SLaTE 2015
Workshop on Speech and Language Technology in Education • September 4–5, 2015 • Leipzig
Program & Abstracts
Program and Abstract Book

SLaTE 2015

Sixth Workshop on Speech and Language Technology in Education

Satellite Workshop of Interspeech 2015 of the ISCA Special Interest Group SLaTE

September 4–5, 2015
Leipzig, Germany

https://www.slate2015.org
Message of the Organizers

In 2015, the Workshop on Speech and Language Technology in Education (SLaTE) takes place in Leipzig, Germany, as a Satellite Workshop of INTERSPEECH 2015 in Dresden. This workshop is the meeting of the correspondent ISCA Special Interest Group and is organized this year by the Pattern Recognition Lab of Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) in cooperation with Hochschule für Telekommunikation Leipzig (HfTL). Meanwhile, it is the sixth edition of the workshop. It all started in 2007 in Farmington, U.S.A., and was continued two years later in Birmingham, UK. It then took place every year, 2010 in Tokyo, Japan, and 2011 in Venice, Italy, and is now again on a bi-annual basis. In 2013, the workshop took place in Grenoble, France. Over the last years, it has become an established international workshop with a strong community and a constant number of submissions. We received a high number of 45 submissions this year and are happy to welcome you to SLaTE 2015 in Leipzig in order to discuss the latest advances in the domain of Speech and Language Technology in Education.

The 45 submissions can be subdivided into 36 full papers with a length of four to six pages, four short papers of two pages each, and five 1-page proposals for demonstrations. Each submission has been reviewed by three reviewers. The Technical Program Chair and the International Scientific Review Committee assured the high quality of the accepted papers. Finally, 30 full papers could be accepted for oral and poster presentations, and two short papers were accepted for poster presentations. One paper was turned into a demonstration, finally ending up with six demonstrations.

As the special interest group SLaTE preferred a two-day workshop opposed to the three-day meeting in 2013 and the number of accepted papers is similar to SLaTE 2013, the program is rather dense this year. The paper presentations are organized in five oral session, two poster sessions, and one session with demonstrations.

For the very first time, SLaTE has its own Satellite Workshop – the Workshop on L1 Teaching, Learning and Technology, which takes place in parallel to the poster and demonstration session on Friday. This Satellite of Satellite (SoS) workshop addresses researchers from didactics, psychology, and pedagogy. The aim is to join the two different communities of SLaTE and the SoS and to foster discussions on how the technologies developed in SLaTE can be applied to educational questions and datasets. The attendees of SLaTE are invited to attend the discussions in the Satellite Workshop and vice versa.

For a social gathering, the attendees of SLaTE are invited to the Feast of the Knights in one of Leipzig’s historic restaurants, the Ratskeller. We are looking forward to a successful and inspiring workshop! Our sincere gratitude goes to everybody who helped to make this workshop a huge success, especially the organizers of SLaTE 2013, the members of the International Scientific Committee of SLaTE and last but certainly not least all our members of the Scientific Review Committee.

Stefan Steidl, General Chair
Anton Batliner, Technical Program Chair
Oliver Jokisch, Local Chair
Scientific Committees

SLaTE 2015 is organized by Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Hochschule für Telekommunikation Leipzig (HfTL) in cooperation with the ISCA-SIG SLaTE. The organizers would like to thank Martin Russell, Helmer Strik, and Maxine Eskenazi for their advice and support.

General Chair

Stefan Steidl
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Germany

Technical Program Chair

Anton Batliner
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Germany
Technische Universität München, Germany

Local Chair

Oliver Jokisch
Hochschule für Telekommunikation Leipzig (HfTL), Germany

International Scientific Committee

Abeer Alwan, University of California at Los Angeles (UCLA), CA, U. S. A.
Pierre Badin, GIPTSA-lab, Grenoble, France
Jared Bernstein, Pearson Knowledge Technologies, Menlo Park, CA, U. S. A.
Catia Cucchiarini, Radboud University, Nijmegen, The Netherlands
Rodolfo Delmonte, Ca’ Foscari University Venice, Italy
Maxine Eskenazi, Carnegie Mellon University (CMU), Pittsburgh, PA, U. S. A.
Björn Granström, KTH, Royal Institute of Technology, Stockholm, Sweden
Valerie Hazan, University College London, UK
Diane Litman, University of Pittsburgh, PA, U. S. A.
Dominic Massaro, University of California at Santa Cruz (UCSC), CA, U. S. A.
Nobuaki Minematsu, University of Tokyo, Japan
Patti Price, PPRICE, Speech and Language Technology Consulting, Menlo Park, CA, U. S. A.
Martin Russell, University of Birmingham, UK
Stephanie Seneff, Massachusetts Institute of Technology (MIT), Cambridge, MA, U. S. A.
Helmer Strik, Radboud University, Nijmegen, The Netherlands

Scientific Review Committee

We thank all our reviewers for their help to make the workshop a big success.

