
**SLaTE 2009**

ISCA International Workshop on Speech and Language Technology in Education

Wroxall Abbey Estate, Warwickshire, England

3rd – 5th September 2009

# Technical Programme and Abstracts

International Speech
Communication
Association

UNIVERSITY OF
BIRMINGHAM

## SLaTE

As Chair of the ISCA Special Interest group, SLaTE, I would like to welcome you to the SLaTE 2009 Workshop. The SLaTE SIG is interested in all research that concerns the use of language technologies for education applications, especially those for language learning. There are many domains that are covered by SLaTE, such as pronunciation error detection and dialogue games, as you will see over the next few days.

The first meeting on this topic was organized by the researchers at KTH and held in Marhollmen Sweden in 1998. That very successful meeting was followed by meetings in Aberdeen and in Venice. SLaTE itself was created by a group of interested researchers at the 2006 Interspeech conference in Pittsburgh, USA. And the first SLaTE SIG workshop was held in Farmington PA, USA in 2007. The present meeting at Wroxhall Abbey, UK carries on the tradition of assembling to hear high quality research, to attend demonstrations of software and to have interesting discussions with colleagues from around the world.

Several of the presentations from the 2007 meeting have resulted in the publication of a special issue of Speech Communication (October 2009). This issue had the highest number of proposed papers of any issue in Speech Communication history.

I would like to especially thank Martin Russell for organizing this wonderful workshop! It is well-planned and is full of interesting events, papers and demos. I hope you all enjoy this workshop and find many new ideas to take home with you.

Maxine Eskenazi
Chair, ISCA SIG SLaTE

## SLaTE 2009

Wecome to the ISCA SLaTE 2009 workshop. The venue for the workshop is Wroxall Abbey Estate, in Warwickshire, England. This was chosen for a number of reasons: It's small, and SLaTE 2009 will be the only major event while we are there, it is unique and steeped in history, it is remote but relatively accessible, and it is close to a number of interesting places, including Warwick and Stratford-upon-Avon (visits to both of these are included in the social programme). It is certainly not a modern hotel. All of the guest rooms are different with their own unique characters, and the conference facilities are more intimate than you would expect in a more modern venue. I hope you like it!

Forty five papers were submitted to SLaTE 2009. This is a significant increase compared with SLaTE 2007 and indicates that our area is growing. Each submission was evaluated by at least three reviewers. Overall the quality was judged to be very high, and we accepted 37 papers. Of these, 19 will be presented as talks in the 5 oral sessions, and 18 as posters in 3 poster sessions. We also have 14 demonstrations in 2 sessions. This has resulted in a busy programme, and it has been necessary to include evening sessions on the Thursday and Friday, between the social events and dinner.

Where possible the papers have been grouped according to themes, however, due to the relatively small number of papers and the travel arrangements of some of their presenters it has not always been possible to adhere tightly to these themes.

I would like to thank the SLaTE committee for their support, and the reviewers for reading and assessing their allocated papers.

I hope that you enjoy the workshop!

Martin Russell
Organiser, SLaTE 2009

**SLaTE Committee**

| | |
|---|---|
| Abeer Alwan | UCLA |
| Jared Bernstein | Ordinate Corp. |
| Rodolfo Delmonte | University of Venice |
| Maxine Eskenazi | Carnegie Mellon University |
| Björn Granström | KTH |
| Valerie Hazan | University College, London |
| Diane Litman | University of Pittsburgh |
| Dominic Massaro | University of California – Santa Cruz |
| Nobuaki Minematsu | University of Tokyo |
| Patti Price | PPrice.com |
| Martin Russell | University of Birmingham |
| Stephanie Seneff | MIT |
| Helmer Strik | Radboud University |

**SLaTE 2009 Scientific Review Committee**

The organisers would like to thank the following individuals who took part in the review of papers submitted to SLaTE 2009.

| | |
|---|---|
| Abeer Alwan | Gregory Aist |
| Anton Batliner | Kay Berkling |
| Jared Bernstein | Alan Black |
| Lei Chen | Catia Cucchiarini |
| Rodolfo Delmonte | Ryan Downey |
| Maxine Eskenazi | Horacio Franco |
| Björn Granström | Diego Giuliani |
| Valerie Hazan | W Lewis Johnson |
| Diane Litman | Nuno Mamede |
| Dominic Massaro | Nobuaki Minematsu |
| Mari Ostendorf | Patti Price |
| Martin Russell | Stephanie Seneff |
| Helmer Strik | Louis ten Bosch |
| Joseph Tepperman | Isabel Trancoso |
| Hugo Van hamme | |

## SLaTE 2009 - Outline Technical and Social Programme

| | Thursday 3rd | Friday 4th | Saturday 5th | Sunday 6th |
|---|---|---|---|---|
| 7.30-8.30am | | Breakfast | Breakfast [plus SLaTE Committee meeting] | Breakfast |
| 9.00-10.30am | | Oral session 3: Creation & Assessment Of Content For SLaTE | Poster session 3: Topics in SLaTE (1) | Coach departs for Brighton 11.30am |
| 10.30-11.00am | | *Coffee* | *Coffee* | |
| 11.00-12.45pm | | Poster session 2: Pronunciation Training & Assessment (2) | Demo session 2 | |
| 12.45-1.45pm | *Lunch* | *Lunch* | *Lunch* | |
| 1.45-3.00pm | Welcome<br>Oral session 1: Virtual Tutors & Games | Oral session 4: Pronunciation Training & Assessment (3) | Oral session 5: Topics in SLaTE (2) | |
| 3.00-3.30pm | | | SLaTE general assembly | |
| 3.30-4.00pm | *Tea* (4.10 – 4:25pm) | *Tea* (3.30-4.00pm) | *Tea* (3.30-4.00pm) | |
| 4.00-5.30pm | Poster session 1: Language Proficiency Assessment | Excursion to Warwick Castle | Excursion to Stratford-upon-Avon | |
| 5.30-6.30pm | Guided tours of Wroxall Abbey Estate | | | |
| 6.30-8.00pm | Oral session 2: Pronunciation Training & Assessment (1) | Demo session 1 | | |
| 8.00pm | Dinner | Dinner | Dinner | |

# Social Programme

**Wroxall Abbey Estate History Talk and Tours - Thursday 3rd September, 5.30 – 6.30pm**

Tony Carr, one of the ministers responsible for Wroxall Abbey Estate Chapel, will give a presentation on the history of Wroxall Abbey Estate. The talk will take place in the Chapel and will be followed by group tours of the estate.

**Visit to Warwick Castle – Friday 4th September, 3.30 – 6.30pm**

An executive coach and English speaking guide would meet the guests at Wroxall Abbey and depart for a visit to England's finest Medieval Castle

This breathtaking monument rises majestically above the banks of the River Avon. First fortified by William the Conqueror in 1068, it spans nine centuries of war and peace, turmoil and calm, upheaval and progress. Down through the years, the Castle has echoed to the sound of feast days, of lavish entertainments, of high society dinners. The guests will enjoy a guided tour which includes the magnificent Great Hall and State Apartments, with Madame Tussaud's "Royal Weekend House Party" and the Kingmaker attraction.

**Visit to Stratford-on-Avon – Saturday 5th September, 4.00 – 8.00pm**

An executive coach and English speaking guide would meet the guests at Wroxall Abbey and depart for a guided tour of Stratford-upon-Avon where they would be met by two English speaking guides for a walking tour of the town centre.

A small, old Elizabethan town on the gently-flowing River Avon, Stratford is famous as the birthplace of the Immortal Bard, William Shakespeare, born there in 1564. The charming town centre boasts an extensive array of 16th and 17th century buildings, with rows of half-timbered houses, including Shakespeare's Birthplace on Henley Street. Stratford is also well known as a stage for Shakespeare's works and is home to the Royal Shakespeare Company.

The experienced Guides will point out the many places of interest, including the five Shakespeare Houses, the three Royal Shakespeare Company theatres, the Church where the Bard is buried, Shakespeare's School, Harvard House, the Guild Chapel and other historic buildings of interest.

