Social and Linguistic Speech Prosody

Proceedings of the 7th international conference on Speech Prosody

SpeechProsody 7

(Trinity College Dublin) May 20-23, 2014

ISSN: 2333-2042

Social and Linguistic Speech Prosody

1 Frontmatter/Preface

1.1 Statistics by Country (showing number of authors)

Authors from 45 countries sent in submissions to Speech Prosody 2014
5 countries didn't make it - we hope they'll try again for SP8!

1.2 Accepted authors by country:

Algeria	1
Australia	6
Austria	1
Bangladesh	1
Belgium	6
Brazil	17
Canada	12
China	16
Costa Rica	1
Czech Republic	7
Denmark	1
Estonia	9
European Union	281
Finland	8
France	59
Germany	70
Hong Kong	6
Hungary	26
India	4
Iran	1
Ireland	12
Israel	3
Italy	15
Japan	29
Mexico	1
Netherlands	16
Norway	1
Poland	8
Portugal	6
Qatar	1
Russian Federation	2
Saudi Arabia	1
Slovakia	3
South Africa	2
Spain	18
Swaziland	1
Sweden	7
Switzerland	12
Taiwan	7
UK	25
USA	76

1.3 Acceptance rates:

We have seen a 35% increase in acceptances since Speech Prosody in Nara, 10 years ago, and a 68% increase in submissions.

2004 (in Nara): 164/180; 29 oral and 135 poster
2014 (in Dublin): 222/303; 42 oral and 180 poster

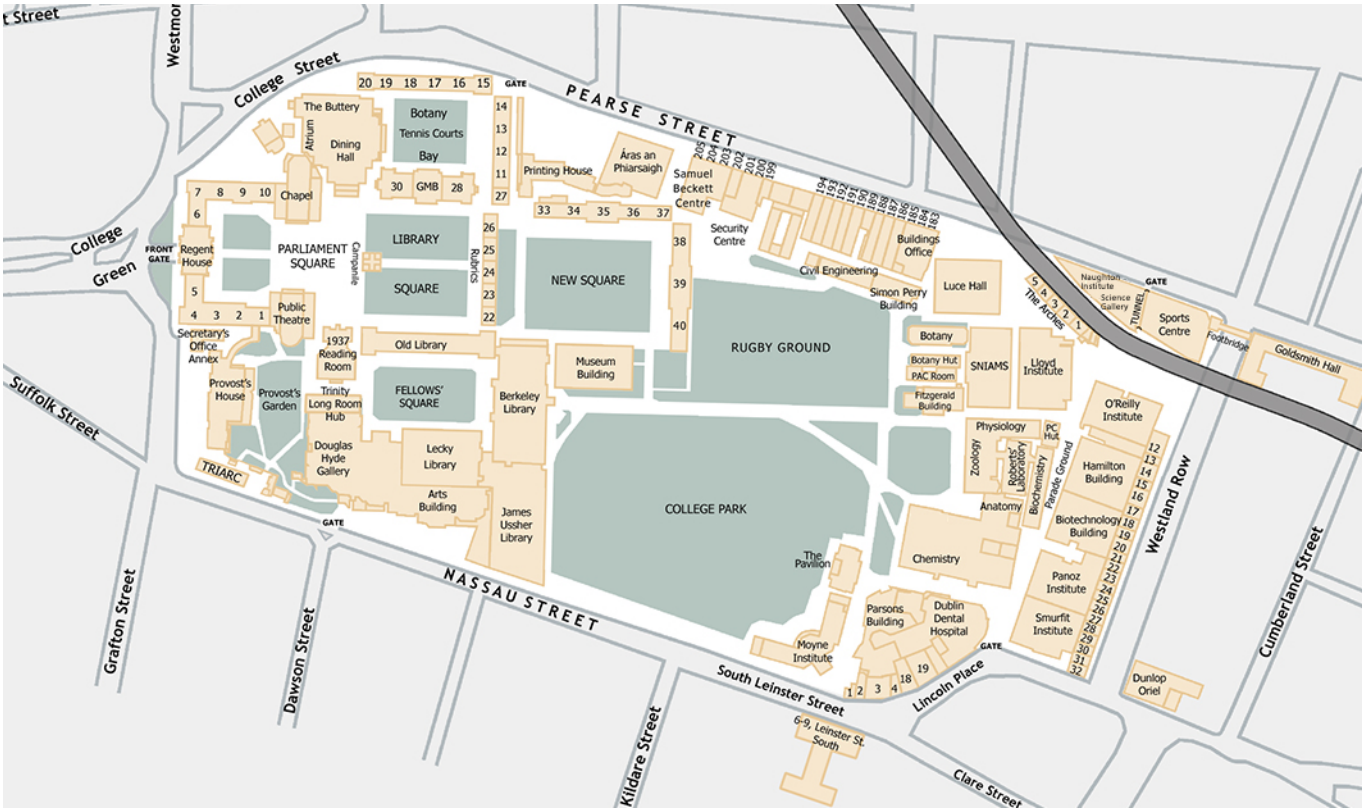
2004: 91% acceptance rate
2014: 73% acceptance rate

2004: 21% oral/poster ratio
2014: 23% oral/poster ratio

altogether 480 authors and 4 keynote speakers are represented at SP7

Speech Prosody - brought to you in Dublin by the Pros Bros!

(photo courtesy of Jolanta Bachan, Brighton 2012)



Finding us!

Speech Prosody 2014 will be hosted in the 'Ed Burke' Theatre, deep in the Arts Building of Trinity College Dublin (*the* University of Dublin ;-)

Trinity College is an academic island of quiet history right in the centre of Dublin City (the Aircoach stops just outside) and the Arts Building has an entrance in Nassau Street, facing Dawson Street but it is much nicer to enter from Fellows' Square (through Front Gate) and enjoy the old College

(please don't forget to visit the Long Room (Old Library) while you are here ;-)

and actually, while you're here, the Chester Beatty Library in Dublin Castle is also well worth a visit

and while we're on the subject of libraries, Marsh's Library is offering a great exhibition on Japan!

. . . . . but this is all for Saturday - we hope you'll be with us full-time until then . . .

An essay on 'ProsBros' (1st version, DG)

'ProsBros', alternatively 'Pros-Bros', 'Pros Bros' … what's that? By induction from context and some help from Monty Python, you will have concluded that 'ProsBros' is the name of a set:

ProsBros = {Nick, Daniel, Dafydd}

That's it, a nickname for the three of us. (Why NICKname, by the way? Not fair.) So why 'ProsBros'? Analogical thinking will no doubt yield the tentative hypothesis that 'Pros' is an abbreviation for 'prosody'. Well done, correct! But 'Bros'? Analogical thinking may lead you to think that it is the plural of 'Bro', which is an abbreviation of 'brother', almost correct, though as you will see it is actually a direct abbreviation for 'brothers'. Now, having clarified the extensional semantics and the morphology, the intensional semantics, pragmatics and phonetics remain to be clarified.

First, semantics. Yes, we wrote our dissertations on prosody (Daniel and Dafydd on intonation, Nick on timing), and have since continued to work in the field, with occasional deviations (the most deviant being Dafydd). Yes, we have all worked with computational phonetic tools, yes, we have all worked extensively with speech corpora (the most well-known being Dan's), yes, we have all worked with speech synthesis (most of all, Nick).

Second, pragmatics. Yes, we are roughly the same generation. Yes we have been friends for decades. We are three expat Brits (hence the undeniable influence of Monty Python on this note) who have each worked mainly on other languages than English: Nick on Japanese, Dan on French, and Dafydd on German (and a collection of African and Asian languages). We have been heavily involved in creating and supporting international infrastructures in these fields: Speech Prosody, COCOSDA, international project consortia. Ironically, we have been often been vicarious representatives of these speech communities in committees and conferences: "What is the Japanese perspective on this, er, Nick?" ... "What is the French perspective on this, er, Dan?" ... "What is the German perspective on this, er, Dafydd?"

Third, phonetics. Yes this one of the main areas in which we work. Now note that 'Pros' and 'Bros' could both rhyme with 'Oz', 'boss', 'rose' or 'gross', yielding 16 possible combinations. Following Occam's Razor, we reject 'rose' and 'gross' as too complex (alternatively: too emotional), leaving 4 combinations, and to restrict the search space we propose a new markedness constraint 'Disyllabic Nickname Rhyme Harmony' (*DNRH) in Optimal Nickname Theory:

**XAYB*, where onset(*X*), onset(*Y*), rhyme(*A*), rhyme(*B*), for *A*≠*B*

The *DNRHC permits only [prɔsbrɔs] and [prɔzbrɔz] (ignoring other segmental details). Controversially, in acknowledgment of a plethora of languages with final devoicing, [prɔsbrɔs] is marginally preferred to [prɔzbrɔz], and together with the English Compound Stress Rule, ['prɔsbrɔs] emerges as the favoured pronunciation, though ['prɔzbrɔz] is a close second.

However, the selection is finally clinched by further analogical thinking, which will initially only be accessible to Brits. So a little cultural history: in 1898 the legendary sartorial hire business 'Moss Bros', well known throughout the UK and beyond, was established in London by the brothers Alfred and George Moss, who deserve the Noble Prize [sic] for achieving the remarkable goal of ensuring that businessmen worldwide (and some businesswomen) wear the same style of suit, shirt and tie. No, we do not normally wear these suits, shirts and ties, but we were strongly influenced by the name of the business (and, as noted above, by Monty Python).

Now the gentle reader may wish to face the challenge of a final exercise in induction and analogical thinking: How is 'Moss Bros' pronounced?

Programme

Nick Campbell	Trinity College Dublin, the University of Dublin, Ireland
Dafydd Gibbon	Universität Bielefeld, Germany
Daniel Hirst	CNRS & Université de Provence, France

International Advisory Committee

Paavo Alku	Aalto University
Véronique Aubergé	Grenoble LIG
Christophe D'Alessandro	LIMSI-CNRS
Plinio Barbosa	University of Campinas
Fred Cummins	University College Dublin
Grazyna Demenko	Adam Mickiewicz University
Hongwei Ding	Tongji University
Jens Edlund	Royal Technical Institute (KTH)
Mária Gósy	Hungarian Academy of Sciences
Mark Hasegawa-Johnson	University of Illinois at Urbana-Champaign
Keikichi Hirose	University of Tokyo
Oliver Jokisch	Leipzig University of Telecommunication
Haruo Kubozono	National Institute for Japanese Language and Linguistics
Philippe Martin	Université Paris Diderot
Hansjörg Mixdorff	Beuth University of Applied Sciences
Bernd Möbius	Dept. of Comp. Ling. and Phonetics, Saarland University
Toshiyuki Sadanobu	Kobe University
Yoshinori Sagisaka	Waseda University
Jan van Santen	Center for Spoken Langauge Processing
M.G.J. Swerts	Tilburg University, School of Humanities
Jürgen Trouvain	Saarland University
Khiet Truong	University of Twente
Chiu-Yu Tseng	Institute of Linguistics, Academia Sinica
Petra Wagner	Universität Bielefeld
Nigel Ward	University of Texas at El Paso
Yi Xu	University College London

Special Session Organisers

Hiroya Fujisaki	University of Tokyo
Toshiyuki Sadanobu	Kobe University
Véronique Aubergé	Laboratory of Informatics of Grenoble (LIG)
Marzena Żygis	Zentrum für Allgemeine Sprachwissenschaft, Berlin
Zofia Malisz	Universität Bielefeld

Programme Committee

IAC (see above)	**all IAC members were active reviewers**
Noam Amir	Tel Aviv university
Bistra Andreeva	Department of Computational Linguistics and Phonetics
Amalia Arvaniti	University of Kent
Véronique Aubergé	LIG Grenoble
Cinzia Avesani	ISTC-CNR
Anton Batliner	Lehrstuhl fuer Mustererkennung
Stefan Baumann	IfL Phonetik, Cologne University
Pier Marco Bertinetto	Scuola Normale Superiore
Roxane Bertrand	Laboratoire Parole et Langage, UMR 6057 CNRS