Pierre Badin          Keelan Evanini          Patti Price
Géaerd Bailly        Björn Granström        Martin Russell
Anton Batliner       Rüdiger Hoffmann       Oscar Saz Torralba
Jared Bernstein      Florian Hönig         Stefan Steidl
Susanne Burger       Hiroaki Kato          Helmer Strik
Lei Chen             Mariko Kondo          Isabel Trancoso
Febe de Wet          Jacques Koreman        Jürgen Trouvaí
Rodolfo Delmonte     Joaquim Listerri       Karl Weilhammer
Ryan Downey          Nobuaki Minematsu     Preben Wik
Donna Erickson       Hansjörg Mixdorff
Supporters

This workshop is supported by

The International Speech Communication Association (ISCA) and the ISCA Archive

Institut für Bildungs- und Wissenschaftsmanagement, Leipzig, Germany

INTERSPEECH 2015 – The 16th Annual Conference of the International Speech Communication Association (ISCA), September 6–10, 2015, Dresden, Germany

TUBS GmbH, TU Berlin ScienceMarketing, Berlin, Germany
Authors List

Ghada Alharbi .............................................. 21
Lawrence Angrave ........................................ 10
Gérard Bailly .............................................. 14
Anton Batliner .............................................. 8, 12
Claudia Baur .................................................. 12, 16, 18
Kay Berking .................................................. 21
Suma Bhat ..................................................... 16
Anne Bonneau ............................................... 17, 19
Marie-Line Bosse ............................................ 14
Pierrette Bouillon .......................................... 12, 18
Enrique Cámara-Arenas ..................................... 18
Valentín Cardeñoso-Payo .................................. 18
Lei Chen ....................................................... 8
Nancy F. Chen .............................................. 10
Wenda Chen ................................................... 10
Cathy Chua ................................................... 12, 16, 18
Minhwa Chung ................................................. 17
Catia Cucchiarini ........................................... 16, 19
Febe de Wet .................................................. 19, 21
David Escudero-Mancebo .................................. 18
Keelan Evanini ............................................... 9
Dominique Fohr ............................................. 11, 19
Eva Frangi .................................................... 17
Mark J. F. Gales ............................................. 8
Mario Ganzeboom .......................................... 16
Emilie Gerbier ............................................... 14
César González-Ferreras ................................... 18
Wentao Gu .................................................... 10
Horst-Udo Hain ............................................ 13
Thomas Hain .................................................. 21
Mark Hasegawa-Johnson ................................... 10
Hiroya Hashimoto .......................................... 12
Khairun-nisa Hassanali ................................... 8
Hiroko Hirano ............................................... 12
Keikichi Hirose ............................................. 9
Juneki Hong .................................................. 18
Florian Höning ............................................. 8, 12
Wenping Hu .................................................. 15
Oliver Jokisch .............................................. 10
Denis Jouvet .................................................. 17
Shiying Kang ............................................... 18
Tatsuya Kawahara ......................................... 15
Daleen Klop .................................................. 21
Kate M. Knüll ............................................... 8
Yves Laprie .................................................. 17
Rémi Lavalley ............................................... 21
Jill Fain Lehman ............................................. 17
Kun Li .......................................................... 18
Shang-Wen Li ............................................... 20
Boon Pang Lim .............................................. 10
Hui Lin ....................................................... 12
Lei Liu ........................................................ 10
Pengfei Liu .................................................. 18
Bin Ma ......................................................... 10
Takehiko Makino .......................................... 9
Laurette Marais ............................................ 21
Karina Matthes .............................................. 13
Odile Mella ................................................... 11, 19
Tareb Melloul .............................................. 10
Helen Meng ................................................... 18
Nobuaki Minematsu ....................................... 9, 12
Maryam Sadat Mirzaei .................................... 15
Bernd Möbius ............................................... 17
Minsoo Na .................................................... 17
Diane Napolitano .......................................... 16
Baldur Neuber .............................................. 10
Raymond W. M. Ng ....................................... 21
Aibhne Ní Chasaide ....................................... 19
Neasa Ní Chiaráin ......................................... 19
Elmar Nöth .................................................... 8, 12
Luigi Palumbo ............................................... 19
Elisa Pellegrino ............................................ 14
Rico Petrick ............................................... 13
Nadine Pflaumer .......................................... 21
Teeraphon Pongkittiphan ................................ 9
Xiaojun Qian ............................................... 18
Yao Qian ..................................................... 15
Bhiksha Raj ................................................... 18
Manny Rayner ............................................. 12, 16, 18
Agha Ali Raza ............................................... 18
Jia Chen Ren ............................................... 10
Roni Rosenfeld ............................................. 18
Martin Russell ............................................. 17
Daisuke Saito ............................................... 9, 12
Frank K. Soong ............................................ 15
Helmer Strik ................................................. 16, 19
Cristian Tejedor-García ................................... 18
Jürgen Trouvain ............................................ 14, 17
Nikos Tsourakis ........................................... 12, 16, 18
Anjana Sofia Vakil ........................................ 12, 14
Rogier C. van Dalen ....................................... 8
Joost van Doremalen ...................................... 16
Debora Vigliano ............................................ 14
Mathias Walther .......................................... 10
Xinhao Wang ............................................... 9
Sebastian Wankerl ........................................ 12
Nikolas Wolfe .............................................. 18
Seung Hee Yang ............................................ 17
Su-Youn Yoon ............................................... 8, 9, 16
Frank Zimmerer .......................................... 18
Victor Zue .................................................... 20
Social Event

On Friday, September 4, 2015, at 7:30 pm, we will meet in one of Leipzig’s historic restaurants for a *Dinner of the Knights*. The restaurant is located in the center of Leipzig. From the workshop venue, it can be reached via the following public transportation:

- get on **Tram 9** heading towards *Thekla* and get off at station *Neues Rathaus*, or
- get on **Tram 11** heading towards *Schkeuditz* and get off at station *Wilhelm-Leuschner-Platz*.