The guides would leave the guests after the tour and the coach would return them to Wroxall Abbey at 20:00

## THURSDAY 3RD SEPTEMBER 2009

**Thu 3/9/09 14:30 Welcome**

Martin Russell

**Thu 3/9/09 14:40 – 16:10 Oral 1**

**VIRTUAL TUTORS AND GAMES**

Chair: Mari Ostendorf

14:40 Chair's Introduction

14:50 [O1.1] *Alicia Sagae, Baylor Wetzel, Andre Valente, W. Lewis Johnson*, "Culture-Driven Response Strategies for Virtual Human Behavior in Training Systems" [PDF]

15:10 [O1.2] *Brandon Yoshimoto, Ian McGraw, Stephanie Seneff*, "Rainbow Rummy: A Web-based Game for Vocabulary Acquisition using Computer-directed Speech" [PDF]

15:30 [O1.3] *Preben Wik, Rebecca Hincks, Julia Hirschberg* , "Responses to Ville: A virtual language teacher for Swedish" [PDF]

15:50 [01.4] *Joost van Doremalen, Helmer Strik, Catia Cucchiarini*, "Utterance Verification in Language Learning Applications" [PDF]

**16:10 – 16:25 Tea and Coffee**

**Thu 3/9/09 16:25 – 17:50 Poster 1**

**LANGUAGE PROFICIENCY ASSESSMENT**

Chair: Helmer Strik

16:00 – 16:20 Introduction to poster session 1

16:20 – 17:30 Posters

[P1.1] *Jared Bernstein, Masanori Suzuki, Jian Cheng, & Ulrike Pado*, "Evaluating Diglossic Aspects of an Automated Test of Spoken Modern Standard Arabic" [PDF]

[P1.2] *Ingunn Amdal, Magne H. Johnsen, Eivind Versvik*, "Automatic evaluation of quantity contrast in non-native Norwegian speech" [PDF]

[P1.3] *Klaus Zechner*, "What did they actually say? Agreement and Disagreement among Transcribers of Non-Native Spontaneous Speech Responses in an English Proficiency Test" [PDF]

[P1.4] *Pieter Müller, Febe de Wet, Christa van der Walt & Thomas Niesler*, "Automatically assessing the oral proficiency of proficient L2 speakers" [PDF]

[P1.5] *Jian Cheng, Brent Townshend*, "A Rule-Based Language Model for Reading Recognition" [PDF]

[P1.6] *Dean Luo, Nobuaki Minematsu*, *Yutaka Yamauchi and Keikichi Hirose*, "Analysis and Comparison of Automatic Language Proficiency Assessment between Shadowed Sentences and Read Sentences" [PDF]

**Thu 3/9/09 17:50 – 18:40**

**Talk on the history of Wroxall Abbey Estate**

**Thu 3/9/09 18:40 – 20:00 Oral 2**

**PRONUNCIATION TRAINING & ASSESSMENT (1)**

Chair: Björn Granström

18:40 [O2.1] *Florian Hönig, Anton Batliner, Karl Weilhammer, Elmar Nöth*, "Islands of Failure: Employing word accent information for pronunciation quality assessment of English L2 learners" [PDF]

19:00 [O2.2] *Alissa M. Harrison, Wai-kit Lo, Xiao-jun Qian, Helen Meng*,

"Implementation of an Extended Recognition Network for Mispronunciation Detection and Diagnosis in Computer-Assisted Pronunciation Training" [PDF]

19:20 [O2.3] *Sandra Kanters*, *Catia Cucchiarini*, *Helmer Strik*, "The Goodness of Pronunciation Algorithm: a Detailed Performance Study" [PDF]

19:40 [O2.4] *Preben Wik, David Lucas Escribano*, "Say 'Aaaaa' Interactive Vowel Practice for Second Language Learning" [PDF]

**Thu 3/9/09 20:00 Dinner at Wroxall Abbey**

## FRIDAY 4TH SEPTEMBER 2009

**Fri 4/9/09 09:00 – 10:30 Oral 3**

**CREATION & ASSESSMENT OF CONTENT FOR SLaTE**

Chair: Martin Russell

09:00 Chair's introduction

09:10 [O3.1] *Yushi Xu, Anna Goldie, Stephanie Seneff*, "Automatic Question Generation and Answer Judging: A Q&A Games for Language Learning" [PDF]

09:30 [O3.2] *Julie Medero, Mari Ostendorf*, "Analysis of Vocabulary Difficulty Using Wiktionary" [PDF]

09:50 [O3.3] *Juan Pino, Maxine Eskenazi*, "Semi-automatic Generation of Cloze Question Distractors: Effect of Students' L1" [PDF]

10:10 [O3.4] *Luís Marujo, José Lopes, Nuno Mamede, Isabel Trancoso, Juan Pino, Maxine Eskenazi, Jorge Baptista, Céu Viana*, "Porting REAP to European Portuguese" [PDF]

**10:30-11:00 Tea and Coffee**

**Fri 4/9/09 11:00 – 12:45 Poster 2**

**PRONUNCIATION TRAINING & ASSESSMENT (2)**

Chair: Isabel Trancoso

11:00 – 11:20 Introduction to poster session 2

11:20 – 13:00 Posters

[P2.1] *Helmer Strik, Frederik Cornillie, Jozef Colpaert, Joost van Doremalen, Catia Cucchiarini*, "Developing a CALL System for Practicing Oral Proficiency: How to Design for Speech Technology, Pedagogy and Learners" [PDF]

[P2.2] *Hansjörg Mixdorff, Daniel Külls, Hussein Hussein, Gong Shu, Hu Guoping, Wei Si*, "Towards a Computer-aided Pronunciation Training System for German Learners of Mandarin" [PDF]

[P2.3] *William R. Rodríguez, Eduardo Lleida*, "Formant Estimation in Children's Speech and its application for a Spanish Speech Therapy Tool" [PDF]

[P2.4] *Natalia Cylwik, Agnieszka Wagner, Grażyna Demenko*, "The EURONOUNCE corpus of non-native Polish for ASR-based Pronunciation Tutoring System" [PDF]

[P2.5] *Hiroyuki Obari, Hiroaki Kojima, Machi Okumura, Masahiro Yoshikawa, Shuichi Itahashi*, "Investigating the Effectiveness of Prontest to Train English Proficiency"[PDF]

[P2.6] *Oscar Saz, Victoria Rodríguez, Eduardo Lleida, W.-R. Rodríguez, C. Vaquero*, "An Experience with a Spanish Second Language Learning Tool in a Multilingual Environment" [PDF]

**Fri 4/9/09 14:00 – 15:30 Oral 4**

**PRONUNCIATION TRAINING & ASSESSMENT (3)**

Chair: Maxine Eskenazi

14:00 Chair's introduction

14:10 [O4.1] *Lei Chen*, "Audio Quality Issue for Automatic Speech Assessment" [PDF]

14:30 [O4.2] *Shizhen Wang, Patti Price, Yi-Hui Lee and Abeer Alwan*, "Measuring Children's Phonemic Awareness through Blending Tasks" [PDF]

14:50 [O4.3] *Minh Duong Jack Mostow*, "Detecting Prosody Improvement in Oral Rereading" [PDF]

15:10 [O4.4] *Alexander Gruenstein, Ian McGraw, Andrew Sutherland*, "A Self-Transcribing Speech Corpus: Collecting Continuous Speech with an Online Educational Game" [PDF]

**15:30 – 15:45 Tea and Coffee**

**Fri 4/9/09 15:45 – 18:30**

**Excursion to Warwick Castle**

15:45 Coach leaves for Warwick Castle

16:00 – 18:00 Tours of Warwick Castle

18:30 return to Wroxall Abbey

**Fri 4/9/09 18:30 – 20:00 Demo 1**

[D1.1] **"**Japanese CALL system based on Dynamic Question Generation and Error Prediction for ASR", *Tatsuya Kawahara, Hongcui Wang, Yasushi Tsubota and Masatake Dantsuji,* Kyoto University, Japan

[D1.2] "Development of a CALL System to Enhance ESL/EFL Learners' Skills of Shadowing and Reading Aloud**",** *Dean Luo[1], Nobuaki Minematsu[1], Yutaka Yamauchi[2] , [1]The University of Tokyo [2]Tokyo International University*

[D1.3] "Computer Aided Pronunciation Training (CAPT) System "AZAR"", *Michael Beilig*

[D1.4] "Spoken dialog systems for learning foreign language communication skills", *Andre Valente and W. Lewis Johnson,* Alelo Inc

[D1.5] "Langofone - Language learning in your pocket", *Preben Wik,* Center for Speech Technology (CTT), Dept. Speech, Music and Hearing, KTH, Stockholm, Sweden

[D1.6] "COMUNICA: MULTILEVEL TOOLS FOR SPANISH CALL", *Oscar Saz, W.-Ricardo Rodríguez, Eduardo Lleida, Carlos Vaquero,* Communications Technology Group (GTC), Aragón Institute for Engineering Research (I3A), University of Zaragoza, Zaragoza, Spain

[D1.7] "Pronunciation Evaluation System for Oral English Testing and Oral Chinese Testing", *Si Wei*

20:00 Dinner at Wroxall Abbey

## SATURDAY 5TH SEPTEMBER 2009

**Sat 5/9/09 09:00 – 10:30 Poster 3**

**TOPICS IN SLaTE (1)**

Chair: Nuno Mamende

09:00 – 09:20 Introduction to poster session 3

09:20 – 10:30 Posters

[P3.1] *Grażyna Demenko, Agnieszka Wagner, Natalia Cylwik and Oliver Jokisch*, "An Audiovisual Feedback System for Acquiring L2 Pronunciation and L2 Prosody" [PDF]

[P3.2] *Rebecca Hincks, Jens Edlund*, "Using speech technology to promote increased pitch variation in oral presentations" [PDF]

[P3.3] *Zöe Handley, Mike Sharples, Dave Moore*, "Training Novel Phonemic

Contrasts: A Comparison of Identification and Oddity Discrimination Training" [PDF]

[P3.4] *Angela M. Wigmore, Gordon J.A. Hunter, Eckhard Pflügel and James Denholm-Price,* "TalkMaths : A Speech User Interface for Dictating Mathematical Expressions into Electronic Documents" [PDF]

[P3.5] *Liu Liu, Jack Mostow, and Gregory Aist*, "Automated Generation of Example Contexts for Helping Children Learn Vocabulary" [PDF]

[P3.6] *Khe Chai Sim*, "Improving Phone Verification Using State-level Posterior Features and Support Vector Machine for Automatic Mispronunciation Detection" [PDF]

**10:30-11:00 Tea and Coffee**

**Sat 5/9/09 11:00 – 12:45 Demo 2**

[D2.1] "Structure-based pronunciation assessment", *Nobuaki Minematsu and Masayuki Suzuki*

[D2.2] "Voice Race and Voice Scatter: Online Educational Games for Collecting Orthographically-Labeled Speech Data", *Alexander Gruenstein, Ian McGraw, and Andrew Sutherland*

[D2.3] **"**REAP.PT, a tutoring system for teaching Portuguese", *L. Marujo, J. Lopes, N. Mamede, I. Trancoso, J. Pino, M. Eskenazi, J. Baptista, C. Viana*

[D2.4] "Virtual Chinese Tutor (VCT) - A Chinese Language Pronunciation Learning Software", *Yow-Bang Wang*[1,2]*, Hsin-Min Wang*[1]*, Lin-Shan Lee*[1,2], [1]Institute of Information Science, Academia Sinica, [2]Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan, R.O.C.