Maria Paola Bissiri	Technische Universität Dresden
Antonio Bonafonte	UPC
Francesca Bonin	Trinity College Dublin
Genevieve Caelen-Haumont	MICA laboratory
Aoju Chen	Utrecht University
Sin-Horng Chen	National Chiao Tung University
Robert Clark	The University of Edinburgh
Jennifer Cole	University of Illinois
Ricardo Cordoba	Grupo de Tecnologia del Habla, Madrid
Snezhina Dimitrova	University of Sofia
Elisabeth Delais-Roussarie	CNRS-Université Paris 7 Paris Diderot,
Gorka Elordieta	University of the Basque Country
John Esling	University of Victoria
Sascha Fagel	zoobe message entertainment GmbH
Zsuzsanna Fagyal-Le Mentec	University of Illinois at Urbana-Champaign
Isabel Falé	Universidade Aberta/CLUL
Janet Fletcher	School fo Languages & Linguistics University of Melbourne
Sónia Frota	Universidade de Lisboa
Hiroya Fujisaki	University of Tokyo
Dafydd Gibbon	Universität Bielefeld
Matt Gordon	UC Santa Barbara
Björn Granström	KTH, Sweden
Carlos Gussenhoven	Radboud University, Nijmegen
Mária Gósy	Research Institute for Linguistics, HAS
Sophie Herment	Université de Provence
Daniel Hirst	CNRS & Université de Provence
Merle Horne	Lund University
David House	KTH, Sweden
Jill House	University College London
Sarmad Hussain	CLE-KICS, UET
Ignasi Iriondo	Enginyeria i Arquitectura La Salle. Universitat Ramon Llull
Stefanie Jannedy	Center for Linguistics (ZAS)
Sun-Ah Jun	UCLA
Maciej Karpiński	Adam Mickiewicz University
Hideki Kawahara	Wakayama University
Tatsuya Kawahara	School of Informatics, Kyoto University, Kyoto, Japan
Roland Kehrein	Uni Marburg, Germany
Esther Klabbers	OHSU
Jody Kreiman	University of California, Los Angeles
Frank Kügler	Potsdam University
Haizhou Li	Institute for Infocomm Research
Yuan-Fu Liao	National Taipei University of Techonology
Joaquim Llisterri	Universitat Autònoma de Barcelona
Madureira	PUCSP
Zofia Malisz	Universität Bielefeld
Piet Mertens	K.U.Leuven
Nobuaki Minematsu	University of Tokyo
Helena Moniz	FLUL/INESC-ID
Hiroki Mori	Utsunomiya University
Shrikanth Narayanan	University of Southern California
Eva Navas	University of the Basque Country
Oliver Niebuhr	Dept. of General Linguistics, University of Kiel
Elmar Nöth	University of Erlangen-Nuremberg
Michael O'Dell	University of Tampere
John Ohala	University of California, Berkeley
Zdena Palkova	Institute of Phonetics, Charles University Prague
Prem C. Pandey	Indian Institute of Technology Bombay
Gabor Pinter	Kobe University
Bernd Pompino-Marschall	HU Berlin
Heather Pon-Barry	Arizona State University
Cristel Portes	Universite de Provence

Brechtje Post	University of Cambridge
Hugo Quené	Utrecht University
César Reis	Universidade Federal de Minas Gerais
Eduardo Rodriguez Ba]nga	University of Vigo
Daisuke Saito	UNversity of Tokyo
Elina Savino	University of Bari
Amy Schafer	University of Hawaii
Antje Schweitzer	Stuttgart University
Jane Setter	University of Reading
Chilin Shih	University of Illinois at Urbana-Champaign
Elizabeth Shriberg	SRI International
Miquel Simonet	University of Arizona
Shari Speer	Ohio State University
Marc Swerts	Tilburg University
Jianhua Tao	Chinese Academy of Sciences
Shu-Chuan Tseng	Institute of Linguistics, Academia Sinica
Alice Turk	University of Edinburgh
Vincent Van-Heuven	University of Leiden
Nanette Veilleux	Simmons College
Céu Viana	inesc-id, Portugal
Yue Wang	Simon Fraser University
Duane Watson	University of Illinois Urbana-Champaign
Stefan Werner	University of Eastern Finland
Laurence White	Plymouth University
Marcin Wlodarczak	Universität Bielefeld
Chai Wutiwiwatchai	Human Language Technology Laboratory, NECTEC
Jiahong Yuan	University of Pennsylvania

Assistant Reviewers

Amengual, Mark
Arnold, Denis
Batista, Fernando
Casillas, Joseph
Correia, Susana
Jügler, Jeanin
Rohena-Madrazo, Marcos
Šimko, Juraj
Steiner, Ingmar
Wang, Yang

Local Organising Committee

Mai & Sarah & Lucy @ Odyssey

Odyssey International Incentives & Meetings
6-8 Garville Lane, Rathgar, Dublin 6, Ireland

Tel : + 353 1 497 4866
Fax : + 353 1 496 1396

with sincere thanks also to Fáilte Ireland and the SFI for their kind help!

Speech Prosody - a brief history

(according to the pros bros)

The first international meeting on Speech Prosody that we know of was a three-day seminar on Intonation and Discourse organised by the British Association for Applied Linguistics in Birmingham in April 1982. Strangely enough, the second meeting was held just two weeks later in Paris: a workshop on Prosody organised by the European Association for Psycholinguistics. This was soon followed by a Working Group on Intonation that was a satellite event preceding the 13th International Congress of Linguists in Tokyo.

And then, after that, nothing for more than ten years...

Before the next prosody meeting, the 12th ICPhS meeting was held in Aix-en-Provence in August 1991 where we were struck by the large number of papers on the topic of Speech Prosody - more than 20% of the papers which were directly related to this topic and the small room assigned was overflowing into the corridors! Dan asked Mario Rossi, the conference chairman, to announce an ad-hoc meeting before one of the plenary sessions and over 100 people turned up, no doubt wondering what to expect. As it happened we had no more idea than they did but among suggestions made then were: setting up an international organisation - organising conferences on prosody - and setting up a prosody mailing list.

George Allen volunteered to set up an email list for prosody and this list was a valuable resource for many years, although the fact that it included both literary prosody (versification) and speech prosody was slightly confusing for some people. There would be a flurry of engineers signing off the list after a posting on mediaeval versification, followed by an equally urgent flurry of linguists quitting the prosody ship after an engineer had posted something on hidden Markov models for speech recognition. We eventually decided to replace the list by one specifically devoted to Speech Prosody.

In the years following this meeting, ICSLP (International Conference on Spoken Language Processing) and ESCA (European Speech Communication Organisation) organised some more workshops on prosody (Lund 1993, Yokohama 1994, Athens 1995, Kraków 1999) and workshops on prosody also became a regular feature of the ICPhS meetings (Stockholm 1995, San Francisco 1999). Despite this welcome increase in the number of meetings, there was still something missing. From one meeting to the next there was no way of telling when or where the next meeting would be held. Each meeting was organised as a separate event with no co-ordination. This made it hard for potential participants to plan to present a paper on prosody. It was particularly true for doctoral students, who had no way of knowing if there would be an appropriate meeting somewhere where they could present their research before the end of their thesis preparation.

In September 1999, ESCA and ICSLP agreed to combine and were renamed ISCA (International Speech Communication Association) and the new organisation soon set up Special Interest Groups to promote research on specific topics of speech communication or for specific languages. With the wider availability of internet, it was now quite easy to contact a large number of specialists on Speech Prosody, asking if they would agree to support the creation of a Special Interest Group on Speech Prosody (SPròSIG). The response was enthusiastic - 72 established researchers in the field from 23 different countries agreed to support the SIG and in January 2000 the group was recognised by ISCA.

The 'Prosody 2000' Kraków meeting, chaired by Wiktor Jassem, was a kind of Proto-Speech Prosody, actually a combination of two workshops - one on Speech Recognition and Synthesis, organised by our prosody sister Grazyna Demenko with the help of Dafydd Gibbon, and one on Prosodic Transcription and Modelling organised by Esther Grabe, Kai Alter and Hansjörg Mixdorff.

The ISCA/SPròSIG event, the First International Conference on Speech Prosody, Aix-en-Provence, April 3 2002 chaired by Daniel Hirst, was a success with 152 submitted papers plus 6 invited keynote speakers, 12 invited co-speakers. It was attended by 317 people. Of course the question everybody asked was: would it last? The list of Speech Prosody meetings speaks for itself: 2004 Nara Japan, 2006 Dresden Germany 2008, Campinas Brazil, 2010 Chicago USA, 2012 Shanghai China, 2014 Dublin, Ireland; and the next will be in 2016 - but where ???

A modification of the constitution of SPròSIG in 2010 added a Permanent Advisory Committee, consisting of the founder officers of SPròSIG (Daniel Hirst, Nick Campbell, Bernard Bel), the current elected officers (currently Keikichi Hirose, Yi Xu, Mark Hasegawa-Johnson, Hansjorg Mixdorff) as well as the chair and co-chair of the last 5 conferences.

At the last meeting in Shanghai (2012), the PAC decided to nominate Hiroya Fujisaki as Honorary Life Member of the PAC. The 2011 board meeting of the International Phonetic Association resolved to co-sponsor the Speech Prosody Special Interest Group - this resolution was adopted formally in October 2012. The Speech Prosody SIG is, consequently, now officially affiliated with both ISCA and IPA.

Speech Prosody is too serious a matter to be left in the hands of just engineers... ... or just linguists. We really need each other, and Speech Prosody is a way to bring us all together!

SP7 - another SPròSIG event - with special thanks to :

The Team:

EasyChair:

Odyssey:

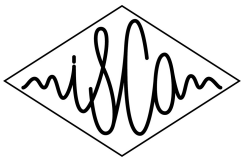
Science Foundation Ireland:

The University of Dublin:

Fáilte Ireland:

SPròSIG:

ISCA



IPA

and YOU!

Daily Event Programme

Speech Prosody will feature 222 papers in plenary oral and poster format: there will be 4 invited keynotes each followed by 3 oral presentations on a related theme, and 3 oral thematic sessions of 6 papers each. Each day will have a poster session and posters can remain in place throughout the day. There is no distinction in rank between an oral and a poster presentation. There will be 2 special sessions, a Round Table, and a Panel Discussion, with a special commemoration of Prof Miyoko Sugito, the honorary co-chair of Speech Prosody 2004, to remember her contributions to our field. Lunch will be provided each day.

Day-1.1 Tuesday May 20th, 1:30pm - 2pm : Opening (3 bros:-welcome!)
Day-1.2 Tuesday May 20th, 2pm - 3:30pm : plenary (1+3 oral) Invited Keynote
Day-1.3 Tuesday May 20th, 4pm - 5:30pm : special (Booster & Round Table)
Day-1.4 Tuesday May 20th, 5:30pm - 6pm : SSSpASSS (20 Special Session posters)
Day-1.5 Tuesday May 20th, 6pm - 7:30pm : (28 Posters with **WELCOME RECEPTION**)

Day-2.1 Wednesday May 21st, 9am - 10:30am : plenary (1+3 oral) Invited Keynote
Day-2.2 Wednesday May 21st, 11am - 1pm : (6 oral) - Speech Rhythm and Timing
Day-2.3 Wednesday May 21st, 2pm - 4pm : special session - Slavic Prosody
Day-2.4 Wednesday May 21st, 4:30pm - 6pm : (48 posters) Perception and Intonation
Day-2.5 Wednesday - evening - **Reviewers' Reception** in Trinity College Long Room

Day-3.1 Thursday May 22nd, 9am - 10:30am : 3-1-plenary (1+3 oral) Invited Keynote
Day-3.2 Thursday May 22nd, 11am - 1pm : (48 posters) Theoretical and Linguistic Prosody
Day-3.3 Thursday May 22nd, 2pm - 3:30pm : Panel Session & memorial
Day-3.4 Thursday May 22nd, 4pm - 6pm : (6 oral) Perception and Production
Day-3.5 Thursday May 22nd, 7pm - 9:30pm : **BANQUET - Old Dining Hall** - *ALL WELCOME!*

Day-4.1 Friday May 23rd, 9am - 10:30am : (48 posters) Intonation and Speaking Style
Day-4.2 Friday May 23rd, 11am - 1pm : (6 oral) Intonation - general
Day-4.3 Friday May 23rd, 2pm - 3:30pm : plenary (1+3 oral) Invited Keynote
Day-4.4 Friday May 23rd, 3:30pm - 4pm : closing

Day-5 Saturday May 24th, **FREE DAY**

Day One

Tuesday May 20th, 1:30pm - 2pm : Opening (welcome to Speech Prosody!)

Tuesday May 20th, 2pm - 3:30pm : plenary (1+3 oral presentations) (p.1)

Invited Keynote: Fred Cummins - followed by 3 oral presentations
Chair: Ailbhe Ní Chasaide

Coffee break - 3:30 - 4pm

Tuesday May 20th, 4pm - 5:30pm : (Booster & Round Table) (p.2)

SSSpASSS: Special Session: **Social prosody: Affective Social Speech Signals**
Chair: Véronique Aubergé
Véronique and team to introduce SSSpASSS posters (booster: 2-min per paper) with Round Table discussion to follow, then posters in the exhibition space with welcome reception

Tuesday May 20th, 5:30pm - 6pm : SSSpASSS Posters) (p.2)

Chair: Gu Wentao

Tuesday May 20th, 6pm - 7:30pm : - Prominence & Phrasing - (28 posters) (p.7)

Chair: Juraj Šimko

Day Two

Wednesday May 21st, 9am - 10:30am : plenary (1+3 presentations) (p.15)

Invited Keynote: Stefanie Shattuck-Hufnagel - followed by 3 oral presentations
Chair: Julia Hirschberg

Coffee break - 10:30 - 11am

Wednesday May 21st, 11am - 1pm : oral (6 presentations) (p.16)

- Speech Rhythm and Timing - Chair: Amalia Arvaniti

Lunch break - 1 - 2pm (lunch provided)

Wednesday May 21st, 2pm - 4pm : special session - Slavic Prosody (p.18)

special session - Slavic Prosody (another SP!) Chairs: Zofia Malisz & Marzena Żygis

Coffee break - 4:00 - 4:30pm

Wednesday May 21st, 4:30pm - 6pm : poster (48 presentations) (p.19)

- Perception and Intonation - Chair: Oliver Jokisch

evening - Reviewers' Reception in Trinity College Long Room (by invitation)

Day Three

Thursday May 22nd, 9am - 10:30am : plenary (1+3 presentations) (p.31)

Invited Keynote: Jürgen Trouvain followed by 3 oral presentations
Chair: Nigel Ward

Coffee break - 10:30 - 11am

Thursday May 22nd, 11am - 1pm : poster (48 presentations) (p.32)

- Theoretical and Linguistic Prosody - Chair: Alice Turk

Lunch break - 1 - 2pm (lunch provided)

Thursday May 22nd, 2pm - 2:30pm : Remembering Sugito Miyoko (p.44)

Chair: Sadanobu Toshiyuki

Thursday May 22nd, 2:30pm - 3:30pm : Terminology in Prosody Research (p.44)

Chair: Fujisaki Hiroya

Coffee break - 3:30 - 4pm

Thursday May 22nd, 4pm - 6pm : oral (6 presentations)

- Perception and Production - Chair: Agnieszka Wagner

Thursday May 22nd, 7pm - 9:30pm : BANQUET - Old Dining Hall - ALL WELCOME!