Ratskeller Leipzig
Lotterstraße 1
04109 Leipzig
phone +49 (341) 123 4567

[http://www.ratskeller-leipzig.de](http://www.ratskeller-leipzig.de)
How Many Speakers, How Many Texts – The Automatic Assessment of Non-native Prosody

Florian Hönig, Anton Batliner, Elmar Nöth

Fri-O1-1

We present an in-depth analysis of a method for automatically scoring the prosody of non-native speech. For studying its suitability for different application scenarios, we perform a systematic comparison of different evaluation schemes such as text (in-)dependence and/or speaker (in-)dependence. The focus lies on methodological issues, with the aim of promoting the careful evaluation of automatic assessment methods. Further contributions are the analysis of (1) a method that utilizes speaker IDs to improve performance, and (2) the analysis of performance as a function of the number of speakers and texts used for training the system.

Automatically Grading Learners’ English Using a Gaussian Process

Rogier C. van Dalen, Kate M. Knill, Mark J. F. Gales

Fri-O1-2

There is a high demand around the world for the learning of English as a second language. Correspondingly, there is a need to assess the proficiency level of learners both during their studies and for formal qualifications. A number of automatic methods have been proposed to help meet this demand with varying degrees of success. This paper considers the automatic assessment of spoken English proficiency, which is still a challenging problem. In this scenario, the grader should be able to accurately assess the learner’s ability level from spontaneous, prompted, speech, independent of L1 language and the quality of the audio recording. Automatic graders are potentially more consistent than humans. However, the validity of the predicted grade varies. This paper proposes an automatic grader based on a Gaussian process. The advantage of using a Gaussian process is that as well as predicting a grade, it provides a measure of the uncertainty of its prediction. The uncertainty measure is sufficiently accurate to decide which automatic grades should be re-graded by humans. It can also be used to determine which candidates are hard to grade for humans and therefore need expert grading. Performance of the automatic grader is shown to be close to human graders on real candidate entries. Interpolation of human and GP grades further boosts performance.

Automatic Scoring of Non-native Children’s Spoken Language Proficiency

Khairun-nisa Hassanali, Su-Youn Yoon, Lei Chen

Fri-O1-3

In this study, we aim to automatically score the spoken responses from an international English assessment targeted to non-native English-speaking children aged 8 years and above. In contrast to most previous studies focusing on scoring of adult non-native English speech, we explored automated scoring of child language assessment. We developed automated scoring models based on a large set of features covering delivery (pronunciation and fluency), language use (grammar and vocabulary), and topic development (coherence). In particular, in order to assess the level of grammatical development, we used a child language metric that measures syntactic proficiency in emerging language in children. Due to acoustic and linguistic differences between child and adult speech, the automated speech recognition (ASR) of child speech has been a challenging task. This problem may increase difficulty of automated scoring. In order to investigate the impact of ASR errors on automated scores, we compared scoring models based on features from ASR transcriptions with ones based on human transcriptions. Our results show that there is potential for the automatic scoring of spoken non-native child language. The best performing model based on ASR transcriptions achieved a correlation of 0.86 with human-rated scores.
Automatic Prediction of Intelligibility of English Words Spoken with Japanese Accents – Comparative Study of Features and Models Used for Prediction

Teeraphon Pongkittiphan, Nobuaki Minematsu, Takehiko Makino, Daisuke Saito, Keikichi Hirose

Fri-O1-4

This study investigates automatic prediction of the words in given sentences that will be unintelligible to American listeners when they are pronounced with Japanese accents. The ERJ intelligibility database contains results of a large listening test, where 800 English sentences read with Japanese accents were presented to 173 American listeners and correct perception rate was obtained for each spoken word. By using this database, in our previous study, an intelligibility predictor was built for each word of input texts or utterances. For prediction, lexical and linguistic features were extracted from texts and pronunciation distance and word confusability were calculated from utterances. CART was used as prediction model. In this paper, new features that are related to speech prosody and three new prediction models of ensemble methods (Adaboost, Random Forest and Extremely Randomized Trees) are tested and compared to the old features and model. Finally, our new system can predict very unintelligible words and rather unintelligible words with F1-scores of 72.74% and 84.78%, respectively.

Word-level F0 Modeling in the Automated Assessment of Non-native Read Speech

Xinhao Wang, Keelan Evanini, Su-Youn Yoon

Fri-O1-5

This study investigates methods for automatically evaluating the appropriateness of F0 contours in the task of automated assessment of non-native read aloud speech. The F0 contour of a test taker’s spoken response is represented as a fixed-dimension vector with a word-level F0 value corresponding to each word in the prompt text. This vector is then correlated with gold standard vectors extracted from native speaker responses. Three different measures are used to describe the F0 contour within a word, including the mean of the F0 values, the difference between the mean values for each word and its neighbor-
Poster Session 1: From Pronunciation to Conversation
11:20–12:30, Friday, September 4, 2015

Towards a Conversational Expert System for Rhetorical and Vocal Quality Assessment in Call Center Talks

Mathias Walther, Baldur Neuber, Oliver Jokisch, Taïeb Mellouli
Fri-P1-1

This article presents the concept and development steps towards a conversational expert system for rhetorical and vocal quality assessment in call center talks. At first the state of the art in quality assessment is discussed. The influencing rhetorical and vocal factors are introduced. In our novel approach, the recognition of vocal factors is modeled by competing classification systems and combined into a multi-classifier system which is based on decision trees. Finally we propose an expert system which incorporates the generated decision rules. The system accuracy can be improved by user-adapted rule sets. Furthermore solutions to the problem of inconsistent rules are discussed.