[D2.5] "Automatic assessment of non-native prosody", *Florian Hönig, Anton Batliner, Karl Weilhammer, Elmar Nöth*

[D2.6] "A Spoken Dialog System for Learners of English", *Karl Weilhammer, Catharine Oertel, Robin Siegemund, Ricardo Sá, Anton Batliner, Florian Hönig und Elmar Nöth*

[D2.7] "CLIMB LEVEL 4 – teaching English for aviation safety", *Maxine Eskenazi, Gary Pelton,* Carnegie Speech Company

**Sat 5/9/09 12.45 – 13.45 Lunch**

**Sat 5/9/09 13:45 – 15:00 Oral 5**

**TOPICS IN SLaTE (2)**

Chair: Stephanie Seneff

13:45 Chair's Introduction

14:00 [O5.1] *Masayuki Suzuki, Luo Dean, Nobuaki Minematsu, Keikichi Hirose*, "Improved Structure-based Automatic Estimation of Pronunciation Proficiency" [PDF]

14:20 [O5.2] *Gregory Aist and Jack Mostow*, "Predictable and Educational Spoken Dialogues: Pilot Results" [PDF]

14:40 [05.3] *John Ingram, Hansjörg Mixdorff and Nahyun Kwon*, "Voice morphing and the manipulation of intra-speaker and cross-speaker phonetic variation to create foreign accent continua: A perceptual study" [PDF]

**Sat 5/9/09 15:00**

**SLaTE General Assembly**

**15:30 – 16:00 Tea and Coffee**

**Sat 5/9/09 16:00**

**Walking tour of Stratford-upon-Avon**

20.00 Return to Wroxall Abbey

# ABSTRACTS

## THURSDAY 3rd SEPTEMBER 2009

**Thu 3/9/09 14:00 – 15:30 Oral 1**

**VIRTUAL TUTORS AND GAMES**

### O1.1 Culture-Driven Response Strategies for Virtual Human Behavior in Training Systems

*Alicia Sagae, Baylor Wetzel, Andre Valente, W. Lewis Johnson*

Alelo, Inc Los Angeles, CA, USA

{asagae, bwetzel, avalente, ljohnson}@alelo.com

**Abstract**

In this work we introduce a set of response strategies that capture the effect of cultural norms on the behavior of conversational agents in language and culture training systems. Response strategies cover behavior such as deception, vagueness, and distraction. Starting from a vocabulary of strategies derived from the literature on compliance and cooperation, we compare this explicit list to evidence of implicit strategizing in hand-authored dialogs captured from a serious game system. As a result, we find that the strategy representation codifies a layer of communicative information that seems to be necessary for believable dialog in the context of teaching cultural communication skills. We also explore how dialog strategies can be explicitly authored and tested, presenting results from an implemented prototype.

### O1.2 Rainbow Rummy: A Web-based Game for Vocabulary Acquisition using Computer-directed Speech

*Brandon Yoshimoto, Ian McGraw, Stephanie Seneff*

MIT Computer Science and Artificial Intelligence Laboratory, 32 Vassar Street, Cambridge, MA 02139

yoshimoto@alum.mit.edu, imcgraw@csail.mit.edu, seneff@csail.mit.edu

**Abstract**

This paper describes a new on-line game we have developed which allows learners of Chinese or English to practice speaking in a communicative setting. Game play resembles gin rummy or Mah Jong, and is intended to be sufficiently engaging to invite persistent replay. Students compete in a social game against other students at remote settings, or they can play against a robotic partner. A user study was conducted on 16 students of Chinese, to assess whether a configuration that utilizes speech recognition is as effective for learning vocabulary as a configuration that only requires the student to listen. Results show that the vocabulary learning gains averaged across subjects were greater following use of the speech-enabled version of the game compared to the listening-only version.

### O1.3 Responses to Ville: A virtual language teacher for Swedish

*Preben Wik[1], Rebecca Hincks[2], Julia Hirschberg[3]*

Centre for Speech Technology, CSC, KTH, Sweden[1], The Unit for Language and Communication, CSC, KTH, Sweden[2], Department of Computer Science, Columbia University, USA[3]

preben@speech.kth.se, hincks@speech.kth.se, julia@cs.columbia.edu

**Abstract**

A series of novel capabilities have been designed to extend the repertoire of Ville, a virtual language teacher for Swedish, created at the Centre for Speech technology at KTH. These capabilities were tested by twenty-seven language students at KTH. This paper reports on qualitative surveys and quantitative performance from these sessions which suggest some general lessons for automated language training.

### O1.4 Utterance Verification in Language Learning Applications

*Joost van Doremalen, Helmer Strik, Catia Cucchiarini*

Department of Linguistics, Radboud University, Nijmegen, The Netherlands

fj.vandoremalen,h.strik,c.cucchiarinig@let.ru.nl

**Abstract**

A CALL system for oral proficiency is being developed in which constrained responses are elicited from L2 learners. In the first phase the best matching utterance is selected from a predefined list of possible responses. Since errors may occur and giving feedback on the basis of incorrectly recognized utterances is confusing, we verify the correctness of the utterance in the second phase. In the current paper we focus on the utterance verification process. Combining duration related features with a likelihood ratio (LR) yielded an equal error rate (EER) of 10.3%, which was significantly better than the EER for LR alone, 14.4%, and the EER for the duration-related features, 25.3%

Index Terms: utterance verification, non-native speech processing, computer-assisted language learning

## Thu 3/9/09 16:00 – 17:30 Poster 1

## LANGUAGE PROFICIENCY ASSESSMENT

### P1.1 Evaluating Diglossic Aspects of an Automated Test of Spoken Modern Standard Arabic

*Jared Bernstein, Masanori Suzuki, Jian Cheng, & Ulrike Pado.*

Pearson Knowledge Technologies, 299 S. California Ave., Palo Alto, California, 94306 U.S.A.

Jared.Bernstein@Pearson.com

**Abstract**

A fully automatic test of facility with spoken Modern Standard Arabic (MSA) was developed and evaluated. The paper notes the diglossic situation of MSA (where colloquial and formal languages are quite distinct), and presents the structure and scoring of the test. Evaluation of the reliability and validity of the test is described, with added analyses that compare not just learners and native speakers, but also educated and uneducated speakers of the formal dialect. Results suggest scores from this commercial test are suitable in selecting MSA speakers.

### P1.2 Automatic evaluation of quantity contrast in non-native Norwegian speech

*Ingunn Amdal, Magne H. Johnsen, Eivind Versvik*

Department of Electronics and Telecommunications, University of Science and Technology, Trondheim, Norway

{ingunn.amdal,mhj}@iet.ntnu.no

**Abstract**

Computer assisted language learning (CAPT) has been shown to be effective for learning non-natives pronunciation details of a new language. No automatic pronunciation evaluation system exists for non-native Norwegian. We present initial experiments on the Norwegian quantity contrast between short and long vowels. A database of native and non-native speakers was recorded for training and test respectively. We have used a set of acoustic-phonetic features and combined them in a classifier based on linear discriminant analysis (LDA). The resulting classification rate was 92.3% compared with a human rating. As expected, vowel duration was the most important feature, whereas vowel spectral content contributed insignificantly. The achieved classification rate is promising with respect to making a useful Norwegian CAPT for quantity.

## P1.3 What did they actually say? Agreement and Disagreement among Transcribers of Non-Native Spontaneous Speech Responses in an English Proficiency Test

*Klaus Zechner*

Educational Testing Service, Princeton, NJ, USA

kzechner@ets.org

**Abstract**

This paper presents an analysis of differences in human transcriptions of non-native spontaneous speech on a word level, collected in the context of an English Proficiency Test. While transcribers of native speech typically agree at a very high level (5% word error rate or less), this study finds substantially higher disagreement rates between transcribers of non-native speech (10%-34% word error rate). We show how transcription disagreements are negatively correlated to the length of utterances (fewer contexts) and to human scores (impact of lower speaker proficiency) and also seem to be affected by the audio quality of the recordings. We also demonstrate how a novel multi-stage transcription procedure using selection and ranking of transcription alternatives by peers can achieve a higher quality gold standard that approaches the quality of native speech transcription.