Day Four

Friday May 23rd, 9am - 10:30am : poster (48 presentations) (p.46)

- intonation and speaking style - Chair: Laura Dilley

Coffee break - 10:30 - 11am

Friday May 23rd, 11am - 1pm : oral (6 presentations) (p.59)

- Intonation - Chair: Hansjörg Mixdorff

Lunch break - 1 - 2pm (lunch provided)

Friday May 23rd, 2pm - 3:30pm : plenary (1+3 presentations) (p.61)

Invited Keynote: Anne Cutler - followed by 3 oral presentations
Chair: Chiu Yu Tseng

Friday May 23rd, 3:30pm - 4pm : closing

Saturday May 24th, Free Day - go explore Dublin!!!

The Authors index

(sorted by given name - for an index sorted by family name see the Author index at the end of the proceedings (p.1180))

Page numbers refer to the abstract entry where a link can be found to the full paper.

Abstracts

NOTE: page numbers refer to the digital form of the proceedings where full papers are included - these can be downloaded from http://www.speechprosody2014.org/proceedings.pdf

1 Day One - May 20th

Tuesday - Opening Session

1:30pm - 2pm : 1-0-opening (3 bros:welcome!etc)

1.1 Tuesday Session One

2pm - 3:30pm : 1-1-plenary (1+3 presentations)

1.1.1 KeyNote 1

Fred Cummins - 30-min
From Prayer to Protest: An Initial Look at Joint Speech
Joint speech is an umbrella term covering choral speech, synchronous speech, chant, and all forms of speech where many people say the same thing at the same time. Prosodists, more than most, should be aware of the incompleteness of a structuralist description of language. Much of our use of language is ignored or missed when linguistic behaviour is viewed through the narrow lens of phonological/syntactic structure. I will discuss Joint Speech, as found in prayer, protest, classrooms, and sports stadia around the world. Despite its deep embedding in practices we value very much, joint speech has not hitherto attracted the attention of scientists as a distinct form of language behavior, because it is uninteresting from a structuralist point of view. If we merely take the time to look, however, there is much to be found in joint speech that is crying out for elaboration and investigation. I will attempt to sketch the terra incognita that opens up and present a few initial findings (phonetic, anthropological, neuroscientific) that suggest that Joint Speech is far from being a peripheral and exotic special case. It is, rather, a central example of language use that must inform our theories of what language and languaging are.

1.1.2 p.65

Zhihua Xia, Rivka Levitan, Julia Hirschberg,
Prosodic Entrainment in Mandarin Chinese and English: A Cross-Linguistic Comparison
Entrainment is the propensity of speakers to begin behaving like one another in conversation. We identify evidence of entrainment in a number of acoustic and prosodic dimensions in conversational speech of American English speakers and Mandarin Chinese speakers. We compare entrainment in the Columbia Games corpus and the Tongji Games Corpus and find some remarkable similarities between the two.

1.1.3 p.70

Rob Voigt, Robert J. Podesva, Dan Jurafsky,
Speaker Movement Correlates with Prosodic Indicators of Engagement
Recent research on multimodal prosody has begun to identify associations between discrete body movements and categorical acoustic prosodic events such as pitch accents and boundaries. We propose to generalize this work to understand more about continuous prosodic phenomena distributed over a phrase - like those indicative of speaker engagement - and how they covary with bodily movements. We introduce movement amplitude, a new vision-based metric for estimating continuous body movements over time from video by quantifying frame-to-frame visual changes. Application of this automatic metric to a collection of video monologues demonstrates that speakers move more during phrases in which their pitch and intensity are higher and more variable. These findings offer further evidence for the relationship between acoustic and visual prosody, and suggest a previously unreported quantitative connection between raw bodily movement and speaker engagement.

1.1.4 p.75

Štefan Beňuš, Marián Trnka,
Prosody, voice assimilation, and conversational fillers
Conversational fillers (CFs), commonly transcribed as uh, um, or er, typically start with a schwa-like vowel, and signal multiple social, interactive, meta-cognitive, and pragmatic functions. They also co-occur with prosodic boundaries, increase saliency of inter-word disjunctures, and participate thus in coding the prosodic structure. Contrary to these functions, CFs are assumed not to participate in the phonological system of a language. This paper uses two types of Slovak conversational speech corpora for investigating the the prosodic and phonological behavior of CFs. In Slovak, the vowel inventory does not include a schwa, and word-final obstruents undergo voice assimilation that is triggered by word-initial vowels but interacts with the strength of the prosodic boundary between the two words. Our data show the propensity of CFs to neutralize word-final voicing, and function thus as prosodic breaks, but also non-negligible number of cases of CFs triggering voicing of word-final obstruents, supporting their relevance for cognitive phonology.

1.2 Tuesday Special Session - SSSpASSS

4pm - 5:30pm : Booster & Round Table

Special Session: Social Prosody: **Affective Social Speech Signals**

Véronique and team to introduce SSSpASSS posters (booster: 2-min per paper) with Round Table then poster session to follow after break (posters with Welcome Reception)

5:30pm - 6pm : (20 poster presentations)

Tuesday Session Two - SSSpASSS Posters

1.2.1 p.81

Marta Maffia, Elisa Pellegrino, Massimo Pettorino,
Labeling expressive speech in L2 Italian: the role of prosody in auto-and external annotation
The present study is intended to compare two approaches of labeling expressive corpora: auto-annotation and annotation by external lay listeners. These two methods have been applied to the semi-spontaneous emotional speech produced by Chinese learners of L2 Italian, involved in the CardTask, a mood-induction procedure that permits to control the context of interaction, preserving the spontaneity of reactions. The emotional responses to the stimuli presented in the task were object of an auto-annotation session. The same samples were then administered only in the auditory mode to 20 Italian and 20 Chinese lay listeners. The results of perceptual tests have underlined some similarities and differences both between auto- and external annotation, and between the rates given by Italian and Chinese external listeners. The labels chosen by native Italians were similar to those selected in the auto-annotation session, particularly in the case of anxiety, fear and disgust. The correspondence between the results of the two annotation methods may be ascribed to the different prosodic patterns characterizing the emotional states. The results of the annotation made by Chinese listeners show that they found it hard to give a specific emotional label to utterances produced in a second language relying only on prosodic patterns.

1.2.2 p.86

Yuko Sasa, Véronique Aubergé,
Socio-affective interactions between a companion robot and elderly in a Smart Home context: prosody as the main vector of the "socio-affective glue"
The aim of this preliminary study of feasibility is to give a glance at interactions in a Smart Home prototype between the elderly and a companion robot that is having some socio-affective language primitives as the only vector of communication. Through a Wizard of Oz platform (EmOz), a robot is introduced as an intermediary between the technological environment and some elderly who have to give vocal commands to the robot to control the Smart Home. The robot vocal productions increases progressively by adding prosodic levels: (1) no speech, (2) pure prosodic mouth noises supposed to be the "glue's" tools, (3) lexicons with supposed "glue" prosody and (4) subject's commands imitations with supposed "glue" prosody. The elderly subjects' speech behaviors confirm the hypothesis that the socio-affective "glue"

effect increase towards the prosodic levels, especially for socio-isolated people.

1.2.3 p.91

Noor Alhusna Madzlan, Jingguang Han, Francesca Bonin, Nick Campbell,
Towards Automatic Recognition of Attitudes: Prosodic Analysis of Video Blogs
Understanding of speakers' attitude is essential for establishing successful human interaction. In this paper we analyse attitude manifestations in video blogs. We describe the main features of this novel communication medium and focus our attention on its possible exploitation as a rich source of information for human-human and human-machine communication. We describe the manual annotation of attitudes and the prosodic analyses. Finally we present a preliminary attitude automatic annotation system that attains 65% accuracy.

1.2.4 p.95

Pan Liu, Marc Pell,
Processing emotional prosody in Mandarin Chinese: A cross-language comparison
To understand how emotional prosody is processed in Mandarin Chinese and whether it differs from that of other languages, we conducted a perceptual-acoustic study on a set of Chinese vocal emotional stimuli and examined how they were perceived and acoustically characterized, in comparison with four other languages, English, Arabic, German, and Hindi, reported by Pell et al. [1]. Chinese pseudo-utterances spoken in seven emotions (anger, disgust, fear, sadness, happiness, pleasant surprise, and neutrality) were first identified by a group of native Mandarin speakers in a seven forced choice task, and then subjected to acoustic analyses. Results revealed that among the seven emotions, neutrality, anger, sadness, and fear tended to be recognized most accurately. Acoustic analysis demonstrated the importance of three acoustic parameters (f0 mean, f0 range, and speech rate) in characterizing vocal emotions in Mandarin. Both the perceptual and acoustic characteristics are highly similar, although not identical, to that observed by Pell et al. [1] in English, Arabic, German, and Hindi, indicating a set of universal principles in vocal emotion communication across languages.

1.2.5 p.100

Carlos Ishi, Hiroaki Hatano, Miyako Kiso,
Acoustic-prosodic and paralinguistic analyses of "uun" and "unun"
The speaking style of an interjection contains discriminative features on its expressed intention, attitude or emotion. In the present work, we analyzed acoustic-prosodic features and the paralinguistic functions of two variations of the interjection "un", a lengthened pattern "uun" and a repeated pattern "unun", which are often found in Japanese conversational speech. Analysis results indicate that there are differences in the paralinguistic function expressed by "uun" and "unun", as well as different trends on F0 contour types according to the conveyed paralinguistic information.

1.2.6 p.105

Jean Philippe Goldman, Tea Prŝir, George Christodoulides, Antoine Auchlin,
Speaking style prosodic variation: an 8-hour 9-style corpus study
This paper presents the results of a prosodic and phonostylistic analysis based on C-PhonoGenre, a 9-hour-long spoken French corpus, including 10 speakers on average recorded in 10 speaking situations. The corpus was automatically segmented at phonetic, syllabic, word levels (EasyAlign), and larger pause-separated units. Part-of-speech annotation (DisMo) and prominent syllable detection (ProsoProm) was added automatically. The corpus was also manually annotated at the syllabic level for stylistic variants, such as post-tonic schwas, liaisons, elisions, disfluencies, audible breaths and noises. Acoustic analyses (ProsoReport, DurationAnalyser) provide more than 100 micro- and macro-prosodic measures, which we correlate with the phonostylistic and linguistic annotation. This analysis finally yields a contrastive, fine-grained prosometric description of phonostylistic and situational variation, over 4 situational gradual dimensions: audience, media, preparation, and interactivity. Further statistical analysis was carried out to explore the discriminative and explanatory power of combinations of prosodic measures.

1.2.7 p.110

Leandra Antunes, Véronique Aubergé, Yuko Sasa,
Certainty and uncertainty in Brazilian Portuguese: methodology of spontaneous corpus collection and data analysis
This work presents a methodology used to collect some spontaneous social affect corpus and preliminary prosodic analysis of certainty and uncertainty in Brazilian Portuguese. The corpus was collected by a Wizard of Oz (Emoz)method, the scenario to induce certainty and uncertainty is based on the situation of a job interview, for which a companion robot (Emox) is supposed to be a trainer. The subjects were convinced to benefit of a free training of this "revolutionary" method to train to job interview. In this scenario the linguistic expressions are partially controlled, in order to focus the certainty/uncertainty expression mainly on paraphrasing and prosody. Data were preliminary analyzed for audiovisual prosody: videos analysis were made regarding eyes, mouth and face/head movements, while audio analysis were made about acoustic prosody parameters of fundamental frequency and duration. The first results show that using Emoz within such a scenario is an efficient way to induct spontaneous but comparable speech production. Prosodic results show that fundamental frequency and duration measurements, as well as eyes, mouth and face/head movements, are differently used in certainty and in uncertainty production in Brazilian Portuguese.