Corpus-based Pronunciation Variation Rule Analysis for Singapore English

Wenda Chen, Nancy F. Chen, Boon Pang Lim, Bin Ma
Fri-P1-2

In this paper, we evaluate a set of linguistic rules for pronunciation variations in Singapore English. We collect and annotate a speech corpus for Singapore English and label it with IPA narrow transcriptions. Data driven pronunciation rules are derived using American English (Buckeye corpus) as a reference. We compare the data driven rules with linguistic rules proposed by phoneticians, and found that some pronunciation variations observed in Singapore English are also observed in American English, but with different frequencies of occurrence. Our analysis verifies the linguistic rules previously proposed for Singapore English and demonstrates an approach to utilizing our findings to improve pronunciation feedback.

Declarative and Interrogative Mandarin Intonation by Native Speakers and Cantonese L2 Learners

Wentao Gu, Lei Liu
Fri-P1-3

This study compared sentence intonation of L1 Mandarin by native speakers and L2 Mandarin by Cantonese learners, with both acoustic analysis and perceptual experiment. Three types of sentences (i.e., statement, intonation question, and particle question) ending with different tones and in different lengths were investigated. The perceptual experiment showed that declarative intonation in L2 speech was better identified than in L1 speech, which could be explained by the more prominent F0 declination in L2 speech. In contrast, interrogative intonation in L2 speech had a lower rate of identification than in L1 speech, and the differences in rate varied with the sentence-final tone. Acoustic analysis showed that global F0 raising for questions was weaker in L2 speech than in L1 speech, especially in longer sentences, while sentence-final F0 raising was relatively well maintained in L2 speech. Perceptual and acoustic studies showed consistent results on L2 intonation errors, which could be explained by the limited language abilities and language transfer effects.

ClassTranscribe: a New Tool with New Educational Opportunities for Student Crowdsourced College Lecture Transcription

Jia Chen Ren, Mark Hasegawa-Johnson, Lawrence Angrave
Fri-P1-4

ClassTranscribe is an open-source, web-based platform that leverages crowdsourcing to address the problem of accurate, reliable and fast transcriptions of college lectures. Completed transcriptions provide search functionality that augments existing lecture recordings and enable enhanced educational features including closed captioning.

Notes
Detection of Phone Boundaries for Non-native Speech Using French-German Models

Dominique Fohr, Odile Mella

Within the framework of computer assisted foreign language learning for the French/German pair, we evaluate different HMM phone models for detecting accurate phone boundaries. The optimal parameters are determined by minimizing on the non-native speech corpus the number of phones whose boundaries are shifted by more than 20 ms compared to the manual boundaries. We observe that the best performance was obtained by combining a French native HMM model with an automatically selected German native HMM model.
Demo Session:
14:00–15:00, Friday, September 4, 2015

An Open Platform that Allows Non-expert Users to Build and Deploy Speech-enabled Online CALL Courses
Manny Rayner, Claudia Baur, Pierrette Bouillon, Cathy Chua, Nikos Tsourakis

We demonstrate Open CALL-SLT, a framework which allows non-experts to design, implement and deploy online speech-enabled CALL courses. The demo accompanies two long papers also appearing at the SLaTE 2015 workshop, which describe the platform in detail.

A CAPT Tool for Training and Research on Lexical Stress Errors in German
Anjana Sofia Vakil

This demonstration presents de-stress: the German (de) System for Training and Research on Errors in Second-language Stress. This prototype Computer-Assisted Pronunciation Training (CAPT) tool provides a variety of options for diagnosis of and feedback on lexical stress errors, and could potentially be a useful component of an intelligent CAPT system.

Pinpointing the Difference – Visual Comparison of Non-native Speaker Groups
Florian Höning, Sebastian Wankerl, Anton Battiner, Elmar Nöth

We apply a tool originally developed for comparing pathological and healthy speakers to non-native speech. The method works on speakers who produce a given word sequence. Using time-alignment, we can display prototypical loudness contours, local tempo variations, and also spectrograms, together with information on variability and group effect size over time. The system, which will be made publicly available, is able to expose typical differences in a group of German and Italian speakers.

Development of a Prosodic Reading Tutor of Japanese – Effective Use of TTS and F0 Contour Modeling Techniques for CALL
Nobuaki Minematsu, Hiroya Hashimoto, Hiroko Hirano, Daisuke Saito

A text typed to a speech synthesizer is generally converted into its corresponding phoneme sequence on which various kinds of prosodic symbols are attached by a prosody prediction module. By using this module effectively, we build a prosodic reading tutor of Japanese, called Suzuki-kun, and it is provided as one function of OJAD (Online Japanese Accent Dictionary). In Suzuki-kun, by using a prosody prediction module, any Japanese text is converted into its reading (Hiragana sequence) on which the pitch pattern that sounds natural is visualized as a smooth curve drawn by the F0 contour generation process model. Further, positions of accent nuclei and unvoiced vowels are illustrated. Suzuki-kun also reads that text out following the prosodic features that are visualized. Since releasing Suzuki-kun, the number of accesses to OJAD has been drastically increased and for the last four months, OJAD received 129,168 accesses, 58.9% of which were from outside Japan.