## P1.4 Automatically assessing the oral proficiency of proficient L2 speakers

*Pieter Müller[1], Febe de Wet[2, 4], Christa van der Walt[3] & Thomas Niesler[1]*

[1]Department of Electrical and Electronic Engineering, [2]Centre for Language and Speech Technology (SU-CLaST), [3]Department of Curriculum Studies, Stellenbosch University, South Africa, [4]HLT Research Group, CSIR Meraka Institute, South Africa.

pfdevmuller@dsp.sun.ac.za, {fdw,cvdwalt,trn}@sun.ac.za

**Abstract**

We consider the automatic assessment of oral proficiency for advanced second language speakers. A spoken dialogue system is used to guide students through a reading and a repeating exercise and to record their responses. Automatically-derived indicators of proficiency that have proved successful in other studies are calculated from their speech and compared with human ratings of the same data. It is found that, in contrast to the findings of other researchers, posterior scores correlate poorly with human assessments of the reading exercise. Furthermore, the repeating exercise is found both to be more challenging and to provide a better means of automatic assessment than the reading exercise for our test population.

### P1.5 A Rule-Based Language Model for Reading Recognition

*Jian Cheng, Brent Townshend*

Knowledge Technologies, Pearson, 299 S. California Ave, Palo Alto, California 94306, USA

jian.cheng@pearson.com

**Abstract**

Systems for assessing and tutoring reading skills place unique requirements on underlying ASR technologies. Most responses to a "read out loud" task can be handled with a low perplexity language model, but the educational setting of the task calls for diagnostic measures beyond plain accuracy. Pearson developed an automatic assessment of oral reading fluency that was administered in the field to a large, diverse sample of American adults. Traditional N-gram methods for language modeling are not optimal for the special domain of reading tests because N-grams need too much data and do not produce as accurate recognition. An efficient rule-based language model implemented a set of linguistic rules learned from an archival body of transcriptions, using only the text of the new passage and no passage-specific training data. Results from operational data indicate that this rule-based language model can improve the accuracy of test results and produce useful diagnostic information.

### P1.6 Analysis and Comparison of Automatic Language Proficiency Assessment between Shadowed Sentences and Read Sentences

*Dean Luo*[1]*, Nobuaki Minematsu*[1]*, Yutaka Yamauchi*[2] *and Keikichi Hirose*[1]

[1]The University of Tokyo

[2]Tokyo International University

dean@gavo.t.u-tokyo.ac.jp

**Abstract**

In this paper, we investigate automatic language proficiency assessment from learners' utterances generated through shadowing and reading aloud. By increasing the degrees of difficulty of learners' tasks for each practice, we examine how the automatic scores, the conventional GOP and proposed F-GOP, change according to the cognitive loads posed on learners. We also investigate the effect and side-effect of MLLR (Maximum Likelihood Linear Regression) adaptation on shadowing and reading aloud. Experimental results show that shadowing can better reflect the learners' true proficiency than reading aloud. Global MLLR adaptation can improve the evaluation performances on reading aloud more significantly than shadowing. But the performance is still better in shadowing. Finally we show that, by selecting native utterances of adequate semantic difficulty, the evaluation performance by shadowing is even improved.

## Thu 3/9/09 18:30 – 20:00 Oral 2

## PRONUNCIATION TRAINING & ASSESSMENT (1)

### O2.1 Islands of Failure: Employing word accent information for pronunciation quality assessment of English L2 learners

*Florian Hönig*[1]*, Anton Batliner*[1]*, Karl Weilhammer*[2]*, Elmar Nöth*[1]

[1] Chair of Pattern Recognition, Department of Computer Science, Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen, Germany

[2] digital publishing, München, Germany

**Abstract**

So far, applied research aiming at computer-assisted pronunciation training has normally concentrated on segmental aspects. Here, we present a database with realizations of nonnative English speakers with German, French, Spanish, and Italian as native language. We concentrate on the acoustic-prosodic modelling of word accent position and use a large prosodic feature vector to automatically recognize erroneous word accent positions produced by non-native English speakers.

## O2.2 Implementation of an Extended Recognition Network for Mispronunciation Detection and Diagnosis in Computer-Assisted Pronunciation Training

*Alissa M. HARRISON, Wai-kit LO, Xiao-jun QIAN, Helen MENG*

The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

{alissa, wklo, xjqian, hmmeng}@se.cuhk.edu.hk

**Abstract**

This paper presents recent extensions to our ongoing effort in developing speech recognition for automatic mispronunciation detection and diagnosis in the interlanguage of Chinese learners of English. We have developed a set of context-sensitive phonological rules based on cross-language (Cantonese versus English) analysis which has also been validated against common mispronunciations observed from the learners interlanguage. These rules are represented as finite state transducers which can generate an extended recognition network (ERN) based on arbitrary canonical pronunciations. The ERN includes not only standard English pronunciations but also common mispronunciations of learners. Recognition with the ERNenables the speech recognizer to phonetically transcribe the learner's input speech. This transcription can be compared with the canonical pronunciations to identify the location(s) and type(s) of phonetic differences, thus facilitating mispronunciation detection and diagnoses. We have developed a prototype implementation known as the CHELSEA system and have validated the approach based on a new, annotated test set of 600 utterances recorded from 100 Cantonese learners of English. The approach achieves a false rejection rate (i.e. system identifies a phone as incorrect when it is actually correctly pronounced) of 13.6%; as well as a false acceptance rate (i.e. system identifies a phone as correct when it is actually mispronounced) of 44.7%. Among the detected errors, the system can correctly diagnose 54.8% of the mispronunciations.

## O2.3 The Goodness of Pronunciation Algorithm: a Detailed Performance Study

*Sandra Kanters*[1], *Catia Cucchiarini*[2], *Helmer Strik*[2]

[1]Customer Contact Solutions, Logica, The Netherlands

[2]Department of Linguistics, Radboud University Nijmegen, The Netherlands

sandra.kanters@logica.com, [c.cucchiarini|h.strik]@let.ru.nl

**Abstract**

An inventory was compiled of pronunciation errors frequently made by foreigners speaking Dutch. On the basis of this inventory artificial errors were created in a native development corpus, which in turn were used to optimize thresholds for the Goodness of Pronunciation (GOP) algorithm. In the current study the GOP algorithm is evaluated in three different ways: (1) using a native test corpus with artificial errors which reflect errors frequently made by non-natives, (2) within an actual application

used by non-natives for practicing pronunciation, and (3) post-hoc, using the recorded interactions of the pronunciation training application, to determine what the performance of the algorithm would have been if optimal speaker and phone specific thresholds had been used. The results show that the performance of the GOP algorithm was satisfactory and that the procedure by which thresholds were determined by simulating realistic pronunciation errors was appropriate, because performance on the artificially introduced errors closely approximated performance on real data. This finding is particularly welcome if we consider that, in general, paucity of data is a common problem in this kind of research. Furthermore, it appeared that post-hoc threshold optimization only led to a slight increase in performance.

**Index Terms**: Goodness of Pronunciation (GOP), pronunciation error detection, Computer Assisted Pronunciation Training (CAPT)

### O2.4 Say 'Aaaaa' Interactive Vowel Practice for Second Language Learning

*Preben Wik, David Lucas Escribano*

Department of Speech Music and Hearing, KTH, Stockholm, Sweden

preben@speech.kth.se, davidle@kth.se

**Abstract**

This paper reports on a system created to help language students learn the vowel inventory of Swedish. Formants are tracked, and a 3D ball moves over a vowel-chart canvas in real time. Target spheres are placed at the target values of vowels, and the students' task is to get the target spheres. A calibration process of capturing data from three cardinal vowels is used to normalize the effects of different size vocal tract, thus making it possible for people to use the program, regardless of age, size, or gender. A third formant is used in addition to the first and second formant, to distinguish the difference between two Swedish vowels.

## FRIDAY 4th SEPTEMBER 2009

**Fri 4/9/09 09:00 – 10:30 Oral session 3**

## CREATION & ASSESSMENT OF CONTENT FOR SLaTE

### O3.1 Automatic Question Generation and Answer Judging: A Q&A Game for Language Learning

*Yushi Xu, Anna Goldie, Stephanie Seneff*

Spoken Language Systems Group, MIT Computer Science and Artificial Intelligence Laboratory, United States

{yushixu, agoldie, seneff}@csail.mit.edu

**Abstract**

We have designed a question and answer game for students learning Mandarin Chinese. The game produces spoken questions from automatically generated statements, and judges the student's answers automatically. The student interacts with the system by speech, so that comprehensive reading, listening and speaking ability can be practiced. This paper focuses on the methods for question generation and answer judgment, as well as the game implementation. Evaluation results have shown that our methods and the game system are both successful.

### O3.2 Analysis of Vocabulary Difficulty Using Wiktionary

*Julie Medero, Mari Ostendorf*

Department of Electrical Engineering, University of Washington, Seattle, WA 98195, USA

{jmedero, mo}@ee.washington.edu

**Abstract**

Assessing vocabulary difficulty is useful for finding and creating texts at low reading levels. Prior work has focused on characteristics such as word length and word frequency. In this work, we explore whether other cues might be useful, using features extracted from Wiktionary entries. Comparing words in comparable articles in Standard and Simple English Wikipedia, we find that words that appear in Standard but not Simple English tend to have shorter definitions, fewer part-of-speech types and word senses, and fewer languages that they have been translated into.