1.2.8 p.115

Jan Volín, Lenka Weingartová, Oliver Niebuhr,
Between Recognition and Resignation The Prosodic Forms and Communicative Functions of the Czech Confirmation Tag "jasně"
Like question tags, confirmation tags such as the Czech affirmative particle jasně can be used with various prosodic characteristics that augment, reverse or otherwise modify their relatively unspecific lexical meaning. We extracted 172 instances of jasně from several dialogues and assessed their discourse function. 36 prosodic correlates in temporal, amplitude and fundamental frequency domains were measured and used in three computational classifiers: linear discriminant analysis, classification trees and artificial neural networks. All three methods significantly reflected the functional assessments and additionally indicated the relative importance of individual predictors in a mutually consistent manner.

1.2.9 p.120

Ting Wang, Hongwei Ding, Qiuwu Ma, Daniel Hirst,
Automatic Analysis of Emotional Prosody in Mandarin Chinese:Applying the Momel Algorithm
Based on the Momel algorithm, a set of acoustic parameters was analyzed automatically on Chinese emotional speech. Global prosodic features were calculated on the sentence level, which showed a concordance with the usual pattern reported in the literature. Local constraints were also considered on the syllable layer. An ANOVA showed that there were interactive effects among emotions, syllable positions and syllable tones on certain parameters. Further more, by examining the pitch movements, no significant difference was found between neutral speech and active emotional speech, which was different from the performance in non-tonal languages. However when reducing the tonal influence by using only tone 1 syllables in the utterance, this inverse effect disappeared. Hence we posited an interpretation that due to the existence of lexical tone in Mandarin Chinese, the paralinguistic use of pitch movements has been reduced.

1.2.10 p.125

Yan Lu, Véronique Aubergé, Albert Rilliard,
Prosodic Profiles of Social Affects in Mandarin Chinese
An acted corpus of 19 prosodic social affects is devoted to this work, which investigates the production side of prosodic attitudes in Mandarin Chinese, with the aim of extracting the more prominent patterns of acoustical variations. Results are then compared to previous perception data obtained on the same expressions. The F0, intensity and duration characteristics of 76 utterances conveying 19 prosodic attitudes are statistically examined in this study. All attitudes are regrouped into 5 clusters according to their prosodic features. The result of the statistical analysis shows that the prominent differentiation between clusters is mostly related to F0 and duration parameters; some similarities are noted between the clustering of attitudes from acoustic features and from perceptual confusions obtained in previous

experiments; inside each cluster, some attitudes show typical characteristics in F0 and duration.

1.2.11 p.130

Houwei Cao, Štefan Beňuš, Ruben C. Gur, Ragini Verma, Ani Nenkova,
Prosodic cues for emotion: analysis with discrete characterization of intonation
In this paper we study the relationship between acted perceptually unambiguous emotion and prosody. Unlike most contemporary approaches which base the analysis of emotion in voice solely on continuous features extracted automatically from the acoustic signal, we analyze the predictive power of discrete characterizations of intonations in the ToBI framework. The goal of our work is to test if particular discrete prosodic events provide significant discriminative power for emotion recognition. Our experiments provide strong evidence that patterns in breaks, boundary tones and type of pitch accent are highly informative of the emotional content of speech. We also present results from automatic prediction of emotion based on ToBI-derived features and compare their prediction power with state-of-the-art bag-of-frame acoustic features. Our results indicate their similar performance in the sentence-dependent emotion prediction tasks, while acoustic features are more robust for the sentence-independent tasks. Finally, we combine ToBI features and acoustic features together and further achieve modest improvements in sentence-independent emotion prediction, particularly in differentiating fear and neutral from other emotion.

1.2.12 p.135

Massimo Pettorino, Elisa Pellegrino, Marta Maffia,
"Young" and "Old" Voice: the prosodic auto-transplantation technique for speaker's age recognition
The present study is intended to figure out the extent to which prosody and intonation entail listeners' ability to estimate the speaker's age. The performance of a 40-year old anchorman and that produced by the same speaker at the age of 80 were spectro-acoustically analyzed in order to identify the prosodic features of the "young" and the "old" voice. The results of the analyses have shown relevant differences between the two voices on suprasegmental level. To test the effects of these differences on perceptual level, through the prosodic transplantation technique, the F0 values and the durations of segments and silences were transferred from the "young" to the "old" voice and viceversa. Two age recognition tests, based on original and transplanted voices, were administered to Italian listeners. The results of perceptual tests have confirmed the strict relationship between some rhythmic and prosodic features and the speaker's age and have demonstrated the effectiveness of the transplantation technique. With advancing age, articulation rate and speech rate slow down, voice register raises and tonal range widens. Moreover, the "old" voice is also characterized by a higher percentage of vocalic portion that determines a shift of Italian rhythm towards the isomoraic pattern

1.2.13 p.140

Julien Magnier, Maya Gratier, Anne Lacheret,
Expressive prosody vs neutral prosody : From descriptive binary to continuous features
In this paper, we propose to compare expressive and neutral oral renditions of a children's tale in french by examining the segmentations performed by twelve high level french readers. We used a software dedicated to this kind of analysis (Analor) which takes into account different parameters (pause, pitch gesture, pitch jump) and their relative strength to determine pertinent prosodic units (phrases). The extraction of these phrases and their features enables us to observe the influence of both the type of oralisation (expressive or neutral) and punctuation signs on the organization of speech flow. Results show that prosodic phrases in the expressive readings are more numerous (specially at comma locations), that their boundaries are more clearly demarcated, and that they have more varied contours than those in neutral readings.

1.2.14 p.144

Heather Pon-Barry, Arun Reddy Nelakurthi,
Challenges for Robust Prosody-based Affect Recognition
Prosody-based affect recognition has great potential impact for building adaptive speech interfaces. For example, in intelligent systems for personalized learning, sensing a student's level of certainty, which is

often signaled prosodically, is one of the most interesting states to interpret and respond to. However, robust uncertainty recognition faces several challenges, including the lack of gold-standard labels, and differences in expressivity among speakers. In this paper we explore the intersection of these two issues. We have collected a corpus of spontaneous speech in a question-answering task. Three kinds of certainty labels are associated with each utterance. First, speakers rated their own level of certainty. Second, a panel of listeners rated how certain the speaker sounded. Third, an externally crowdsourced difficulty score is generated for each stimulus (the question). We present an analysis of the prosodic characteristics of individual speaking styles, as they relate to these three different measurements of certainty.

1.2.15 p.149

Dominique Fourer, Takaaki Shochi, Jean Luc Rouas, Jean Julien Aucouturier, Marine Guerry,
Prosodic analysis of spoken Japanese attitudes
The aim of this paper is to provide cues for prosodic characterization of attitudes in Japanese. This is comparable with similar researches made on others languages (e.g. American english, Brazilian portuguese, etc.). The presented work focuses on an objective analysis results of the Japanese based on the audio signal structure. In the proposed experiments, the speech signal of several Japanese native speakers is analyzed, The used signals were recorded in a particular context where the corresponding attitude is clearly identified and segmented. The presented results are based on a previous study where 16 attitudes were defined to describe the emotional content of human spoken language. Thus we compare the signal properties which can characterize each attitude.

1.2.16 p.154

Noam Amir, Eitan Globerson,
On the Role of Pitch in Perception of Emotional Speech
Two experiments investigated the role of intonation in perception of basic emotions. In the first experiment, pitch contours of stimuli from a corpus containing portrayals of anger, joy, fear and sadness were manipulated with respect to range, mean and smoothness. In the second experiment, pitch contours of identical words portraying different emotions were exchanged. In each experiment, the emotional category and intensity of the original and manipulated stimuli were evaluated by two separate groups of 20 participants. Results of the first experiment show mainly that pitch mean and range should vary congruently to portray activation correctly, and demonstrate the interaction in varying these two parameters. Results of the second experiment show that a pitch contour conveying high activation is not sufficient in conveying the appropriate emotion, if the other paralinguistic cues are not also in accordance. A pitch contour indicating low activation, on the other hand, is apparently a more powerful cue and thus less reliant on other cues.

1.2.17 p.159

Sven Grawunder, Marianne Oertel, Cordula Schwarze,
Politeness, culture, and speaking task - paralinguistic prosodic behavior of speakers from Austria and Germany
This paper tests previous findings for polite speech of low pitch, low intensity, higher number of hesitation markers and filled pauses against those parameters in a different socio-cultural background. Two similar groups of (19+13) participants, from Austria and from Germany, were recorded. The adopted experimental approach used 16 tasks aiming at different speech acts in situations that evoke either polite or informal speech. The analyzed acoustic and electroglottographic signals reveal main effects for lower pitch, lower intensity and HNR only for the German group. Open quotient values differ only for female speakers. In both groups significantly lower word rate and lower speaking rate as well as higher rates of filled pauses and hesitation markers are found in formal (polite) conditions. However individual speakers can show indifferent or opposing behavior for a given parameter with compensatory utilisation of other parameters in order to express politeness (formality).

1.2.18 p.164

Miguel Oliveira Jr, Ayane Nazarela Santos De Almeida, René Alain Santana De Almeida, Ebson Wilkerson Silva,

Speech rate in the expression of anger: a study with spontaneous speech material
The study of the acoustic expression of emotion is, in general, the analysis of whether prosodic variables such as intonation (F0), speech rate, pauses, rhythm, intensity and duration, are reliable clues for the characterization of the emotional states of the speaker. The present paper aims to verify whether an association exists in Brazilian Portuguese between the basic emotion of "anger" and the prosodic variable "speech rate", as the literature often suggests there is for other languages. The corpus consisted of fragments of spontaneous speech recorded from a radio program. The fragments were selected on the basis of a perceptual test. For the production analysis, only excerpts that were identified by more than 75% of the participants of the perceptual test as associated to the categories "anger" and "neutral" were selected. The results demonstrated that, for the data that were used for the analysis, there is a general reduction in speech rate when utterances are associated with the emotion of "anger", if compared to utterances spoken in a "neutral" mode by the same speaker, contrary to what literature often indicates for other languages.

1.2.19 p.169

Yan Lu, Véronique Aubergé, Nicolas Audibert, Albert Rilliard,
Audiovisual perception of expressions of Mandarin Chinese social affects by French L2 learners
This study focuses on confusions made by French L2 learners vs. native subjects in the perception of 11 audiovisual Mandarin Chinese attitudes, selected from a broader set of 19 attitudes previously evaluated in audio condition by both native Chinese and naïve French listeners. Two groups of French L2 learners of Mandarin Chinese were selected according to their level assessed by the Common European Framework of Reference for Languages: 9 beginners (A1) vs. 10 intermediate learners (A2). Subjects evaluated the 11 attitudes in audio, visual and audiovisual condition. Comparison of confusions between learners of level A1 vs. A2 indicates few significant differences, mostly in audiovisual condition and without a clear gain for one group over the other: confusions patterns are closer to the native reference for group A1 in expression of doubt, and for group A2 in expression of contempt. The comparison of French L2 learners pooled together vs. native speakers reference sheds light on major confusions to be targeted by specific methods and exercises. In audio-only condition, neutral surprise and politeness are less recognized by learners, who confuse contempt with question and question with obviousness. In visual-only condition, obviousness is more confused with declaration, contempt with irritation, and disappointment with doubt. In audio-visual condition, recognition of neutral surprise is lower, while infant-directed speech is better recognized; neutral surprise is more confused with irritation and contempt with disappointment. Cross-modality comparisons suggest a limited contribution of informations conveyed by acoustic prosody in the identification of audiovisual social affects by L2 learners.

1.2.20 p.174

Nuzha Moritz, Christophe Damour,
Coordination between gesture and prosody in two versions of"The Great Gatsby": 1974, 2013"
The cross-disciplinary study (phonetics and film study) aims at highlighting the coordination between posture and prosody in two versions of "The Great Gatsby". The central aim of the study is to understand how prosodic variations are related to gesture in different acting schools. Formal and functional analysis of gesture and their relation to prosody, shows striking contrast between the acting styles

1.3 Tuesday Session Three - Poster

- Prominence & Phrasing -

1.3.1 p.183

Fred Cummins, Judit Varga,
Explorations in the prosodic characteristics of synchronous speech, with specific reference to the roles of words and stresses
We examine the prosodic characteristics of read speech produced alone or in synchrony with a co-speaker in English. Previous work has demonstrated a marked difference between these two speaking conditions in Mandarin, but not English. We employ word lists that are either simple sequences of trochees, or

complex lists with regular stress alterantion but irregular word boundaries. Inter-onset intervals are examined and no major differences between solo and synchronous interval sequences are found. Viewed from the perspective of two generative models, however, there is weak evidence for some small difference in the dependence of interval duration on serial position.

1.3.2 p.187

Zhen Qin, Annie Tremblay,
Effects of native dialect on Mandarin listeners' use of prosodic cues to English stress
This study investigates the effect of native dialect on the use of prosodic cues to English stress by Standard Mandarin (SM) listeners, Taiwanese Mandarin (TM) listeners, and English listeners. Both SM and TM use fundamental frequency (F0) to realize lexical tones, but only SM uses duration together with F0 to realize lexically contrastive full-full vs. full-reduced stress patterns. Native English listeners and second language learners of English who spoke SM or TM as native language and were at similar proficiencies in English completed a sequence-recall task. English disyllabic non-words that differed in stress placement were resynthesized to contain only F0 cues, only duration cues, or converging F0 and duration cues. The results showed that SM-speaking learners used duration more than TM-speaking learners to recall English non-words. Native dialect is suggested to be considered in second language speech processing models.