Massive Pronunciation Training via Mobile Applications
Hui Lin

In this paper, we introduce a mobile application (app) that helps people practicing their English pronunciation using mobile devices. Equipped with an embedded assessment engine, the app offers accurate pronunciation assessment and feedback to a learner instantly. Moreover, game elements and mechanics are introduced to make the training experience fun, rewarding and engaging. The app turns out be phenomenal and leads to massive pronunciation training at an unprecedented scale. Since its launch, the app has accumulated more than 20 million users. Hundreds of years of speech data are collected from more
than 11 million different speakers, which is probably the largest speech corpus for Chinese spoken English in the world.

**Lingunia World of Learning**

*Karina Matthes, Rico Petrick, Horst-Udo Hain*

In this paper we present a soft toy named Lingufino for preschool children that uses speech input and output for communication. It takes the child onto a journey to an adventure world: Lingunia. Based on a story that is shown in a picture book the toy explains different topics like animals, colours, numbers, seasons etc. and involves the child into the fictional situation with the help of question-and-answer games. By asking for words and facts which previously have been mentioned, the child becomes part of the adventure and – on the fly – improves its active vocabulary.
Automatic Classification of Lexical Stress Errors for German CAPT

Anjana Sofia Vakil, Jürgen Trouvain

Lexical stress plays an important role in the prosody of German, and presents a considerable challenge to native speakers of languages such as French who are learning German as a foreign language. These learners stand to benefit greatly from Computer-Assisted Pronunciation Training (CAPT) systems which can offer individualized corrective feedback on such errors, and reliable automatic detection of these errors is a prerequisite for developing such systems. With this motivation, this paper presents an exploration of the use of machine learning methods to classify non-native German lexical stress errors. In classification experiments using a manually-annotated corpus of German word utterances by native French speakers, the highest observed agreement between the classifier’s output and the gold-standard labels exceeded the inter-annotator agreement between humans asked to classify lexical stress errors in the same data. These results establish the viability of classification-based diagnosis of lexical stress errors for German CAPT.

Self-imitation in Prosody Training: a Study on Japanese Learners of Italian

Elisa Pellegrino, Debora Vigliano

The proficiency in a second language is fully attained only if students have learnt to modulate the rhythmic and prosodic parameters equivalent to those of the native speakers. This study is aimed to test the pedagogical effectiveness of the selfimitation technique for the purpose of developing a native-like prosodic competence. Seven intermediate Japanese learners of Italian (NNSs) and 2 native Italian speakers (NSs) were involved in a read speech activity. NSs and NNSs were asked to read and record two Italian sentences conveying three different pragmatic functions (granting, order, request). NNSs performed the task twice, before and after the self-imitation prosodic training. The items used for the training were obtained by transferring the suprasegmental features of the native speakers, used as donors, to the Japanese learners, considered as the receivers. During the training phase, Japanese learners mimic their utterances previously modified to match the prosody of the reference native speaker, and then recorded the new performance. Seventeen native Italian listeners rated pre- and post-training productions for pragmatic function and accentedness. The results indicate that selfimitation promoted an improvement in learners’ performances in terms of communicative effectiveness. Conversely, average rate of accentedness does not change significantly before and after training.

Using Karaoke to Enhance Reading While Listening: Impact on Word Memorization and Eye Movements

Emilie Gerbier, Gérard Bailly, Marie-Line Bosse

This article reports the use of a karaoke technique to drive the visual attention span (VAS) of subjects reading a text while listening to the text spelled aloud by a reading tutor. We tested the impact of computer-assisted synchronous reading (S+) that emphasizes words when they are uttered, vs. non-synchronous reading (S-) in a reading while listening (RWL) task. Thirty-five 6th grade pupils read 12 stories, each involving one pseudoword presented four times, and each displayed in either condition. They were then unexpectedly tested on their memory for the orthography and for their acquired semantic knowledge of the pseudowords. Although no benefit was observed in the orthographic task, the synchronous condition significantly boosted the semantic memory by 10 points compared to the non-synchronous one (28% vs. 17% correct). We also provide some preliminary analysis on the gaze data collected during reading, suggesting differences between both conditions in terms of first fixation duration, fixation position on the word and onset delay relative to the corresponding speech onset.

Notes
This study introduces a tool, partial and synchronized caption (PSC), for training second language (L2) listening skill. PSC uses an automatic speech recognition (ASR) system to realize word-level alignment between text and speech while it refers to the corpora to effectively select a subset of words for inclusion in the caption. The selection criteria are based on three features contributing to L2 listening difficulties: speech rate, word frequency and specificity. Our findings reveal that PSC in its current state leads to the same level of comprehension as the full caption condition. PSC, however, outperforms the full caption when it comes to preparing learners for listening without using any textual clues as in real-life situations. To enhance this system the incorporation of other features is a necessity. However, the relationship between these factors and their contribution to listening difficulty is complex. This study conducts a root cause analysis on the ASR errors to better understand the underlying features that make recognition difficult for such systems and compares these features with L2 listening influential factors. Our preliminary analysis revealed an interesting similarity between features leading to L2 difficulty and those resulting in ASR errors. Such insightful findings shed light on the future improvements for PSC.