### O3.3 Semi-Automatic Generation of Cloze Question Distractors Effect of Students' L1

*Juan Pino, Maxine Eskenazi*

Language Technologies Institute, Carnegie Mellon University

{jmpino,max}@cs.cmu.edu

**Abstract**

We describe a method to semi-automatically generate incorrect choices, or distractors, for cloze (fill-in-the-blank) questions. We generated distractors aimed at revealing what type of misunderstanding a student was having. English as a Second Language learners answered a series of cloze questions that presented distractors generated by our method. We analyzed their answers in order to see how native languages influence the type of distractor that is chosen. With this preliminary study, we intend to further individualize the use of an intelligent tutoring system for vocabulary learning.

### O3.4 Porting REAP to European Portuguese

*Luís Marujo[1], José Lopes[1], Nuno Mamede[1], Isabel Trancoso[1], Juan Pino[2], Maxine Eskenazi[2], Jorge Baptista[3], Céu Viana[4]*

[1]INESC-ID Lisboa / IST, Portugal, [2]LTI / CMU, USA, [3]Univ. Algarve Portugal, [4]CLUL, Portugal

Luis.Marujo@inesc-id.pt

**Abstract**

This paper describes the early stages of porting REAP, a tutoring system for vocabulary learning, to European Portuguese. A large number of linguistic resources and filtering tools have already been integrated into the ported version. We modified the current system to also target oral comprehension.

**Fri 4/9/09 11:00 – 12:45 Poster 2**

## PRONUNCIATION TRAINING & ASSESSMENT (2)

### P2.1 Developing a CALL System for Practicing Oral Proficiency: How to Design for Speech Technology, Pedagogy and Learners

*Helmer Strik[a], Frederik Cornillie[b], Jozef Colpaert[b], Joost van Doremalen[a], Catia Cucchiarini[a]*

[a]Department of Linguistics, Radboud University, Nijmegen, The Netherlands

[b]Linguapolis - Institute for Language and Communication, University of Antwerp, Antwerp, Belgium

h.strik|j.vandoremalen|c.cucchiarini@let.ru.nl; frederik.cornillie|jozef.colpaert@ua.ac.be

**Abstract**

Automatic recognition of non-native speech is problematic. A key challenge in developing spoken CALL systems is to design exercises that enable learning but which are still technically feasible. This especially applies to systems intended for practicing grammar. In the current paper we focus on the issue of matching design and speech technology. On the one hand we are developing and testing speech technology modules to determine what is feasible. On the other we use this knowledge in designing a CALL system for practicing pronunciation and grammar.

## P2.2 Towards a Computer-aided Pronunciation Training System for German Learners of Mandarin

*Hansjörg Mixdorff[1], Daniel Külls[1], Hussein Hussein[1], Gong Shu[2], Hu Guoping[2], Wei Si[2]*

[1]Department of Informatics and Media, BHT University of Applied Sciences, Berlin, Germany

[2]Dept. EEIS, University of Science and Technology of China, Hefei, Anhui, P.R.China

mixdorff@bht-berlin.de, kuells@bht-berlin.de, hussein@bht-berlin.de, shugong@mail.ustc.edu.cn, gphu@iflytek.com, siwei@iflytek.com

**Abstract**

The current paper discusses first investigations aimed to lay the groundwork for the development of computer-aided pronunciation training for teaching Mandarin to Germans. We conducted a contrastive analysis of the two languages leading to a set of tokens for a production and perception experiment involving German first-year students of Mandarin. Their data were perceptually evaluated by a teaching expert for Mandarin, native speakers of Mandarin as well as processed by a Mandarin automatic speech recognition system.

## P2.3 Formant Estimation in Children's Speech and its application for a Spanish Speech Therapy Tool

*William R. Rodríguez, Eduardo Lleida*

Communications Technology Group (GTC), Aragon Institute for Engineering Research (I3A), University of Zaragoza, Zaragoza, Spain

fwricardo,lleidag@unizar.es

**Abstract**

This paper addresses the problem of how to estimate reliable formant frequencies in high-pitched speech (typical in children), and how to normalize these estimations, independent from vocal tract shape or length. The normalized formant frequencies are used to improve the performance of a Computer-Aided Speech Therapy Tool (CASTT) in Spanish. For this purpose, a study was conducted to see what is the relationship between child's height and their vocal tract length, using traditional technologies in speech processing like linear prediction LPC, homomorphic analysis and modeling of the vocal tract. Results of this study show a high correlation between child's height and their vocal tract length. The study is based on speech from 235 healthy children (110 females and 125 males) which contains Spanish vowels utterances, and enables calibration of a CASTT system for children with speech disorders.

## P2.4 The EURONOUNCE corpus of non-native Polish for ASR-based Pronunciation Tutoring System

*Natalia Cylwik, Agnieszka Wagner, Grażyna Demenko*

Adam Mickiewicz University, Institute of Linguistics, Department of Phonetics, Poznań, Poland

{nataliac, wagner, lin}@amu.edu.pl

**Abstract**

This paper gives a detailed information on the design of the speech corpus for the purpose of developing an ASR-based pronunciation tutoring system. In the first place, assumptions on the structure of the corpus are presented. Then collection of text material, recordings and procedure of annotation of the resulting speech corpus are described. In the end, preliminary results of the analysis of pronunciation errors are discussed. They provide information which is important for ASR training and testing on the one hand, and automatic error detection on the other hand.

## P2.5 Investigating the Effectiveness of Prontest Software to Train English Proficiency

*Hiroyuki Obari(Aoyama Gakuin University), Hiroaki Kojima(AIST), Machi Okumura(Prontest Inc.), Masahiro Yoshikawa(University of Tsukuba), Shuichi Itahashi(NII/AIST)*

College of Economics Aoyama Gakuin University, Tokyo, Japan

hobari@gmail.com

**Abstract**

This paper is to investigate the effectiveness of Prontest software to improve English pronunciation and proficiency for Japanese EFL learners. Several parameters such as speech duration, speech power, F0 (pitch), the ratio of vowel and consonant length and power were introduced to find out how much students made progress in English pronunciation and overall English proficiency. The study concluded that the average score of CASEC computer test improved from 532(SD 109.2) in April to 583(SD 83.1) in July after having used this software for six lessons. The differences of parameters between pre and post-recorded readings indicated that this software helped students to improve English pronunciation.

## P2.6 An Experience with a Spanish Second Language Learning Tool in a Multilingual Environment

*Oscar Saz[1], Victoria Rodríguez[2], Eduardo Lleida[1], W.-R. Rodríguez[1], C. Vaquero[1]*

[1]Communications Technology Group (GTC), Aragón Institute for Engineering Research (I3A), University of Zaragoza, Zaragoza, Spain

[2] Vienna International School, Vienna, Austria

oskarsaz@unizar.es,vrodriguez@vis.ac.at,flleida,wricardo,cvaquerog@unizar.es

**Abstract**

This paper presents the results of an experience with "VocalizaL2", an application for Second Language (L2) learning of Spanish, in a multilingual environment at the Vienna International School (VIS). For the experiment, a group of 6th-graders at the school practiced with the application during 5 sessions altogether with their regular classes. The results of the experiment show on one hand, the great motivation power that computer-based L2 tools have for the pronunciation training of young learners, while also resulting useful for the teachers. On the technical aspect, the tool and the algorithms within are described and a preliminary analysis points out their ability to correct and motivate non-native Spanish pronunciation.

**Fri 4/9/09 14:00 – 15:30 Oral 4**

## PRONUNCIATION TRAINING & ASSESSMENT (3)

### O4.1 Audio Quality Issue for Automatic Speech Assessment

*Lei Chen*

Educational Testing Service, Princeton, NJ, USA

LChen@ets.org

**Abstract**

Recently, in the language testing field, automatic speech recognition (ASR) technology has been used to automatically score speaking tests. This paper investigates the impact of audio quality on ASR-based automatic speaking assessment. Using the read speech data in the International English Speaking Test (IEST) practice test, we annotated audio quality and compared scores rated by humans, speech recognition accuracy, and the quality of features used for the automatic assessment under high and low audio quality conditions. Our investigation suggests that human raters can cope with low-quality audio files well, but speech recognition and the features extracted for the automatic assessment perform worse on the low audio quality condition.

### O4.2 Measuring Children's Phonemic Awareness through Blending Tasks

*Shizhen Wang[1], Patti Price[2], Yi-Hui Lee[1] and Abeer Alwan[1]*

[1]Department of Electrical Engineering, University of California, Los Angeles

[2]PPRICE Speech and Language Technology Consulting

szwang@ee.ucla.edu, pjp@pprice.com, yihuilee@ucla.edu and alwan@ee.ucla.edu

**Abstract**

In this paper, speech recognition techniques are applied to automatically evaluate children's phonemic awareness through three blending tasks (phoneme blending, onset-rhyme blending and syllable blending). The system first applies disfluency detection to filter out disfluent phenomena such as false-starts, sounding out, self-repair and repetitions, and to localize the target answer. Since most of the children studied are Hispanic, accent detection is applied to detect possible Spanish accent. The accent information is then used to update the pronunciation dictionaries and duration models. For valid words, forced alignment is applied to generate sound segmentations and produce the corresponding HMM log likelihood scores. Normalized spectral likelihoods and duration ratio scores are combined to assess the overall quality of the children's productions. Results show that the automatic system correlates well with teachers, and requires no human supervision.