1.3.3 p.192

Caterina Petrone, Mariapaola D'Imperio, Susanne Fuchs, Leonardo Lancia,
The interplay between prosodic phrasing and accentual prominence on articulatory lengthening in Italian
The distribution of preboundary lengthening within the phrase-final word is controversial. In CV syllables immediately preceding a prosodic boundary, the acoustic duration of the syllable onset C is less involved than that of the following rime V in the lengthening phenomenon. Moreover, preboundary lengthening might be extended to the stressed/accented rime within the phrase final word. On the other hand, articulatory constriction gesture for the onset consonant can be lengthened despite not being immediately adjacent to a boundary. In this study, we explore the effects of prosodic boundary and prominence in Italian, at both acoustic and articulatory level. Bilabial consonants in CV onset position were examined. The consonants were inserted in unstressed (word final) and stressed (penultimate vs. antepenultimate) syllables occurring in the vicinity of prosodic boundaries of different levels. In final syllables, the acoustic duration of the onset consonant was not affected by the prosodic boundary manipulation whereas the closing gesture duration showed a pattern of lengthening which was stronger for higher level prosodic boundaries. In non-final syllables, no acoustic/articulatory effect was found for onset consonants but only on the stressed vowels in penultimate position. Structural, phonological and phonetic constraints might be at work in determining preboundary lengthening.

1.3.4 p.197

Francisco Torreira, Miquel Simonet, José Ignacio Hualde,
Quasi-neutralization of stress contrasts in Spanish
We investigate the realization and discrimination of lexical stress contrasts in pitch-unaccented words in phrase-medial position in Spanish, a context in which intonational pitch accents are frequently absent. Results from production and perception experiments show that in this context durational and intensity cues to stress are produced by speakers and used by listeners above chance level. However, due to substantial amounts of phonetic overlap between stress categories in production, and of numerous errors in the identification of stress categories in perception, we suggest that, in the absence of intonational cues, Spanish speakers engaged in online language use must rely on contextual information in order to distinguish stress contrasts.

1.3.5 p.202

Miquel Simonet, Joseph Casillas, Yamile Díaz,
The effects of stress/accent on VOT depend on language (English, Spanish), consonant (/d/, /t/) and linguistic experience (monolinguals, bilinguals)
This study examines Voice Onset Times of coronal stops in utterance-initial position in two languages.

Crucially, the effects of lexical stress (stressed, unstressed syllable) on VOT are analyzed. The study investigates aspirated stops (English /t/), short-lag voiceless stops (English /d/, Spanish /t/) and prevoiced stops (Spanish /d/). Three groups of speakers provide data: English monolinguals, Spanish monolinguals, and proficient Spanish-English bilinguals. The study finds that lexical stress lengthens aspiration (English /t/) and prevoicing (Spanish /d/) but it does not alter significantly short-lag stops (Spanish /t/, English /d/). Monolinguals and bilinguals differ slightly in their phonetic behavior. Implications for gestural coordination as well as for feature theory are discussed.

1.3.6 p.??

Bogdan Ludusan, Guillaume Gravier, Emmanuel Dupoux,
Incorporating Prosodic Boundaries in Unsupervised Term Discovery
We present a preliminary investigation on the usefulness of prosodic boundaries for unsupervised term discovery (UTD). Studies in language acquisition show that infants use prosodic boundaries to segment continuous speech into word-like units. We evaluate whether such a strategy could also help UTD algorithms. Running a previously published UTD algorithm (MODIS) on a corpus of prosodically annotated English broadcast news revealed that many discovered terms straddle prosodic boundaries. We then implemented two variants of this algorithm: one that discards straddling items and one that truncates them to the nearest boundary (either prosodic or pause marker). Both algorithms showed a better term matching F-score compared to the baseline and higher level prosodic boundaries were found to be better than lower level boundaries or pause markers. In addition, we observed that the truncation algorithm, but not the discard algorithm, increased word boundary F-score over the baseline.

1.3.7 p.212

Chiu Yu Tseng, Chao Yu Su,
Binary Contrast and Categorical DifferentiationProsodic Characteristics of English Word Stress in Broad and Narrow Focus Positions
Assuming that categorical differentiation is major acoustic characteristics of English lexical stress through binary instead of more complex 3-way distinction, we investigated lexical stress in broad and narrow focus positions and found how binary distinction is achieved by the concomitancy of secondary stress defined by its position and distance in relation to primary stress. Similar results are found in broad (sentence initial) and narrow focus as well. These results suggest that binary categorical contrast is the optimal choice while differentiation is dependent on robust contrast patterns in the speech signal.

1.3.8 p.217

Adrian Leemann, Marie José Kolly, Volker Dellwo,
Crowdsourcing regional variation in speaking rate through the iOS app 'Dialäkt Äpp'
It is a common stereotype in Switzerland that speakers from Bern speak slowly and speakers from Zurich speak quickly. Are these differences in perception at all mirrored in production? We present a new method of crowdsourcing speaking rate through a free of charge iOS application. Astonishingly, results indicate that the temporal structure of a few words alone as spoken by a few hundred speakers are sufficient to tell apart the two dialects in speaking rate. In line with previous literature, females articulate more slowly than males. Further potential fields of application of the introduced method are discussed.

1.3.9 p.222

Núria Esteve-Gibert, Ferran Pons, Laura Bosch, Pilar Prieto,
Are gesture and prosodic prominences always coordinated? Evidence from perception and production
This study explores the temporal coordination between gesture and speech by addressing two main questions: (1) Are speakers sensitive to the misalignment between gesture prominence and prosodic prominence? (2) Is this sensitivity modulated by the semantic information conveyed by gesture and speech modalities in production? Experiment 1 tested question (1) and Experiment 2 tested question (2). Results from Experiment 1 revealed that the combinations in which prominences were misaligned were less acceptable than combinations with aligned prominences, and that the metrical pattern of the target word had an effect on the speakers' sensitivity: unsynchronized trochees (with the gesture prominence at the

post-tonic syllable) were frequently accepted, while unsynchronized iambs (with the gesture prominence at the pre-tonic syllable) were rejected. Results from Experiment 2 revealed that when the pointing gesture adds information to speech, i.e. it is supplementary to speech, the prominences are frequently misaligned (with gesture occurring after the speech), as if two different speech acts were produced. These findings suggest that the semantic content of gesture-speech combinations might influence the speakers' sensitivity of the misalignment between prosodic and gesture prominences.

1.3.10 p.227

Stefan Baumann, Anna Roth,

Prominence and Coreference On the Perceptual Relevance of F0 Movement, Duration and Intensity

We conducted a web-based experiment on German testing the perception of an element's prosodic prominence in relation to its status as a potential coreferent of an antecedent. Data were elicited by asking subjects to judge the probability of a coreference relation between a context noun (antecedent) and a target word (anaphor), whose lexically stressed syllable was manipulated as to the parameters F0 movement, duration and intensity. Results suggest a direct but inverse relationship between prominence and coreference judgements indicating that the likelihood of a coreference interpretation decreases with increasing prosodic prominence. F0 movement turned out to be the dominant cue for prominence as the main trigger for the perception of pitch accents with rises being perceived as more prominent than falls. In turn, lack of tonal movement probably led to perceived deaccentuation and thus favoured the evaluation of a target word as being coreferential with an antecedent. Duration was found to be a significant factor as well, while intensity did not prove to be relevant for the task given. Thus, the present study with its revised methodology adds new aspects to the debate of which parameters are crucial for prominence perception, directly linking it to the investigation of information structure.

1.3.11 p.232

Pärtel Lippus, Eva Liina Asu, Mari-Liis Kalvik Mari,

An acoustic study of Estonian word stress

This study investigates the acoustic correlates of word stress in Estonian. It forms part of a broader international collaboration the aim of which is to develop a universal language independent model for evaluating lexical stress regardless of the phonological structure of a given language. To this aim the characteristics of word stress in a range of languages is studied using unified methodology. For the present study, four acoustic measures were analysed as a function of speaking style and stress: vowel duration, F0 mean, F0 standard deviation, and spectral emphasis. The results show that the strongest correlate of style and stress in Estonian is vowel duration, but stress has a strong interaction with the Estonian three-way quantity system.

1.3.12 p.236

Lenka Weingartová, Kristýna Poesová, Jan Volín,

Prominence Contrasts in Czech English as a Predictor of Learner's Proficiency

The study investigates prominence patterns in Czech-accented English comparing the production of non-native speakers of English at two distinct stages of phonological acquisition (beginners and intermediates) with a native performance. Word stress in Czech is entirely different from English, it has a fixed position, a delimitative function and rather impalpable acoustic manifestations. Alternations in the realization of word stress were analyzed by measuring the ratios or differences of acoustic correlates of prominence: duration, fundamental frequency, sound pressure level and spectral slope. Since word stress is a relational phenomenon, these characteristics were measured in two adjacent syllables one of which was a canonical stress bearer. The results reveal a clear difference between native and non-native treatment of word stress in all parameters examined. In the non-native sample distinct interferences of L1 across the two groups were detected: the subjects displayed different exploitation of duration, spectral slope and SPL with relation to their proficiency in L2 English. Out of these, duration ratio proves to be the most significant correlate. Furthermore, our findings indicate a strong effect of prosodic context coinciding with the prominence features, particularly in intonation declination and phrase-final lengthening.

1.3.13 p.241

Alice Turk, Stefanie Shattuck Hufnagel,
A sketch of an extrinsic timing model of speech production
In this paper, we motivate and present a sketch of an extrinsic timing model of speech production. It is a three-stage model, involving 1) a phonological planning stage, where symbolic segmental representations are sequenced and slotted into an appropriate prosodic structure, and where appropriate acoustic cues are selected for each segment in its context, and 2) a phonetic planning stage, where cues are mapped onto sets of articulators and appropriate values for spatial and temporal parameters of movement are computed, and 3) a motor-sensory implementation stage, where articulator movements are generated and tracked. We cite model components from the literature that accomplish many of the functions this type of model requires.

1.3.14 p.246

Nicolas Obin, Julie Beliao, Christophe Veaux, Anne Lacheret,
SLAM: Automatic Stylization and Labelling of Speech Melody
This paper presents SLAM : a simple method for the automatic Stylization and LAbelling of speech Melody. This main contributions over existing methods are : the alphabet of melodic contours is fully data-driven, an explicit time-frequency representation is used to derive complex melodic contours, and melodic contours can be determined over arbitrary prosodic/syntactic units. Additionally, the system can handle some specificities of spontaneous speech (e.g., multi speakers, speech turns and speech overlaps). A preliminary experiment conducted on 3 hours of spoken French indicates that a small number of contours is sufficient to explain most of the observed contours. The method can be easily adapted to other stressed languages. The implementation is open-source and freely available.

1.3.15 p.251

Antonio Simoes,
Lexical Stress in Brazilian Portuguese in Contrast with Spanish
This study discusses stress assignment in prosodic, non-verbal words in Brazilian Portuguese, in comparison with descriptions of stress assignment for Spanish [9, 13, 15, 16, 17, 18]. Given the conflicting claims regarding stress assignment in Brazilian Portuguese (see [11, 1, 2, 10, 3]) there is still a need to revisit discussions on stress assignment in Portuguese. In general, stress assignment in Spanish has been satisfactorily explained through the interplay between the morphological and phonological domains. Similar descriptions for Portuguese still requires far more abstraction and use of artifacts than in Spanish, which makes Mattoso Cmara Jr.'s [4, 5] claim that lexical stress is unpredictable in Brazilian Portuguese surprisingly unchallenged.

1.3.16 p.256

Rena Nemoto,
Prosodic Characteristics of Vocalic Hesitations in Comparison with Overlong Vowels in Estonian
The goal of this paper is to investigate vocalic hesitations in Estonian and compare them to the related vowels of overlong (Q3) quantity degree. We wonder if there are some languagespecific characteristics of hesitations. If yes, which kind of characteristics can be observed in Estonian language? We analyze duration, fundamental frequency (f0), intensity, and first two formants using 39.5 hours of manually transcribed monoor dialogue speech from a spontaneous speech corpus. Investigated vocalic hesitations and Q3 vowels are: /ee, ää, aa, , öö/. The characteristics of hesitations as compared to those of Q3 vowels show that hesitations have longer duration range. Hesitations generally include lower f0 and intensity values. However, the values vary in terms of vowels. First two formants of hesitations tend to be located at more centralized positions in a vocalic triangle than related Q3 vowels.

1.3.17 p.261

Alexsandro Meireles, Plínio Barbosa,
Articulatory Reorganizations of Speech Rhythm due to Speech Rate Increase in Brazilian Portuguese
This paper examines how speech rate increase acts to change speech rhythm at the articulatory level.