An Improved DNN-based Approach to Mispronunciation Detection and Diagnosis of L2 Learners’ Speech

Wenping Hu, Yao Qian, Frank K. Soong

We extend the Goodness of Pronunciation (GOP) algorithm from the conventional GMM-HMM to DNN-HMM and further optimize the GOP measure toward L2 language learners’ accented speech. We evaluate the performance of the new proposed approach at phone-level mispronunciation detection and diagnosis on an L2 English learners’ corpus. Experimental results show that the Equal Error Rate (EER) is improved from 32.9% to 27.0% by extending GOP from GMM-HMM to DNN-HMM and the EER can be further improved by another 1.5% to 25.5% with our optimized GOP measure. For phone mispronunciation diagnosis, by applying our optimized DNN based GOP measure, the top-1 error rate is reduced from 61.0% to 51.4%, compared with the original DNN based one, and the top-5 error rate is reduced from 8.4% to 5.2%. On a continuously read, L2 Mandarin learners’ corpus, our approaches also achieve similar improvements.
Becoming Literate While Learning a Second Language – Practicing Reading Aloud

Catia Cucchiarini, Mario Ganzeboom, Joost van Doremalen, Helmer Strik

Sat-O3-1

The DigLin project aims at providing concrete solutions for low-literate and illiterate adults who have to learn a second language (L2). Besides learning the L2, they thus also have to acquire literacy in the L2. To allow intensive practice and feedback in reading aloud, appropriate speech technology is developed for the four targeted languages: Dutch, English, German and Finnish. Since relatively limited resources are available for this application for the four studied languages, this had to be taken into account while developing the speech technology. Exercises with suitable content were developed for the four languages, and are tested in four countries: Netherlands, United Kingdom, Germany, and Finland. Preliminary results are presented in the paper, and suggestions for future directions are discussed.

Supervised Learning of Response Grammars in a Spoken Call System

Manny Rayner, Claudia Baur, Cathy Chua, Nikos Tsourakis

Sat-O3-2

We summarise experiments carried out using a system-initiative spoken CALL system, in which permitted responses to prompts are defined using a minimal formalism based on templates and regular expressions, and describe a simple structural learning algorithm that uses annotated data to update response definitions. Using 1,927 utterances of training data, we obtained a relative improvement of 20% in the system’s ability to react differentially to correct and incorrect input, measured on a previously unseen test set. The results are significant at p < 0.005.

Automatic Detection of Grammatical Structures from Non-native Speech

Suma Bhat, Su-Youn Yoon, Diane Napolitano

Sat-O3-3

This study focuses on the identification of grammatical structures that could serve as indices of the grammatical ability of non-native speakers of English. We obtain parse trees of manually transcribed non-native spoken responses using a statistical constituency parser and evaluate its performance on noisy sentences. We then use the parse trees to identify the grammatical structures of the Index of Productive Syntax (IPSyn), previously found useful in evaluating grammatical development in the context of native language acquisition. Empirical results of this study show: a) parsing ungrammatical sentences using a probabilistic parser suffers some degradation but is still useful for further processing; and b) automatic detection of the majority of the grammatical structures measured by IPSyn can be performed on non-native adult spoken responses with recall values more than 90%. To the best of our knowledge, this is the first study which explores the relationship between parser performance and the automatic generation of grammatical structures in the context of second language acquisition.
Modeling Pronunciation Variations for Non-native Speech Recognition of Korean Produced by Chinese Learners

Seung Hee Yang, Minsoo Na, Minhwa Chung

Sat-O4-1

Recognition accuracy for non-native speech is often too low to make practical use of ASR technology in interfaces such as CAPT systems. This paper describes how we adapted Korean ASR system to Chinese speakers for building a Korean CAPT system for L1 Mandarin Chinese learners by modeling pronunciation variations frequently produced by Chinese learners. Based on pronunciation variation rules describing substitutions, insertions, and deletions together with phonological knowledge rules realized in different phonemic contexts, the probability of occurrence of each rule is calculated. These rules are used to generate extended pronunciation lexicon. For each learner level, ASR experiment is conducted, where 21.2\% relative WER reduction is obtained. This verifies that variation analysis is useful for modeling Korean produced by Chinese learners.

Analysis of Phone Confusion Matrices in a Manually Annotated French-German Learner Corpus

Denis Jouvet, Anne Bonneau, Jürgen Trouvain, Frank Zimmerer, Yves Laprie, Bernd Möbius

Sat-O4-3

This paper presents an analysis of the non-native and native pronunciations observed in a phonetically annotated bilingual French-German corpus. After a forced-choice automatic annotation a large part of the corpus was checked and corrected manually on the phone level which allows a detailed comparison of the realized sounds with the expected sounds. The analysis is reported in terms of phone confusion matrices for selected error-prone classes of sounds. It revealed that German learners of French have most problems with obstruents in word-final position whereas French learners of German show complex interferences with the vowel contrasts for length and quality. Finally, the correct pronunciation rate of the sounds, for several phonetic classes, is analyzed with respect to the learner’s level, and compared to native pronunciations. One outcome is that different sound classes show different correct rates over the proficiency levels. For the German data the frequently occurring syllabic [ːn] is a prime indicator of the proficiency level.
Helping Non-expert Users Develop Online Spoken CALL Courses

Manny Rayner, Claudia Baur, Pierrette Bouillon, Cathy Chua, Nikos Tsourakis

Sat-P2-1

We introduce Open CALL-SLT, a Web 2.0 framework which allows non-experts to design, implement and deploy online speech-enabled CALL courses. Course functionality is divided into six increasingly sophisticated levels; the lowest levels assume only basic web-literacy, while the higher ones require some acquaintance with simple software concepts like regular expressions and XML. We describe the different levels of functionality and the deployment process, which permits multiple developers to compile and run courses on a set of shared servers. The framework has recently been opened up for alpha testing, and we briefly summarize early experiences.