### O4.3 Detecting Prosody Improvement in Oral Rereading

*Minh Duong Jack Mostow*

Project LISTEN, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA

mnduong@cs.cmu.edu mostow@cs.cmu.edu

**Abstract**

A reading tutor that listens to children read aloud should be able to detect fluency growth – not only in oral reading rate, but also in prosody. How sensitive can such detection be? We present an approach to detecting improved oral reading prosody in rereading a given text. We evaluate our method on data from 133 students ages 7-10 who used Project LISTEN‟s

Reading Tutor. We compare the sensitivity of our extracted features in detecting improvements. We use them to compare the magnitude of recency and learning effects. We find that features computed by correlating the student‟s prosodic contours with those of an adult narration of the same text are generally not as sensitive to gains as features based solely on the student‟s speech. We also find that rereadings on the same day show greater improvement than those on later days: statistically reliable recency effects are almost twice as strong as learning effects for the same features.

### O4.4 A Self-Transcribing Speech Corpus: Collecting Continuous Speech with an Online Educational Game

*Alexander Gruenstein[1], Ian McGraw[1], Andrew Sutherland[1,2]*

[1]MIT Computer Science and Artificial Intelligence Lab, Cambridge, MA, USA

[2]Quizlet.com, Albany, CA, USA

alexgru@mit.edu, imcgraw@mit.edu, asuth@mit.edu

**Abstract**

We describe a novel approach to collecting orthographically transcribed continuous speech data through the use of an online educational game called Voice Scatter, in which players study flashcards by using speech to match terms with their definitions. We analyze a corpus of 30,938 utterances, totaling 27.63 hours of speech, collected during the first 22 days that Voice Scatter was publicly available. Though each individual game covers only a small vocabulary, in aggregate speech recognition hypotheses in the corpus contain 21,758 distinct words. We show that Amazon Mechanical Turk can be used to orthographically transcribe utterances in the corpus quickly and cheaply, with near-expert accuracy. Moreover, we present a filtering technique that automatically identifies a sub-corpus of 39% of the data for which recognition hypotheses can be considered human-quality transcripts. We demonstrate the usefulness of such self-transcribed data for acoustic model adaptation.

**Fri 4/9/09 18:30 – 20:00**

## Demonstration session 1

### D1.1 Japanese CALL system based on Dynamic Question Generation and Error Prediction for ASR

*Tatsuya Kawahara, Hongcui Wang, Yasushi Tsubota and Masatake Dantsuji*

Kyoto University, Japan

**Abstract**

We have developed a new Computer Assisted Language Learning (CALL) system to aid students learning Japanese as a second language. The system offers students the chance to practice elementary Japanese by creating their own sentences based on visual prompts, before receiving feedback on their mistakes. It is designed to detect lexical and grammatical errors in the input sentence as well as pronunciation errors in the speech input. Questions are dynamically generated along with sentence patterns of the lesson point, to realize variety and flexibility of the lesson. Students can give their answers with either text input or speech input. To enhance speech recognition performance, a decision tree-based method is incorporated to predict possible errors made by non-native speakers for each generated sentence on the fly.

## D1.2 Development of a CALL System to Enhance ESL/EFL Learners' Skills of Shadowing and Reading Aloud

*Dean Luo[1], Nobuaki Minematsu[1], Yutaka Yamauchi [2]*

[1] The University of Tokyo [2] Tokyo International University

**Abstract**

The CALL system developed in our project enables ESL/EFL learners to enhance their skills of shadowing and reading aloud. Learners are required to record their shadowing and reading aloud into the computer while listening to passages read by a native speaker of English. After recording, they can listen to their voices and observe the sound waves of their own recording and the model one. Through auditory and visual comparison of the two recordings, they can understand the shortcomings of their performances and where they should practice more.

Especially in shadowing practice, learners' shadowed speech is automatically analyzed and evaluated by the computer using speech information processing technology like GOP (goodness of pronunciation). Their English proficiency levels measured by TOEIC (Test of English as International Communication) are also predicted and presented. Based on the results of automatic scoring, the learners can understand how well they have conducted shadowing objectively and also grasp their own proficiency levels.

From the viewpoint of material development, this CALL system enables instructors to choose any speech data obtained from CDs, DVDs, Web sites, etc, and use them as practice materials. For instance, if both audio and text files of President Barack Obama's inauguration address are available, the learners can practice shadowing and reading aloud using his famous speech. Thus instructors' selection of speech data suitable for the learners' interests and proficiency levels can increase student motivation and continuous use of this system, hence improving both aural and oral skills.

**References**

Dean Luo, Nobuaki Minematsu, Yutaka Yamauchi and Keikichi Hirose. "Analysis and comparison of automatic language proficiency assessment between shadowed sentences and read sentences," Proc. SLaTE 2009

## D1.3 Computer Aided Pronunciation Training (CAPT) System "AZAR"

*Michael Beilig*

**Abstract**

The AzAR functionality provides several audio-visual modes of user feedback, e. g. showing animated articulatory organs to correct wrong movements of tongue, lips, etc. or playing back reference utterances but the core function is marking mispronounced phones within the spoken utterance using a coloured scale from red ("bad") to green ("good"):


The marking of mispronounced parts on a user's utterances is based on different phonetic-phonologic and prosodic distance measures - identifying typical cross-lingual influences from the native L1 source language on the L2 target language taught, such as:

- Confusion of specific phoneme classes,
- Wrong phoneme duration,
- Articulation mistakes e. g. voicing unvoiced phonemes.

The practical implementation uses confidence measures on the segmental level from a HMM based speech recognizer. The AzAR programme structure follows an extensive phonetic curriculum, containing contrastive exercises, insertion tests, etc. which has been compiled from real lessons and is supplemented by a glossary. The AzAR 2 prototype was tested and optimized for Russian migrants in Dresden learning German and runs on PC (Linux, Windows, Mac OSX).

It involves a reference speech database for the given language pair combination Russian (L1)/German (L2).

---

### D1.4 Spoken dialog systems for learning foreign language communication skills

*Andre Valente and W. Lewis Johnson*

Alelo Inc

**Abstract**

Alelo develops a variety of products for learning foreign languages and cultures. These systems make heavy use of speech and language technologies, particularly in spoken dialog with animated non-player characters. This demonstration will feature two of these systems. We will show one of the Operational Language and Culture family of products.

These are self-paced computer-based learning environments that help learners acquire mission-related communication skills. Versions of these learning environments are in use by military forces in the United States, Australia, and other countries. Both personal computers (with speech recognition) and iPods (without speech recognition) are supported. Languages supported include Arabic, Pashto, Dari, Urdu, French, and Indonesian. We will also show the GoEnglish Web site. Developed for Voice of America, GoEnglish is intended to help language learners around the world learn American language and culture. Spoken language exercises help learners at a range of levels to develop proficiency in spoken colloquial American English. Versions for multiple first languages are available.

---

### D1.5 Langofone - Language learning in your pocket

*Preben Wik*

Center for Speech Technology (CTT), Dept. Speech, Music and Hearing, KTH, Stockholm, Sweden

**Abstract**

We demonstrate Langofone - a language learning system for mobile phones, developed by The Centre for Speech Technology at KTH, together with Luli Media group, and Sirocco. Langofone consists of three sections:

PhraseBook: Listen to recordings of phrases, organized by topic into packages. Record your own attempts to say a phrase, send your recordings to the analysis server, and get feedback consisting of four individual scores and a weighted average.

Quiz: Reading and listening comprehension on a list of phrases you would like to practice through multiple-choice questions

Translate: An API to Google translate, allowing you to translate any word or sentence from and to 12 languages

Langofone is compatible with approximately 90 per cent of all cell phones on the European market.

Read more on www.langofone.com.

### D1.6 COMUNICA: MULTILEVEL TOOLS FOR SPANISH CALL

*Oscar Saz, W.-Ricardo Rodríguez, Eduardo Lleida, Carlos Vaquero*

Communications Technology Group (GTC), Aragón Institute for Engineering Research (I3A), University of Zaragoza, Zaragoza, Spain

**Abstract**

This demo aims to bring a closer view of all the free-distribution tools gathered in "Comunica" (http://www.vocaliza.es). "Comunica" is set, initially, to provide speech therapy to children with special needs during their processes of language acquisition in Spanish. The three tools ("PreLingua", "Vocaliza" and "Cuéntame") cover different levels of language, from phonation and articulation to linguistic and descriptive abilities. The demo is focused on the more recent advances in "Comunica" presented at the SLaTE workshop: A novel tool for the training of the Spanish vowels with on-line formant normalization and a new version of "Vocaliza", with phoneme-level evaluation and feedback, oriented to young children who approach Spanish as a second language.