Main results show that speech rate increase worked to change articulatory parameters in the following way: a) decrease of acceleration duration; b) decrease of y-extremum; c) decrease of constriction displacement; d) decrease in modulus of peak and/or valley velocity; e) decrease of gestural duration; and f) constant proportional time-to-peak (or valley) velocity. Besides, results have shown that speech rate tends to affect all gestures in an utterance independently of their phrasal position. Nevertheless, there was evidence that some articulatory parameters could, if properly manipulated, provide cues for rhythmic restructurings in speech. Finally, results show that the dynamical speech rhythm model (Barbosa, 2007) is more appropriate to deal with Brazilian Portuguese acoustical data than the pi-gesture model (Byrd &Saltzman, 2003), and that both models could explain articulatory reorganizations due to speech rate increase.

1.3.18 p.265

Sabine Zerbian, Jane Kühn, Christoph Schroeder, Svenja Schuermann,
Prosody in Turkish learners of German as a Foreign Language
Results of a pilot study are reported which investigates the prosodic realization of information structure by six learners of German as a Foreign Language (GFL) with Turkish as first language. Question-answer pairs were read out loud which systematically varied the position of narrow focus in the response by means of a preceding wh-question. A qualitative analysis of the results shows deaccentuation of postfocal constituents in the case of subject focus for 4/6 GFL speakers but no consistent pitch increase on focused constituents. Two speakers did not change prosody due to information structure. The results are discussed in connection with the acquisition of prosody as a marker of information structure. Deaccentuation has been reported to cause problems in L2 prosody. In Turkish, deaccentuation occurs postfocally. The claim will be motivated that the occurrence of deaccentuation in the L1 is a necessary but not sufficient condition for early acquisition of deaccentuation in a foreign language.

1.3.19 p.270

Sandrine Brognaux, Thomas Drugman, Marco Saerens,
Synthesizing sports commentaries: One or several emphatic stresses?
Emphatic stresses are known to fulfill essential functions in expressive speech. Their integration in speech synthesis usually relies on a prosodic annotation of the training corpus. Emphasized syllables are then assigned a single label or can receive several labels according to their acoustic realization. While it is more complex to predict those various labels for a new text to synthesize, it might allow for a better rendering of the stress in the synthesized speech. This paper examines whether the use of more than one emphatic label improves the perceived expressivity of the synthesized speech. It relies on a manually-annotated expressive corpus of sports commentaries. Statistical acoustic analyses show that four distinct realizations of emphatic stresses can be distinguished. However, perceptual tests indicate that the integration of this distinction in HMM-based speech synthesis does not lead to a significant improvement in expressivity. This seems to imply that the different acoustic realizations of the stress are not required to be explicitly annotated in the training corpus.

1.3.20 p.275

Jürgen Trouvain, Bernd Möbius,
Sources of variation of articulation rate in native and non-native speech: comparisons of French and German
Speech tempo including articulation rate is often considered as a good predictor in the diagnosis of foreign language proficiency and its comprehension. In this study we investigate various sources of variation of articulation rate such as the L2 proficiency level, individual tempo habits in L1 and L2, and more extensive exposure to native speech. In addition, we also discuss the difficulty of the most informative unit for rate metrics which allows comparisons between French and German. The materials used are French and German read sentences, produced as L1 and L2 speech. In contrast to other studies individual habits of articulation rate in the L1 was only partially observed in the corresponding L2 data (a slow L1 speaker does not necessarily articulate slowly in the L2). The convergence of most French learners to the German model speakers shows the advantage of having additional input for phonetic exercises. The fastest German learners also converge to the rather slow French model speaker.

1.3.21 p.280

Helena Moniz, Ana Isabel Mata, Julia Hirschberg, Fernando Batista, Andrew Rosenberg, Isabel Trancoso,

Extending AuToBI to prominence detection in European Portuguese

This paper describes our exploratory work in applying the Automatic ToBI annotation system (AuToBI), originally developed for Standard American English, to European Portuguese. This work is motivated by the current availability of large amounts of (highly spontaneous) transcribed data and the need to further enrich those transcripts with prosodic information. Manual prosodic annotation, however, is almost impractical for extensive data sets. For that reason, automatic systems such as AuToBi stand as an alternate solution. We have started by applying the AuToBI prosodic event detection system using the existing English models to the prediction of prominent prosodic events (accents) in European Portuguese. This approach achieved an overall accuracy of 74% for prominence detection, similar to state-of-the-art results for other languages. Later, we have trained new models using prepared and spontaneous Portuguese data, achieving a considerable improvement of about 6% accuracy (absolute) over the existing English models. The achieved results are quite encouraging and provide a starting point for automatically predicting prominent events in European Portuguese.

1.3.22 p.285

Fabio Tamburini, Chiara Bertini, Pier Marco Bertinetto,

Prosodic prominence detection in Italian continuous speech using probabilistic graphical models

Prosodic prominence, a speech phenomenon by which some linguistic units are perceived as standing out from their environment, plays a very important role in human communication. In this paper we present a study on automatic prominence identification using Probabilistic Graphical Models, a family of Machine Learning Systems able to properly handle sequences of events. We tested the most promising members of such models on utterances selected from a manually annotated Italian speech corpus, obtaining very good recognition results crucially converging with the prominence detection responses provided by a pool of native speakers.

1.3.23 p.290

Robert Fuchs,

Integrating variability in loudness and duration in a multidimensional model of speech rhythm: Evidence from Indian English and British English

Most research on speech rhythm has focussed on duration. For example, [1] suggested the normalised Pairwise Variability Index for vocalic intervals (nPVI-V) in order to measure the variability of vocalic durations. This paper argues that speech rhythm research should also take into account other correlates of prominence as well as their interaction. The duration-based nPVI, or nPVI-V(dur), is supplemented by an nPVI-V(avgLoud) that measures variability in average loudness. These two metrics account for variability in duration and loudness, but cannot measure if loudness and duration reinforce each other by varying simultaneously in the same direction. This simultaneous variability is accounted for by the combined nPVI-V(dur+avgLoud), which is higher than the average of the other two measures, if vocalic intervals that are longer than average are also louder than average. The three metrics are subsequently applied to recordings of a reading task performed by 20 speakers of Indian English (IndE) and 10 speakers of British English (BrE). Results indicate that IndE has less variability in duration and less variability in loudness than BrE. In addition, IndE has less simultaneous variability in duration and loudness than BrE. This indicates that duration and loudness are less often used together as cues to prominence in IndE than in BrE.

1.3.24 p.295

Hongwei Ding, Rüdiger Hoffmann,

A Durational Study of German Speech Rhythm by Chinese Learners

This study focuses on the temporal and metrical features of the German speech produced by Chinese speakers. German is described to be a stress-timed language, while standard Chinese is regarded as a syllable-timed language. It has been suggested that the rhythm of the target language can be influenced by the learners' native language. In this study we conducted an investigation of ten sentences with 18

Chinese students in the low intermediate proficiency level in comparison with six native German speakers. We compared the duration values in terms of pairwise variability indices, and found that most of these Chinese speakers have a lower nPVI-V and a higher rPVI-C than the German speakers. We illustrate that the conventional duration measures of nPVI-V can be influenced by the syllable structures of the utterance and the classification approach of vocalic intervals, and a comparable nPVI-V can hardly be expected from different investigations. Furthermore, we argue that duration values alone cannot fully capture the rhythmic patterns of speech because other prosodic parameters such as pitch and energy also join to contribute to rhythmic characteristics of the speech.

1.3.25 p.300

Donna Erickson, Shigeto Kawahara, J.C. Williams, Jeff Moore, Atsuo Suemitsu, Yoshiho Shibuya,
Metrical Structure and Jaw Displacement: An Exploration
The current experiment using EMA shows that the amount of jaw displacementor mandible movement-may reflect the metrical organization of English sentences. The experiment also supports F1 as a reliable acoustic correlate of jaw displacement, hence metrical organization, but also demonstrates that F0 does not have a similar relationship to mandible movement.

1.3.26 p.305

Erwan Pépiot,
Male and female speech: a study of mean f0, f0 range, phonation type and speech rate in Parisian French and American English speakers
Many studies have been conducted on acoustic differences between female and male speech. However, they have generally been led on speakers of only one language, and have focused on a single acoustic parameter. The present study is an acoustic analysis of dissyllabic words or pseudo-words produced by 10 Northeastern American English speakers (5 females, 5 males) and 10 Parisian French speakers (5 females, 5 males). Several prosodic parameters were measured: mean f0, f0 range, phonation type (through H1-H2 intensity differences) and words' duration. Significant cross-gender differences were obtained for each tested parameter. Moreover, cross-language variations were observed for f0 range, and H1-H2 differences. These results suggest that cross-gender acoustic differences are partly language-dependent and could be socially constructed.

1.3.27 p.310

Hansjörg Mixdorff, Angelika Hönemann, Oliver Niebuhr, Christoph Draxler,
Perceived Prominence Reflected by Imitations of Words with and without F0 Continuity
This paper continues our work on the perception of prominence as a function of F0 continuity. In an earlier study the first author had shown that F0 intervals occurring at lexically stressed syllables and measured using the amplitude of Fujisaki model accent commands strongly contribute to the perceived prominence of that syllable. More recent work explored how F0 continuity influenced prominence ratings of single word utterances. The outcome indicated that listeners made use of the physically available F0 information and therefore words containing gaps in the contour were perceived as less prominent. It was also shown that subjects were able to interpolate missing parts as long as the F0 peak was still present. The current study explores whether subjects compensate the lack of prominence in words containing F0 gaps by asking them to produce a word with the same accent strength as that of a spoken word stimulus, the spoken word being either the same or different from the one they are asked to utter. We evaluated word durations, F0 intervals and intensities of the responses as correlates of prominence and found that listeners indeed seem to adjust depending on the kind of stimulus they have heard.

1.3.28 p.315

Toshiyuki Sadanobu,
The Structure of Japanese Phrase in Accordance with Speaking Modes
While English is often spoken in an increment of clause (i.e. subject and predicate), Japanese of a smaller phrase called "bunsetsu" (e.g. noun phrase and case particle). Previous studies on Japanese language, however, have traditionally been focusing on clause structure, and little attention has been paid on the

structure of “bunsetsu” (non-predicate one, especially). This paper describes the basic structure of non-predicate “bunsetsu” from grammatical point of view, and elucidates that the structure of non-predicate “bunsetsu” varies in accordance with four speaking modes ((i) Sentence mode A; (ii) Sentence mode B; (iii) “Bunsetsu” mode; and (iv) Character mode), which are identified on the criteria of compatibility among seven phenomena attested in Japanese speech. To be more concrete, this paper shows that it is only the mode (iii) that enables copula, “bunsetsu”-final particle (“Kantoujoshi” in Japanese), final leaping, and combination of breaking and prolongation in non-predicate “bunsetsu”).

2 Day Two - May 21st

2.1 Wednesday Session One

May 21st, 9am - 10:30am : 2-1-plenary (1+3 presentations)

2.1.1 KeyNote 2 (p.64)

Stefanie Shattuck-Hufnagel 30-min
Cue-based analysis of speech: implications for prosodic labelling systems

Over the past few decades it has become clear that an adequate account of systematic context-driven variation in word forms requires representations below the level of the abstract symbolic phoneme or even the allophone. One proposal for this sub-allophonic level of description is in terms of feature cues, such as the cues to articulator-free features and articulator-bound features proposed by Halle (1992) and by Stevens (2002), also assumed in the concept of enhancing cues in Stevens and Keyser (2010), Keyser and Stevens (2006) and Stevens, Keyser and Kawasaki (1986). This proposal of a level of representation of discrete feature cues, along with continuous-valued cue parameters, has the potential to bridge the gap between abstract symbolic categories of the phonology and the concrete spatial and temporal specifications that drive the articulatory-acoustic implementation of word forms in continuous communicative speech. Such an approach suggests that phonetic transcription might benefit from a focus on capturing the individual cues to feature contrasts that are realized in the speech signal. Does this approach to understanding phonetic variation in word forms have implications for prosodic labelling? We will explore this possibility, taking as our point of departure Arbisi-Kelm’s (2006) proposal for labelling the separate correlates of prosodic disfluency in stuttered speech, adapted by Brugos and Shattuck-Hufnagel (2012) for prosodic disfluencies in utterances produced by typical speakers. Our hypothesis is that variation in cue selection and cue parameter values is systematically governed by context, and that cue-level transcription may be needed to capture systematicity in the phonetic implementation of prosodic phonology as well as of lexical phonology.

2.1.2 p.321

Yuki Asano,
Stability in perceiving non-native segmental length contrasts
Previous studies have demonstrated that listeners show high sensitivity in discriminating non-native prosodic contrasts thanks to auditory memory (Hayes and Masuda 2008; Hirano 2011). We tested the limits of discriminating Japanese consonantal length contrasts with three groups of listeners (German learners of Japanese, German non-learners and Japanese natives) under increasing task demands. We increased auditory memory load through a longer inter-stimulus interval (=ISI) (2500ms vs. 300ms) and added psycho-acoustic complexity (trials with task-irrelevant pitch falls that occurred simultaneously with the consonant vs. with monotonous pitch). Results showed very good discrimination in all groups when task demands were lowest. With increasing task demands, only non-natives’ discrimination abilities decreased: non-learners were strongly affected by both ISI and pitch, while learners only by pitch. The psycho-acoustic complexity of the stimuli had a stronger impact on performance than increased memory load. Our findings suggest that L2 learners can establish novel phonological representations, but the ability to use them can be applied still only under favorable listening conditions with no distracting acoustic information. The non-native listeners’ reduced sensitivity under increasing task demands appears to be the reason why even advanced learners still face difficulties in natural learning situations.

