrasts were realized most robustly by L1 English speakers. A general pattern of contrast underdifferentiation was found in L2 speakers with respect to F0, duration and intensity, with the most striking difference found in F0. These results corroborate our earlier findings on L1 Mandarin speakers’ production of on-focus/post-focus contrasts in their realization of English narrow focus. Taken together, these results demonstrate that underdifferentiation of prosodic contrasts at both the lexical and phrase levels is a major prosodic feature of Taiwan English; future research will determine whether it can also be found in the L2 English of other syllable-timed or tone language speakers.

• Mario Carranza

Intermediate phonetic realizations in a Japanese accented L2 Spanish corpus

Abstract: This paper addresses the issue of manual transcription of non native speech in an attempt to establish rule-based strategies for labelling intermediate realizations. The problems of transcribing non canonical realizations of L2 sounds which present shared features of the target (Spanish) and the source language (Japanese) will be considered. We introduce a Japanese accented non native L2 Spanish corpus, and exemplify the use of decision trees in manual transcriptions as a systematic method for dealing with ambiguous realizations. This approach could help a potential error detection system to detect both canonical and erroneous realizations, contributing to the development of CAPT tools.
• Jacques Koreman, Preben Wik, Olaf Husby & Egil Albertsen

Universal contrastive analysis as a learning principle in CAPT

Abstract: Cross-linguistic comparison is good starting point for CAPT. For segmental contrasts, the IPA categories with their claim to phonemic universality enable a structure which ensures communicative effectiveness for learners who manage to perceive and pronounce all phoneme categories in an L2. The CALST system implements contrastive analysis in two types of exercises for phonetic and abstract listening, where users can train the perception of sounds which are not part of the L1 inventory. Since substitutions for unfamiliar sounds depend on L1, the selection of sound contrasts which are trained in the exercises should also depend on L1. We shall argue for a pragmatic approach to the selection of exercises.

• Greg Short, Keikichi Hirose & Nobuaki Minematsu

Automatic recognition of vowel length in Japanese for a CALL system motivated by perceptual experiments

Abstract: Acquisition of the Japanese vowel length contrast can be problematic for non-native speakers. For these speakers, a CALL system which can automatically recognize whether a particular vowel is being pronounced as long or short could be of great benefit for pointing out their errors and issuing corrective feedback. However, a method that can adequately do this has not been proposed yet. Vowel length recognition is made difficult because the vowel length distinction is dependent on the surrounding vowel durations. HMMs, the standard way of recognizing this distinction, do not make use of this information. Other methods have been proposed to recognize this, but they do not appear viable unless knowledge about the length of other vowels is present. Thus, to derive a method, we conduct perceptual experiments to analyze the mechanism of vowel length perception. From this analysis, we develop an automatic recognition algorithm for vowel length that uses support vector machines (SVMs). We tested this method on a speaking rate corpus, native speech, and non-native speech. The method produced recognition results that are overall superior to HMMs and also more robust against speaking rate differences with an average of a 0.83 correct recognition rate for 3 datasets. The error and non-error classification rates on non-native speech for this are 0.86 and 0.84 respectively recognition rate for 3 datasets. The error and non-error classification rates on non-native speech for this are 0.86 and 0.84 respectively.
• Han-Ping Shen, Nobuaki Minematsu, Takehiko Makino, Steven H Weinberger, Teeraphon Pongkittiphan & Chung-Hsien Wu

Speaker-based accented English clustering using a world English archive

Abstract: English is the only language available for global communication. Due to the influence of speakers’ mother tongue, however, those from different regions often have different accents in their pronunciation of English. The ultimate goal of our project is automatic creation of a global pronunciation map of World Englishes on an individual basis, for speakers to use to locate similar English pronunciations. Creating the map mathematically requires a matrix of pronunciation distances among all the speakers considered. Our previous study proposed a good algorithm for that purpose [1], where, using reference pronunciation distances calculated from labeled data, a pronunciation distance predictor was trained and built for unlabeled data. Due to space limit in [1], the procedure for calculating reference distances was not described in detail. Then in this paper, detailed descriptions are given and 498 world-wide native and non-native speakers in the Speech Accent Archive are clustered using the reference distances. Results show high accentual validity of the reference inter-speaker distances.