### D1.7 Pronunciation Evaluation System for Oral English Testing and Oral Chinese Testing

*Si Wei*

## SATURDAY 5th SEPTEMBER 2009

**Sat 5/9/09 09:00 – 10:30 Poster 3**

## TOPICS IN SLaTE (1)

### P3.1 An Audiovisual Feedback System for Acquiring L2 Pronunciation and L2 Prosody

*Grażyna Demenko[1], Agnieszka Wagner[1], Natalia Cylwik[1] and Oliver Jokisch[2]*

[1]Adam Mickiewicz University, Institute of Linguistics, Department of Phonetics, Poznań, Poland

[2]TU Dresden, Laboratory of Acoustics and Speech Communication, Dresden, Germany

[1]{lin, wagner, nataliac}@amu.edu.pl, [2]oliver.jokisch@tu-dresden.de

**Abstract**

In recent years the application of computer software to the learning process has been acknowledged an indisputably effective tool supporting traditional teaching methods. A particular focus has been put on the application of computational techniques based on speech and language processing to second language learning. At present, a number of commercial self-study programs using speech synthesis and recognition are available. Most of them, however, focus on segmental features only. The paper presents technical and linguistic specifications for the Euronounce project [1] which aims at creating an intelligent tutoring system with multimodal feedback functions for acquiring not only foreign languages' pronunciation but also prosody. The project focuses on German as a target language for native speakers of Polish, Slovak, Czech and

Russian and vice versa. The paper outlines the Euronounce feedback system and presents the Pitch Line program which can be implemented in the prosody training module of the Euronounce tutoring system.

## P3.2 Using speech technology to promote increased pitch variation in oral presentations

*Rebecca Hincks[1], Jens Edlund[2]*

Unit for Language and Communication, CSC, KTH, Sweden[1]

Centre for Speech Technology, CSC, KTH, Sweden[2]

hincks@speech.kth.se, edlund@speech.kth.se

**Abstract**

This paper reports on an experimental study comparing two groups of seven Chinese students of English who practiced oral presentations with computer feedback. Both groups imitated teacher models and could listen to recordings of their own production. The test group was also shown flashing lights that responded to the standard deviation of the fundamental frequency over the previous two seconds. The speech of the test group increased significantly more in pitch variation than the control group. These positive results suggest that this novel type of feedback could be used in training systems for speakers who have a tendency to speak in a monotone when making oral presentations.

## P3.3 Training Novel Phonemic Contrasts: A Comparison of Identification and Oddity Discrimination Training

*Zöe Handley[a], MikeSharples[a], Dave Moore[b]*

[a]Learning Sciences Research Institute, University of Nottingham, UK

[b]Medical Research Council (MRC) Institute of Hearing Research (IHR), Nottingham, UK

zoe.handley@nottingham.ac.uk, mike.sharples@nottingham.ac.uk, dave@ihr.mrc.ac.uk

**Abstract**

High Variability Pronunciation Training (HVPT) is a highly successful alternative to ASR-based pronunciation training. It has been demonstrated that HVPT is effective in teaching the perception of non-native phonemic contrasts, and that this skill generalizes to the perception of unfamiliar words and talkers, transfers to pronunciation, and is retained long-term. HVPT is, however, not efficient and hence not motivating for the learner. In this study, we therefore compare HVPT with an alternative, namely oddity discrimination training. This comparison, in which Mandarin-Chinese speakers were trained to pronounce the English /r/-/l/ phonemic contrast, provides preliminary evidence to support the use of discrimination tasks in addition to identification tasks to add variety to HVPT.

## P3.4 *TalkMaths* : A Speech User Interface for Dictating Mathematical Expressions into Electronic Documents

*Angela M. Wigmore, Gordon J.A. Hunter, Eckhard Pflügel and James Denholm-Price*

Faculty of Computing, Information Systems and Mathematics, Kingston University, KT1 2EE, U.K.

{k0330947, G.Hunter, E.Pfluegel, j.denholm-price}@kingston.ac.uk

**Abstract**

We describe the development of a speech-driven user interface system, *TalkMaths*, which enables the dictation of mathematical expressions into electronic documents without

the user needing extensive knowledge of any specialized markup language. This system should be of value to many students and teachers, particularly those with disabilities - for whom typing mathematical text is currently very difficult.

### P3.5 Automated Generation of Example Contexts for Helping Children Learn Vocabulary

*Liu Liu, Jack Mostow, and Gregory Aist*

Project LISTEN, School of Computer Science , Carnegie Mellon University, Pittsburgh, Pennsylvania USA

{liuliu, mostow}@cs.cmu.edu, gregory.aist@alumni.cmu.edu

**Abstract**

This paper addresses the problem of generating good example contexts to help children learn vocabulary. We construct candidate contexts from the Google N-gram corpus. We propose a set of constraints on good contexts, and use them to filter candidate example contexts. We evaluate the automatically generated contexts by comparison to example contexts from children's dictionaries and from children's stories.

### P3.6 Improving Phone Verification Using State-level Posterior Features and Support Vector Machine for Automatic Mispronunciation Detection

*Khe Chai SIM*

School of Computing, Department of Computer Science, National University of Singapore, Singapore.

simkc@comp.nus.edu.sg

**Abstract**

An important aspect of a Computer-Assisted Language Learning (CALL) system for pronunciation acquisition is the automatic detection of mispronunciations. This problem can be formulated as a phone verification task. For each phone to be verified, the system generates a verification score and a decision threshold is applied to accept or reject the pronunciation of that phone. Most verification systems use the HMM phone acoustic models to compute the log posterior probabilities (LPPs) as the verification score. A discriminative back-end using the Support Vector Machine (SVM) can also be applied to the vector of LPPs to further improve the verification performance. This paper investigates the use of a NN/HMM hybrid phone recognizer to obtain the LPP scores. The NN/HMM hybrid framework has been shown to yield superior phone recognition performance over the conventional GMM/HMM based systems. In addition, this paper also examines the use of frame-level phone or state posterior features directly with SVM. Experimental results reported on the TIMIT database show that state-level average posterior features with SVM yielded 9.5% relative Equal Error Rate (EER) improvement over the NN/HMM system.

## Sat 5/9/09 11:00 – 12:45

## Demonstration session 2

### D2.1 Structure-based pronunciation assessment

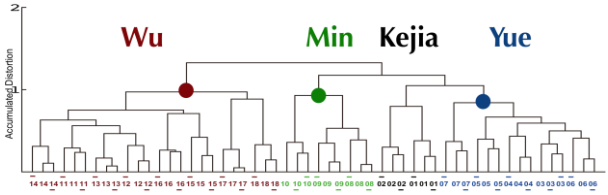
*Nobuaki Minematsu and Masayuki Suzuki*

**Abstract**

We present two demonstration systems both using the same and new speech technologies for pronunciation assessment. In one demo, a

learner's pronunciation of English vowels is assessed by comparing the vowels to those of a teacher who is selected by that learner based on his/her preference. The system instructs which vowel should be corrected at first to become like the selected teacher. In the other demo, Chinese speakers are classified purely based on their dialects. Here, the demo system does not identify the dialect of that speaker but classifies the input speaker and the speakers in the database only based on their dialectal accents, not influenced by age and gender of the speakers. The two demo systems are commonly built on structure-based pronunciation assessment, where a sound system (structure) underlying a speaker's pronunciation is estimated and the sound system (structure) is compared to that of another speaker. Then, differences between the two systems (structures) are quantitatively calculated. It should be noted that the differences do not include any differences caused by extra-linguistic factors because pronunciation structures are extracted from utterances by removing extra-linguistic features from speech acoustics. For example, the pronunciation of a child can be compared directly to that of a very tall male speaker although their voice quality is totally different. In the same way, dialect-based speaker classification is possible among children and adults. After recording some utterances of a participant, the result of pronunciation assessment is printed out and handed out to the participant within a minute.


*Teacher selection window*




*Classification of Chinese adults and children based on dialects*

## D2.2 Voice Race and Voice Scatter: Online Educational Games for Collecting Orthographically-Labeled Speech Data

*Alexander Gruenstein, Ian McGraw, and Andrew Sutherland*

**Abstract**

Voice Race and Voice Scatter are online education games available on the popular flashcard website Quizlet.com. Quizlet users can make and share sets of virtual flashcards, which each contain a term on one side and a definition on the other. Quizlet boasts 420,000 registered users who have created over 875,000 sets of flashcards, which altogether contain more than 24 million individual flashcards. Voice

Race and Voice Scatter use the publicly available WAMI Javascript API, which makes it easy to incorporate speech recognition capabilities into Web applications, to provide a fun way for users to study flashcards on the website by speaking. Moreover, by using recognition confidence scores and contextual information from the games, it is possible to automatically orthographically label a large portion of the collected utterances with near-human accuracy. As such, the games provide diversion, educational value, and labeled speech data.

### D2.3 REAP.PT, a tutoring system for teaching Portuguese

*L. Marujo, J. Lopes, N. Mamede, I. Trancoso, J. Pino, M. Eskenazi, J. Baptista, C. Viana*

**Abstract**

This demo shows the baseline version of REAP.PT, a tutoring system for vocabulary learning, which has been ported from the English version developed by CMU to European Portuguese. Students learn from authentic materials, on topics of their preference. A large number of linguistic resources and filtering tools have already been integrated into the ported version, which has been modified to also target oral comprehension.