2.1.3 p.326

Jessica Siddins, Jonathan Harrington, Ulrich Reubold, Felicitas Kleber,
Investigating the relationship between accentuation, vowel tensity and compensatory short-

ening
The aim of this study was to investigate the relationship between compensatory shortening and coarticulation in German tense and lax vowels and to determine whether this relationship was influenced by prosodic accentuation. While previous studies focussed on temporal vowel reduction due to compensatory shortening, and often found conflicting results, our study extends previous results by including a formant analysis of spatial reduction in two types of compensatory shortening. Polysyllabic shortening was tested in monosyllabic versus disyllabic words, while incremental coda shortening was tested in words with final singleton versus final cluster. Speakers produced minimal pairs differing in vowel tensity in accented and deaccented contexts for both shortening conditions. Vowel duration was influenced primarily by vowel tensity as well as by accentual lengthening for tense but not lax vowels. While vowel duration was not affected by compensatory shortening, formant analyses revealed an effect of coda cluster for tense vowels as well as clear effects of accentuation and vowel tensity. There was no effect of polysyllabic shortening on formants. Further to previous studies on compensatory shortening, these results reveal that compensatory shortening is not limited to temporal reduction, but can have an impact on vowel quality as well.

2.1.4 p.331

Amanda Ritchart, Amalia Arvaniti,
The form and use of uptalk in Southern Californian English
This study examines the phonetics, phonology and pragmatic function of uptalk, utterance-final rising pitch movements, as used in Southern Californian English. Twelve female and eleven male speakers were recorded in a variety of tasks. Instances of uptalk were coded for discourse function (statement, question, confirmation request, floor holding) based on context. The excursion of the pitch rise and the distance of the rise start from the onset of the utterance's last stressed vowel were also measured. Confirmation requests and floor holding showed variable realization. Questions, on the other hand, showed a rise that typically started within the stressed vowel and had a large pitch excursion, while uptalk used with statements exhibited both a smaller pitch excursion and a later rise that often started after vowel offset. This pattern suggests that statements have a L* L-H% melody while questions have L* H-H%. Gender differences were also found: female speakers used uptalk more often than males, and showed greater pitch excursion and later alignment, all else being equal. Other social parameters, however, such as social class and linguistic background did not affect the use of uptalk.

2.2 Wednesday Session Two

11am - 1pm : 2-2-oral (6 presentations)
- speech rhythm and timing -

2.2.1 p.337

Agnieszka Wagner,
Rhythmic structure of utterances in native and non-native Polish
This paper presents results of an ongoing study concerning speech rhythm in native and non-native Polish. The goal of the analyses described in the paper was to characterize rhythmically Polish utterances realized by native and non-native speakers with German and Korean accent. The analyses are limited to the domain of duration, but in the future other prosodic parameters will also be investigated. In the current study, different rhythm metrics (%V, V, C, PVIs and Varcos) were applied to provide quantitative description of temporal patterning in native and non-native Polish. Following the assumption that perceived speech rhythm is the effect of meter and grouping which are closely related to prominence and phrasing, durational marking of various levels of prominence and prosodic edges was also analyzed between the three accents (native Polish and German- and Korean-accented Polish). The analyses aimed also at rhythmic classification of Polish - for that purpose the results of quantitative description with rhythm metrics and phonotactic properties of the speech material used in the current study were compared with the data for other languages presented in the literature.

2.2.2 p.342

Hugo Quené, Rosemary Orr,
Long-term convergence of speech rhythm in L1 and L2 English

When talkers from various language backgrounds use L2 English as a lingua franca, their accents of English are expected to converge, and talkers' rhythmical patterns are predicted to converge too. Prosodic convergence was studied among talkers who lived in a community where L2 English is used predominantly. Speech rhythm was operationalized here as the peak frequency in the spectrum of the intensity envelope, normalized to the speaking rate (in syll/s). Results indicate that talkers produced intensity contours with maximum periodicity at frequencies of about 0.32 times their syllable rates, i.e., peaks in intensity tend to occur every 1/0.32 syllables. These results were collected repeatedly, from 5 recordings conducted over 3 years with the same talkers. We found that variance between talkers in their rhythm decreases over time, thus confirming the predicted convergence in speech rhythm in L2 English. These findings show that speech rhythm in L2 English tends to converge, and that this prosodic convergence continues to proceed over several years, as well as over communicative settings.

2.2.3 p.346

Andreas Windmann, Juraj Šimko, Petra Wagner,

Probing Theories of Speech Timing using Optimization Modeling

We implement two theories about the temporal organization of speech in an optimization-based model of speech timing and conduct simulation experiments in order to test whether both theories can account for the phenomenon of foot-level shortening (FLS) observed in English speech corpora. Results suggest that a model that induces compensatory timing relations between syllables and feet predicts empirical results very accurately. However, we also observe that the FLS effect can equally well be explained under the assumption that suprasegmental timing is confined to localized lengthening effects at the heads and edges of prosodic domains. Implications for theories of speech timing are discussed.

2.2.4 p.351

Sandra Peters, Felicitas Kleber,

The influence of accentuation and onset complexity on gestural timing within syllables

This paper presents results from a production experiment using electromagnetic articulography. The main aim of the study was to investigate how phrasal accent and the number of onset consonants influence the gestural timing of syllable constituents in German. Five speakers of German with sensors attached to the tongue tip, tongue body and lower lip were recorded reading sentences with either accented or unaccented target words that contained simplex (one consonant) and complex (two consonants) onsets. The nucleus was always /a/ and the coda consonant was always /p/. We analyzed acoustic segment duration and gestural overlap (in terms of lag measurements). Onset complexity influenced both CV and VC overlap and accentuation affected gestural overlap to a greater extent than acoustic vowel duration. However, the extent of overlap differed between segment sequences and accentuation patterns: while for CC and VC sequences trends for greater overlap in deaccented than in accented condition were found, CV overlap decreased with deaccentuation. Shorter plateau durations in this context explain the diminished CV overlap in a prosodically weak context. The findings are discussed with respect to the predictions made by articulatory phonology regarding gestural timing and with respect to timing stability in weak versus strong prosodic contexts.

2.2.5 p.356

Núria Esteve-Gibert, Joan Borrs-Comes, Marc Swerts, Pilar Prieto,

Head gesture timing is constrained by prosodic structure

There is an increasing consensus to regard gesture and speech as parts of an integrated communication system, in part because of the findings related to their temporal coordination at different levels. In general, results for different types of gestures show that the most prominent part of the gesture (the apex) is typically aligned with accented syllables. The aim of the present study is to test for this coordination by focusing on head movements taken from a semi-spontaneous setting in order to look at the effects of upcoming phrase boundaries on their timing. Our results show that while apexes of head gestures are synchronized with accented syllables, upcoming phrase boundaries have an effect on the timing of three gestural points, namely the start, apex, and end time of head gestures. Crucially, these points are aligned differently with respect to the stressed syllable for trochees as compared with iambs/monosyllables, showing that head nods are retracted before upcoming phrase boundaries. This result corroborates previous results by Esteve-Gibert &Prieto for pointing gestures in laboratory settings.

2.2.6 p.361

Helen Türk, Pärtel Lippus, Karl Pajusalu, Pire Teras,
The ternary contrast of consonant duration in Inari Saami
The three-way distinction of quantity occurs in several Finnic and Saami languages. The paper focuses on the length contrast of consonants in Inari Saami. Similarly to Estonian and other Finno-Ugric languages where three quantities are described, in Inari Saami the distinction between single consonants, short geminates or consonant clusters, and long geminates or consonant clusters appears only on the boundary of a stressed and unstressed syllable of a disyllabic foot. Our results show that in Inari Saami the duration of consonants is inversely related to the duration of both preceding and following vowels, and there is a tendency towards foot isochrony. The results are in line with previous studies on quantity opposition in Inari Saami and in other Finnic languages, showing the ternary distinction of consonant quantities as a foot-level feature of the language.

2.3 Wednesday Session Three

2pm - 4pm : Special Session - Slavic Prosody

2.3.1 p.366

Štefan Beňuš - 30-min (invited)
Slovak prosody in the phonetics-phonology debate: Yers and emergent prosodic breaks.
Prosody is central for understanding the cognitive system underlying human speech and relates to both more granular aspects of our phonological competence as well as more continuous aspects of observable articulatory movements and resulting acoustic characteristics. The understanding, and formal treatment, of the relationship between these two inter-related components of human speech is at the core of the cognitive approach to speech. In this presentation I contribute to this discussion by drawing links between two seemingly unrelated lines of my research on Slovak, and argue that understanding the continuous prosodic nature of speech is critical for improving our understanding of cognitive competence underlying it. The first aspect concerns yer vowels as the prototypical problem of Slavic phonology, the second involves the nature of prosodic boundaries.

2.3.2 no paper

Tamara Rathcke - 30-min (invited)
Time and timing in intonational phonology: analysing pitch categories in Russian (and other Slavonic languages)
Many Slavonic languages are still lacking a comprehensive description of their intonational phonologies. Given that decisions regarding the number of relevant categories and their types are the key issues of any phonological analysis, this presentation will concentrate on how time and timing can inform intonational phonology. Evidence from Russian (and also Bulgarian, Czech and Polish) will demonstrate that time pressures arising from intermittent voicing and an upcoming phrase boundary have different effects on Slavonic vs. Germanic languages. Potential implications of these findings for prosodic typology will be discussed.

2.3.3 p.368

Jaye Padgett, - 30-min (invited)
On the origins of the prosodic word in Russian
The Prosodic Word is a foundational notion in phonological theories, being relevant for the statement of many phonological generalizations. In spite of their importance, there are basic open questions about prosodic words. Where do they come from? Can their structure in one language vs. another be predicted? In this paper I suggest a research program that attempts to address such questions by viewing prosodic words as emergent over time from the interaction of phonetics, phonologization, and syntactic structure.

2.3.4 p.372

Bistra Andreeva, Jacques Koreman, William Barry,
Local and Global Acoustic Correlates of Information Structure in Bulgarian
In this study the prosodic exponents of information structure are examined in the production of six Bulgarian sentences under different focus conditions (broad focus and non-contrastive and contrastive

narrow focus). Results show that speakers consistently discriminate broad and narrow focus by both local and global acoustic cues. Local cues are the phonetic properties of the accented syllables, while global cues reflect broader phonetic patterns in the intervals before and after the accented syllable, which vary independently of the tonal accent. Contrastive and non-contrastive accents are differentiated exclusively by local cues, but only when the focus is early in the sentence.

2.3.5 p.377

Agnieszka Wagner,
Description of Polish speech rhythm using rhythm metrics and time-delay approach: A comparative study
The goal of this study is to provide a multidimensional description of rhythmic structure of Polish utterances. For this purpose a time-delay approach proposed in [1] is applied and results of qualitative and quantitative analyses based on time-delay plots are compared with results obtained with selected rhythm metrics. The study shows that description that relies on a combination of rhythmic scores is inconclusive and difficult to interpret, because it does not account for rhythmic structuring nor grouping. The time-delay approach, on the contrary, appears to be very efficient in exploring short-time and long-term timing variability that determines Polish speech rhythm.

3:40pm - 4:00: Panel discussion

2.4 Wednesday Session Four

4:30pm - 6pm : 2-4-poster (48 presentations)

- perception and intonation -

2.4.1 p.383

Marion Aguilera, Radouane El Yagoubi, Robert Espesser, Corine Astésano,
Event-Related Investigation of Initial Accent Processing in French
This study investigates stress processing through the Event-Related brain Potential (ERP) technique. It aims at evaluating whether French listeners can perceive and discriminate the Initial Accent (IA) and whether IA is encoded in the phonological representation. Participants listened to trisyllabic words in two stress-pattern conditions, with (+IA) or without (-IA) initial accenting, in an oddball paradigm. The EEG was recorded in both a passive and an active listening task, and in two different oddball versions: one where standard stimuli were +IA words and deviants -IA words, and the reverse for the other version (-IA standard, +IA deviant). Behavioral results show faster processing and less errors for +IA stimuli. ERP results show larger MisMatch Negativity component for -IA words, pointing out 1) that French listeners are sensitive to f0 manipulation, and 2) that +IA is the preferred stress template in French. Altogether, our results indicate that French listeners not only discriminate stress patterns but that IA is encoded in long-term memory, hence phonologically relevant.