Integrating Acoustic and State-transition Models for Free Phone Recognition in L2 English Speech Using Multi-distribution Deep Neural Networks

Kun Li, Xiaojun Qian, Shiyong Kang, Pengfei Liu, Helen Meng

Sat-P2-2

This paper investigates the use of Multi-Distribution Deep Neural Networks (MD-DNNs) for integrating acoustic and statetransition models in free phone recognition of L2 English speech. In Computer-Aided Pronunciation Training (CAPT) system, free phone recognition for L2 English speech is the key model of Mispronunciation Detection and Diagnosis (MDD) in the cases of allowing freely speaking. A simple Automatic Speech Recognition (ASR) system can be approached with an Acoustic Model (AM) and a State-Transition Model (STM). Generally, these two models are trained independently, hence contextual information may be lost. Inspired by the Acoustic-Phonological Model, which achieves greatly improvements by integrating the AM and Phonological Model (PM) in MDD for the case that L2 learners practice their English by following the prompts, we propose a joint Acoustic-State-Transition Model (ASTM) which uses a MD-DNN to integrate the AM and STM. Preliminary experiments with basic parameter configurations show that the ASTM obtains a phone accuracy of about 68% on the TIMIT data. It is better than the system of using separate AM and STM, whose accuracy is only about 52%. Further finetuning the ASTM achieves an accuracy of about 72% on the TIMIT data. Similar performance is obtained if we train and test the ASTM on our L2 English speech corpus (CU-CHLOE).

Implementation and Test of a Serious Game Based on Minimal Pairs for Pronunciation Training

David Escudero-Mancebo, Enrique Cámara-Arenas, Cristian Tejedor-García, César González-Ferreras, Valentín Cardeñoso-Payo

Sat-P2-3

This paper introduces the architecture and interface of a serious game intended for pronunciation training and assessment for Spanish students of English as second language. Users will confront a challenge consisting in the pronunciation of a minimal-pair word battery. Android ASR and TTS tools will prove useful in discerning three different pronunciation proficiency levels, ranging from basic to native. Results also provide evidence of the weaknesses and limitations of present-day technologies. These must be taken into account when defining game dynamics for pedagogical purposes.

Rapid Development of Public Health Education Systems in Low-literacy Multilingual Environments: Combating Ebola through Voice Messaging

Nikolas Wolfe, Juneki Hong, Agha Ali Raza, Bhiksha Raj, Roni Rosenfeld

Sat-P2-4

One of the main challenges in combating the spread of the Ebola outbreak in West Africa is a lack of ef-
effective public health education among affected populations in Guinea, Sierra Leone, and Liberia. Difficulties include resistance to official sources of information, mistrust of government, cultural norms, linguistic barriers, and illiteracy. In this paper we describe the development and initial deployment of a voice-based, multilingual mobile phone application to spread reliable public health information about Ebola via peer-to-peer sharing. Our hypothesis is that we can overcome mistrust and disseminate important health information via the power of social learning and suggestion from friends, family, and local communities. In collaboration with partners on the ground in Conakry, Guinea, we have launched two parallel mobile phone services known as Polly Game and Polly Health to enable message sharing in several Guinean languages. We discuss a variety of strategies we have tried to encourage the spread of the application and data on uptake to date.

Web-based Mini-games for Language Learning that Support Spoken Interaction
Helmer Strik, Luigi Palumbo, Febe de Wet, Catia Cucchiarini

The European ‘Lifelong Learning Programme’ (LLP) project ‘Games Online for Basic Language learning’ (GOBL) aimed to provide youths and adults wishing to improve their basic language skills access to materials for the development of communicative proficiency in Dutch, French, and English through web-based mini-games. These mini-games were tested in four countries: The Netherlands (Dutch), Belgium (French), United Kingdom and South-Africa (English). Four types of mini-games were developed, and in two of them users can use ‘automatic speech recognition’ (ASR) to support spoken interaction. In the current paper we will focus on the English versions of these two games that were tested in the United Kingdom and South-Africa. The analyses that are presented in this paper were conducted to determine what users’ perceptions are about mini-games with and without speech input and ASR and which aspects of the speech-enhanced games are strongly related to each other.

Inter-annotator Agreement for a Speech Corpus Pronounced by French and German Language Learners
Odile Mella, Dominique Fohr, Anne Bonneau

This paper presents the results of an investigation of inter-annotator agreement for the non-native and native French part of the IFCASL corpus. This large bilingual speech corpus for French and German language learners was manually annotated by several annotators. This manual annotation is the starting point which will be used both to improve the automatic segmentation algorithms and derive diagnosis and feedback. The agreement is evaluated by comparing the manual alignments of seven annotators to the manual alignment of an expert, for 18 sentences. Whereas results for the presence of the devoicing diacritic show a certain degree of disagreement between the annotators and the expert, there is a very good consistency between annotators and the expert for temporal boundaries as well as insertions and deletions. We find a good overall agreement for boundaries between annotators and expert with a mean deviation of 7.6 ms and 93% of boundaries within 20 ms.