---

### D2.4 Virtual Chinese Tutor (VCT) - A Chinese Language Pronunciation Learning Software

*Yow-Bang Wang*[1,2]*, Hsin-Min Wang*[1]*, Lin-Shan Lee*[1,2]

[1]Institute of Information Science, Academia Sinica

[2]Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan, R.O.C.

**Abstract**

Virtual Chinese Tutor (VCT) is a successfully operating online Chinese pronunciation learning software, specifically designed for giving the students opportunities to practice both their listening and speaking skills as many time as they wish anytime and anywhere. It was developed by a joint effort between Academia Sinica, National Taiwan University and some industry partners. The first version of VCT has been completed and made available on-line at http://chinese.ntu.edu.tw/. Virtual Chinese Tutor (VCT) is able to evaluate the pronunciation of each utterance produced by an individual learner from four different aspects: pronunciation (i.e. the Initial (initial consonant) and Final (vowel or diphthong part plus an optional nasal ending) of each individual syllable (or character)), pitch (i.e. the lexical tone or neutral tone of each individual syllable (or character)), timing (i.e. the duration distribution among different syllables (or characters) in the utterance), and emphasis (i.e. the energy distribution among different syllables (or characters) in the utterance), as well as an overall score for the entire utterance. For those phonemes with scores below a threshold, a 3-dimensional video will show on the screen to demonstrate the actions of the vocal tract shape, including the relative positions among the lip, tongue and other articulators. After a learning session is finished, the diagnostic statistics for the learner is automatically summarized, indicating the important directions for improvements. This software platform can be used with any course content as long as the text and voice files for the course content are given. For each utterance produced by the learner, forced alignment with the corresponding utterance produced by the instructor is first performed, and the pronunciation, pitch, timing and emphasis scores for each syllable are then evaluated by a set of acoustic models, tone models, and prosodic models. The overall score is then the weighted sum of all those scores. The scoring algorithm was trained with the scores given by real professional Chinese teachers, over a corpus produced by a group of real learners whose mother tongues are not Chinese. Both the above training corpus and course content currently used in this software were contributed by the International Chinese Language Program of National Taiwan University. The details of the technologies will be explained when demonstrating the system in the technical program.

VCT is now also a part of the program offered by NTUtorMing (http://www.ntutorming.com/), which is an online teaching institution that was jointly established by National Taiwan

University (NTU) and TutorABC, a company located in Taipei focusing on language education. Chinese Learners around the world can have real-time interaction with professional Chinese teachers through TutorABC's online learning platform, and VCT offers after-class practice or homework for the students.

### D2.5 Automatic assessment of non-native prosody

*Florian Hönig, Anton Batliner, Karl Weilhammer, Elmar Nöth*

**Abstract**

Wrong placement of word accents as well as any `non-native' prosody such as the transfer of `syllable-timed' rhythm onto English which is `stress-timed' can have a strong impact on (native) listeners and should be avoided. Thus, they should be addressed in CAPT applications.

To this aim, we present a demonstrator that is able to estimate the position of the word accent and the probability of an erroneous word accent position. Moreover, the quality of intonation and rhythm is estimated; the system is trained with annotations obtained from American and British natives.

### D2.6 A Spoken Dialog System for Learners of English

*Karl Weilhammer, Catharine Oertel, Robin Siegemund, Ricardo Sá, Anton Batliner, Florian Hönig und Elmar Nöth*

**Abstract**

Current E-Learning software presents texts, pictures, audio sounds and videos to the learner, but in many cases the user interface only allows writing or clicking on items. Some more elaborate systems for foreign-language learning allow the user to interact via speech, but only provide reading or repeating pre-specified sentences.

We present a spoken-dialog system for learners of English, which is designed for conversational training. The research prototype will demonstrate a hotel-reception scenario. The system plays the role of the receptionist and the user, the hotel guest at check-in. With a system like this, learners of English can prepare themselves for real-life situations before actually travelling to Britain or the US.

Current speech dialog systems (e.g. for telephone banking or flight information) are tailored to perform a task very efficiently, using a fixed policy that guides the user along a strict path through the task. In foreign language learning this approach works well only for beginners. Our system is targeted towards more advanced learners and implements a different dialog strategy. In each system state the dialog manager randomly selects one of many plausible, but different system actions. This keeps the task variable and interesting for the learner, even when repeated several times.

### D2.7 CLIMB LEVEL 4 – teaching English for aviation safety

*Maxine Eskenazi, Gary Pelton*

*Carnegie Speech Company*

**Abstract**

Many aviation accidents and near misses have been caused by mis-communications between pilots and air traffic controllers. As part of their efforts to remediate this situation, the international air safety organization, ICAO, has mandated that all pilots and air traffic controllers pass English fluency tests by the end of 2011. Due to the nomadic nature of their job, pilots need non-classroom English training to pass these tests. Carnegie Speech Company has released, in cooperation with Mayflower

College of England, a product called Climb Level 4 (www.climb-level4.com) that trains pilots and air traffic controllers to be more fluent English speakers.

**Sat 5/9/09 13:45 – 15:00 Oral session 5**

## TOPICS IN SLaTE (2)

### O5.1 Improved Structure-based Automatic Estimation of Pronunciation Proficiency

*Masayuki Suzuki, Luo Dean, Nobuaki Minematsu, Keikichi Hirose*

The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, 113-8656, Japan

{suzuki,dean,mine,hirose}@gavo.t.u-tokyo.ac.jp

**Abstract**

Automatic estimation of pronunciation proficiency has its specific difficulty. Adequacy in controlling the vocal organs is often estimated from spectral envelopes of input utterances but the envelope patterns are also affected by alternating speakers. To develop a good and stable method for automatic estimation, the envelope changes caused by linguistic factors and those by extra-linguistic factors should be properly separated. In our previous study [1], to this end, we proposed a mathematically guaranteed and linguistically-valid speaker-invariant representation of pronunciation, called speech structure. After the proposal, we have tested that representation also for ASR [2, 3, 4] and, through these works, we have learned better how to apply speech structures for various tasks. In this paper, we focus on a proficiency estimation experiment done in [1] and, using the recently developed techniques for the structures, we carry out that experiment again but under different conditions. Here, we use a smaller unit of structural analysis, speaker-invariant substructures, and relative structural distances between a learner and a teacher. Results show higher correlation between human and machine rating and also show extremely higher robustness to speaker differences compared to widely used GOP scores.

### O5.2 Predictable and Educational Spoken Dialogues: Pilot Results

*Gregory Aist[1] and Jack Mostow[2]*

[1]Language Technologies Institute and [2]Robotics Institute, Carnegie Mellon University, Pittsburgh, USA

Gregory.Aist@alumni.cmu.edu

**Abstract**

This paper addresses the challenge of designing spoken dialogues that are of educational benefit within the context of an intelligent tutoring system, yet predictable enough to facilitate automatic speech recognition and subsequent processing. We introduce a design principle to meet this goal: construct short dialogues in which the desired student utterances are external evidence of performance or learning in the domain, and in which those target utterances can be expressed as a well-defined set. The key to this principle is to teach the human learner a process that maps inputs to responses. Pilot results in two domains – self-generated questions and morphology exercises – indicate that the approach is promising in terms of its habitability and the predictability of the utterances elicited. We describe the results and sketch a brief taxonomy classifying the elicited utterances according to whether they evidence student performance or learning, whether they are amenable to automatic processing, and whether they support or call into question the hypothesis that such dialogues can elicit spoken

utterances that are both educational and predictable.

## O5.3 Voice morphing and the manipulation of intra-speaker and cross-speaker phonetic variation to create foreign accent continua: A perceptual study

*John Ingram*[1]*, Hansjörg Mixdorff*[2] *and Nahyun Kwon*[1]

[1]School of EMSAH, University of Queensland, Brisbane, Australia

[2]Department of Informatics and Media, BHT University of Applied Sciences, Berlin, Germany

j.ingram@uq.edu.au, mixdorff@beuth-hochschule.de

**Abstract**

The STRAIGHT system of voice morphing was used to create voice continua of (Korean) accented Australian English, intended to simulate phonetic variation ranging from 'heavily accented' to 'unaccented' (native-like) Australian English, employing dimensions of intra-speaker and cross-speaker variation to yield a range of synthetic voices. These synthetic voices were evaluated against actual samples of Korean accented English, both re-synthesized and non-re-synthesized, in a series of three perceptual rating experiments by native listeners of Australian English. The questions of central interest in this preliminary investigation are: (a) the method of creating the phonetic continua and the respective roles of intra- versus cross-speaker variability in simulating degrees of foreign accent, (b) the success of the STRAIGHT method for creating hybrid voices, compared with 'natural' tokens of accented utterances, and (c) the impact of the re-synthesis method (required for voice morphing) upon perceptual ratings of foreign accent by native listeners. The ultimate objective of this research is to assess the impact of segmental and prosodic features on the perception of foreign accent and intelligibility of L2 learners' speech, where the source (Korean) and target (English) languages pose significant difficulties of segmental and prosodic transfer.