• Hyejin Hong, Sunhee Kim & Minhwa Chung

A corpus-based analysis of Korean segments produced by Japanese learners

Abstract: This paper examines variations of Korean segments produced by Japanese learners of Korean. For corpus-based statistical analysis, we have used Korean read speech corpus produced by Japanese learners. Contrastive analysis of the target language and the source language is performed to provide information for interpreting the results of corpus analysis. Segmental variations are analyzed by aligning canonical phonetic transcriptions with auditory phonetic transcriptions of the corpus. The results show that (1) Japanese learners tend to demonstrate substitutions due to differences in the phonetic systems of the two languages; and (2) they are likely to omit a consonant or insert a vowel to deal with the different syllable structures. These results with detailed statistical data are useful for designing a computer-assisted pronunciation training and assessment system for Japanese learners of Korean.
Authors list

A, Rongna
Ahmed, Imran
Albertsen, Egil
Bacalu, Ciprian
Bang, Jeesoo
Beckman, Mary
Bernstein, Jared
Bodnar, Stephen
Cai, Carrie
Calado, António
Cameran, Christel-Joy
Carranza, Mario
Chen, Lei
Chung, Minhwa
Claus, Felix
Colpaert, Jozef
Cucchiarini, Catia
Davis, Elizabeth
de Vries, Bart Penning
de Vries, Nic J.
de Wet, Febe
Delmonte, Rodolfo
Deutekom, Jan
Dias, Miguel Sales
Drozdova, Polina
Eskenazi, Maxine
Evanini, Keelan
Gamboa Rosales,
Hamurabi
Glass, James
Hain, Horst-Udo
Hämäläinen, Annika
Harbusch, Karin
Härtel, Johannes
Hashimoto, Hiroya
Hayashi, Ryoko
Hirano, Hiroko
Hirose, Keikichi
Hoffmann, Rüdiger
Højfeldt Rasmussen, Morten
Hong, Hyejin
Husby, Olaf
Júdice, Ana
Kang, Sechun
Kim, Hae-Ri
Kim, Sunhee
Kitamura, Tatsuya
Kopparapu, Sunil
Koreman, Jacques
Kweon, Soo-Ok
Lai, Catherine
Lee, Lin-Shan
Lee, Kyusong
Lee, Gary Geunbae
Lee, Lin-Shan
Lee, Ann
Litman, Diane
Makino, Takehiko
Meng, Helen
Miller, Robert
Minematsu, Nobuaki
Mixdorff, Hansjörg
Miyakoda, Haruko
Munro, Murray
Nakagawa, Chieko
Nakamura, Ibuki
Nakamura, Noriko
Neumeyer, Kayla
Pandharipande, Meghna
Patil, Vaishali
Petrick, Rico
Pinto, Fernando Miguel
Pongkittiphan, Teeraphon
Qian, Xiaojun
Rao, Preeti
Rayner, Manny
Rodrigues, Silvia
Saz, Oscar
Schultz, Katharyn
Seneff, Stephanie
Shen, Han-Ping
Short, Greg
Silva, Sandra Morgado
Strik, Helmer
Su, Pei-Hao
Su, Ya-Yunn
Su, Chao-Yu
Suzuki, Masayuki
Tagawa, Yukinori
Tan, Zheng-Hua
Todic, Ognjen
Tseng, Chiu-Yu
Tsourakis, Nikos
van de Craats, Ineke
van Hout, Roeland
Visceglia, Tanya
Wang, Hao
Weinberger, Steven H
Wik, Preben
Wu, Chung-Hsien
Xiong, Wenting
Yu, Tien-Han
Zechner, Klaus
Zhao, Liang
Sponsors

The SLaTE-2013 local organizing committee is happy to acknowledge the financial support of the following sponsors.

![Sponsors logos](image-url)