Evaluating Synthetic Speech in an Irish CALL Application: Influences of Predisposition and of the Holistic Environment
Neasa Ní Chiaráin, Ailbhe Ní Chasaide

This paper reports an evaluation of Irish synthetic voices in the context of a virtual reality CALL application. In addition to eliciting subjects’ ratings of the synthetic voices, the evaluation focusses particularly on (1) the extent to which prior attitudes to synthetic voices affected users’ satisfaction ratings and (2) the extent to which reactions to the synthetic voices were influenced by users’ engagement with the other (non-speech) dimensions of the CALL application. The particular application, Fáilte go TCD, was developed specifically for this purpose and uses virtual reality scenes where the animated characters converse in Irish (synthetic voices). Evaluations were carried out using Likert scale-based questionnaires.
Results showed broadly positive ratings of the Irish synthetic voices in terms of intelligibility, quality and naturalness. They further indicate that (1) users’ prior attitudes towards synthetic speech had a major influence on their reaction to the CALL application and (2) satisfaction levels with the synthetic voices are highly correlated with the rating accorded to other, non-speech dimensions of the platform, suggesting that the different aspects are judged in a holistic way.

Linking MOOC Courseware to Accommodate Diverse Learner Backgrounds

Shang-Wen Li, Victor Zue

Massive Open Online Courses (MOOCs) brings great opportunities to millions of learners. However, the size of the learner population and the heterogeneity of the learners’ backgrounds make conventional one-size-fits-all pedagogy insufficient. For example, learners lacking in prior knowledge may struggle with different concepts. In this paper, we propose a framework - educational content linking, to address the challenges. By linking and organizing scattered educational materials for a given MOOC into an easily accessible structure, this framework can provide guidance and recommendation of these contents, as well as improve navigation. Thus, learners can select appropriate supporting materials to suit their individualized needs and achieve self-exploring remediation. This paper describes an end-to-end case study, which found that learners, especially novices, can search learning materials faster without sacrificing accuracy, and can retain concepts more readily with our proposed approach. We have also obtained encouraging preliminary results that suggest that content linking can be achieved automatically using human language technology and stochastic modeling techniques.

Notes
Annotating Meta-discourse in Academic Lectures from Different Disciplines

Ghada Alharbi, Raymond W. M. Ng, Thomas Hain
Sat-O5-1

The use of discourse structure was shown to be effective in various applications. Meta-discourse is often used as an expression to signal discourse structure. Previous work focused on using the meta-discourse structure in written texts, or spoken material in very clean conditions. This paper presents a metadiscourse annotated corpus in a more challenging educational context. The corpus comprises of academic lectures from two different disciplines: physics and economics. The schema used focuses on five categories: Introduction, Conclusion, Previewing, Reviewing and Enumerating. The annotation task is described in detail, including instructions and strategies used by expert annotators. Annotation results are reported in terms of inter-annotator agreement, self-reported confidence and number of occurrences. Results show that meta-discourse is frequently used in academic lectures and this is observed in the two selected disciplines. Further analysis of the corpus is conducted showing that some of these categories, namely Introduction and Previewing, are correlated with labelled topic boundaries, which is also consistent in both disciplines. This finding shows the potential for using meta-discourse information in topic segmentation task.

German Phonics Game Using Speech Synthesis – A Longitudinal Study about the Effect on Orthography Skills

Kay Berkling, Nadine Pfauumer, Rémi Lavalley
Sat-O5-2

Acquisition of orthography is an important problem in German elementary schools. Today, few, if any, schoolbooks can claim to use knowledge of the deep syllable structure of German and its patterns for explicit orthography instruction. To fill this gap, a game described here allows children to explore the complete complexity of the German syllable patterns in analogy to phonics instruction used for English. The game is deployed on iPad and uses the Apple speech synthesis to allow children to listen to all possible letter combinations for legal German syllable structures. Through the explicit teaching of contextual Grapheme-Sound interaction, children are expected to generalize to new words in their writings that are automatically evaluated using speech technology. In the study presented here, 16 children participated in a weekly after-school session of one hour to play the German phonics game. The difference in achievement between the students and their control group were noticeably reduced by the end of the study.

Text-to-speech Enhanced eBooks for Emerging Literacy Development

Febe de Wet, Laurette Marais, Daleen Klop
Sat-O5-3

The purpose of this study was to measure the efficacy of an eBook to improve the vocabulary and word recognition skills in an Afrikaans speaking group of lower socio-economic status of 6- to 7-year old children with poor vocabulary. The main goals were to investigate if exposure to an interactive eBook would result in the acquisition of new vocabulary and sight word reading in the study participants. A randomised pre-test/post-test between-subjects design was used. An experimental group that received an intervention was compared to a control group before the control group received a delayed intervention. Both groups were reassessed, eight weeks after the interventions to assess the retention of their newly acquired skills. Results show a significant improvement in recognition and vocabulary skills in the experimental group compared to their initial assessments, as well as compared to the control group.