2.4.2 p.388

Alejna Brugos, Jonathan Barnes,
Effects of dynamic pitch and relative scaling on the perception of duration and prosodic grouping in American English
Results of two perception experiments suggest that using timing measures alone to compute prosodic structure misses valuable information from pitch. Previous research showed that pitch can distort perceived duration: tokens with dynamic or higher f0 are perceived as longer than comparable level-f0 or lower-f0 tokens, and silent intervals bounded by tokens of widely differing pitch are heard as longer than those bounded by tokens closer in pitch (the kappa effect). Phrase edges (signalled by increased duration, pause, phrase tones, and f0 reset) set the scene for pitch to modulate perceived duration. Two new experiments used the same duration and f0 manipulations (level vs. varying-slope rises, at varying pitch ranges) of segmentally-identical base files, in two separate tasks: 1) a linguistic grouping task using an ambiguously-structured phrase and 2) a psychoacoustic study on perceived duration. Results show that effects on perceived duration due to dynamic pitch can be either strengthened or nullified depending on relative scaling of compared tokens. These same manipulations push grouping judgments beyond what

would be expected from distortions of perceived duration. This suggests that listeners integrate pitch and timing cues when judging linguistic structure, supporting measures of relative boundary size that combine duration and pitch measures.

2.4.3 p.393

Canan Ipek, Sun-Ah Jun,
Distinguishing Phrase-Final and Phrase-Medial High Tone on Finally Stressed Words in Turkish
The goal of this paper is to investigate the nature of the high tones realized on finally stressed words in Turkish. Following Ipek &Jun's [1] AM model of intonational phonology of Turkish, it was hypothesized that the high tone realized on the last syllable of a phrase (i.e., Intermediate Phrase (ip)) is realized differently from that of a phrase-medial prosodic word (PW), reflecting the prosodic hierarchy. Acoustic data show that an ip-final High tone shows larger f0 rise than a PW-final High tone, and the ip-final syllable is longer than the PW-final syllable. Furthermore, the degree of coarticulation is weaker across an ip boundary than a PW boundary. These findings support the prosodic structure and tonal categories proposed in Ipek &Jun's [1] model of Turkish intonation.

2.4.4 p.398

Willemijn Heeren, Vincent van Heuven,
The interaction of accent and boundary tone in perception of whispered speech
We investigated how the perception of Dutch whispered boundary tones depends on the presence of an accent in the utterance-final word, i.e. the boundary tone landing site. Listeners performed near ceiling in normal speech, whereas the same listeners' performance dropped about 30% in whisper, while processing speed decreased in whisper compared to normal speech. Accent position furthermore influenced boundary tone perception. Initial-stress words showed a question bias that affected recognition of that speech act when accent and boundary tone did not coincide. On final-stress words, in which boundary tone and accent coincided, statements and questions were identified equally well.

2.4.5 p.403

Katarina Bartkova, Denis Jouvet,
Links between Manual Punctuation Marks and Automatically Detected Prosodic Structures
This paper presents a study of the links between punctuation and automatically detected prosodic structures, as observed on large speech corpora that were manually annotated during speech transcription evaluation campaigns in French. These corpora contain more than 3 million words and almost 350 thousands punctuation marks. The detection of the prosodic boundaries and of the prosodic structures is based on an automatic approach that integrates little linguistic knowledge and mainly uses the amplitude and the inversion of the F0 slopes as described in [1], as well as phone durations. The paper first analyzes the occurrences of the punctuation marks with respect to various sub-corpora, which also highlights the variability among annotators. Then, the paper focuses on analyzing prosodic parameters with respect to the punctuation marks, followed or not by a pause, and on analyzing the links between the automatically detected prosodic structures and the manually annotated punctuation marks.

2.4.6 p.408

Timo Roettger, Rachid Ridouane, Martine Grice,
Perception of Peak Placement in Tashlhiyt Berber
Previous production studies on Tashlhiyt Berber have demonstrated that questions and statements have similar intonation contours, i.e., a final rise to a F0 peak and subsequent fall. The contours tended to differ in overall pitch register and peak location: questions (a) revealed a stronger tendency to be realized with the F0 peak on the final syllable than statements and (b) even within the same syllable, peaks were often aligned later in questions than in statements. The peak location, however, was reported to vary strongly both within and across speakers, interpreted as free alternation of tonal association. Given this high degree of variation, the question arises as to how relevant this variation is for communication. The present perception study shows that both pitch register (low vs. high) and tonal placement (peak on

penultimate vs. final syllable) affect listeners' judgments on sentence modality as well as reaction times. Whereas peak alignment within the syllable (early vs. late) did not affect judgments, it did have an effect on reaction times. By demonstrating their perceptual impact, this study confirms that the patterns found in production are communicatively relevant.

2.4.7 p.413

Cristel Portes, Uwe Reyle,
The meaning of French "implication" contour in conversation
French intonational contours inventory has a rising-falling tune which presents very interesting semantic properties. It has been called "intonation d'implication" by Delattre (1966) suggesting that the contour triggers an implicit meaning, i.e. an implicature in Gricean terms. Besides, the "implication" contour have been claimed to convey various attitudinal meanings from obviousness to exasperation, and also to mark contrastive focus. The aim of the present paper is to give a unified account of these seemingly differing semantic descriptions of the "implication" contour in French, using a dynamic semantic framework, namely Discourse Representation Theory (DRT). We claim that the main semantic component of the "implication" contour is to convey a contradiction (or a contrast). We first present our DRT-theoretical approach, and then apply it to occurrences of the "implication" contour in a corpus of conversational dialogue.

2.4.8 p.418

César González Ferreras, Carlos Vivaracho-Pascual, David Escudero-Mancebo, Valentín Cardeñoso-Payo,

Combination of variations of pairwise classifiers applied to multiclass ToBI pitch accent recognition
In this paper we present some experiments on multiclass ToBI pitch accent classification. The system is based on the fusion of pairwise classifiers, which are specialized in the distinction of pairs of prosodic labels. Several machine learning techniques, including neural networks, decision trees and support vector machines, are combined in different ways in order to find the best overall combination. Variations of pairwise classifiers are introduced in order to take into account the influence of the samples of the remaining classes during the training of the binary classifiers. The use of these techniques allowed us to improve the results, both the overall classification accuracy and the balance across the different ToBI pitch accent classes.

2.4.9 p.423

Aoju Chen,
Production-comprehension (A)Symmetry: individual differences in the acquisition of prosodic focus-marking
Previous work based on different groups of children has shown that four- to five-year-old children are similar to adults in both producing and comprehending the focus-to-accentuation mapping in Dutch, contra the alleged production-precedes-comprehension asymmetry in earlier studies. In the current study, we addressed the question of whether there are individual differences in the production-comprehension (a)symmetricity. To this end, we examined the use of prosody in focus marking in production and the processing of focus-related prosody in online language comprehension in the same group of 4- to 5-year-olds. We have found that the relationship between comprehension and production can be rather diverse at an individual level. This result suggests some degree of independence in learning to use prosody to mark focus in production and learning to process focus-related prosodic information in online language comprehension, and implies influences of other linguistic and non-linguistic factors on the production-comprehension (a)symmetricity.

2.4.10 p.428

Sandrine Brognaux, Thomas Drugman,
Phonetic variations : Impact of the communicative situation
While speech synthesis research is now focussing on the generation of various speaking styles or emotions,

very few studies have considered the possibility of including phonetic variations according to the communicative situation of the targeted speech (sports commentaries, TV news, etc.). This paper proposes a phonetic analysis of large French corpora to assess the influence exerted by three situational 'traits': read/spontaneous, media/non-media and expressive/non-expressive. It shows that some variations, like elision, tend to be more frequent in spontaneous and non-media speech, conversely to liaisons which appear more often in read and media speech. Interestingly, no phonetic variation draws a clearcut distinction between expressive and non-expressive speech. Finally, a prosodic analysis indicates that the phonetic variations are not directly correlated with the rhythmic features of their corresponding situational 'trait'.

2.4.11 p.433

Netta Weinstein, Konstantina Zougkou, Silke Paulmann,

Differences between the acoustic typology of autonomy-supportive and controlling sentences

The current study was first to describe distinct patterns of prosody that discriminate motivationally laden speech. To do this we applied self-determination theory, a widely used motivational framework. Participants in the US and UK were asked to read out loud either autonomy-supportive sentences (that support choice and volition) or controlling (pressuring and coercive) sentences. Data analyses were conducted using a conservative hierarchical linear modeling approach to account for nesting of sentences within individuals. Across both countries and controlling for gender, autonomy-supportive sentences were read using lower pitch, less intensity, and a slower speech rate than were controlling sentences. Multiple regression analyses showed links between these patterns of prosody for each participant and his or her current level of motivation, providing additional validity to results. Findings inform both the motivation and prosody literatures and offer a first description of how different kinds of motivational speech may sound.

2.4.12 p.438

Jasmin Pfeifer, Silke Hamann, Mats Exter,

Congenital Amusia in linguistic and non-linguistic pitch perception: What behavior and reaction times reveal

Congenital Amusia is a developmental disorder that has a negative influence on pitch perception. While it used to be described as a disorder of musical pitch perception, recent studies indicate that congenital amusics also show deficits in linguistic pitch perception. This study investigates the perception of linguistic and non-linguistic pitch by ten German amusics and their matched controls. To test the influence of amusia on linguistic pitch perception, the present study parametrically varied pitch differences in steps of one semitone in resynthesized statement-question pairs. In addition, we looked at the influence of stimulus duration, continuity of pitch and direction of pitch change (statement or question). Performance accuracy and reaction times were recorded. Behavioral results show that amusics performed worse than controls over all conditions. The reaction time analysis supports these findings, as amusics were significantly slower across all conditions. Both groups were faster in discriminating statements than questions. Performance accuracy supports these findings, as questions were also harder to discriminate. The present results warrant further investigation of the linguistic factors influencing amusics' perception of intonation.

2.4.13 p.443

Jue Yu, Dafydd Gibbon, Katarzyna Klessa,

Computational annotation-mining of syllable durations in speech varieties

There are many techniques for modelling properties of speech duration patterns, including models of rhythm as oscillation, partial models of rhythm types as departures from isochrony, models of tempo acceleration and deceleration, and models of duration hierarchies and their relation to hierarchies in word and phrase structure. Except for oscillator modelling, many approaches use data extraction from speech annotations, often with mainly manual methods. We employ computational data-mining for phonetic research, as opposed to phonological research on the one hand or speech technological research on the other, and explore the potential of the computational annotation data-mining paradigm for improving efficiency and scope of analysis. We show consistent variation in syllable duration patterns in selected speech varieties in English, Chinese and Polish, chosen for their known different prosodic typological properties. Results include a possible limen of 50ms for relevant timing patterns. For data-mining we use the Time Group Analysis (TGA) methodology, directly in the TGA online tool and integrated into

the Annotation Pro+TGA desktop software.

2.4.14 p.448

Izabel Seara, Juan Manuel Sosa, Vanessa Nunes,
Sentence type and prenuclear contours in Brazilian Portuguese: production and perception
In this paper we examine how the interrogative sentence mode is encoded in some dialects of Brazilian Portuguese (BP) and how questions differ from their declarative counterparts. Our aim is to identify which specific prosodic features, including prenuclear pitch range values, are systematically associated with the interrogative mode of enunciation. In the interdialectal comparison, the speakers from Blumenau (SC) and Aracaju (SE) distinguish themselves from the speakers of the other varieties in their prenuclear patterns, significantly higher for the yes/no interrogatives than the declarative counterparts. This is not the case with other dialects in our study. Perception tests corroborated the production results.

2.4.15 p.453

Ji Young Kim,
Use of suprasegmental information in the perception of Spanish lexical stress by Spanish heritage speakers of different generations
The present study examines the perception of Spanish lexical stress by Spanish heritage speakers of different generations and compares their performance to that of Spanish native controls and English second language (L2) learners of Spanish. Previous studies have shown that English L2 learners experience great difficulty in perceiving Spanish lexical stress. Such difficulty is argued to be derived from English listeners using different strategies from Spanish listeners in the perception of stress. Given that Spanish heritage speakers share the same dominant language with English L2 learners (English), but differ from them with regard to the first language (Spanish), the present study intends to seek whether heritage speakers show similar or different patterns when compared with L2 learners. The present study also intends to account for the heterogeneity among heritage speakers by comparing heritage speakers of different generations. Using a forced-choice identification task with stressed minimal pairs of paroxytone and oxytone verbs, results showed that while 1st generation US-born heritage speakers pattern like Spanish native controls by paying more attention to the acoustic cues of the stimuli, 1.5/2nd generation US-born heritage speakers pattern like English L2 learners by showing bias towards paroxytone verbs.

2.4.16 p.457

David Escudero, Lourdes Aguilar, César González Ferreras, Valentín Cardeñoso, Yurena Gutierrez,
Applying a fuzzy classifier to generate Sp_ToBI annotation: preliminar results
One of the goals of the Glissando research project1 is to enrich a radio news corpus [1] with Sp_ToBI labels. In this paper we present the application of the automatic predictions of a fuzzy classifier to speed the labeling process. The strategy is proposed after completing the following steps: a) manual annotation of a part of the Glissando corpus with Sp_ToBI labels and checking of the coherence of the labels; b) training of the automatic system; c) validation or correction of the automatic system's predictions by a human expert. The automatic judgments of the classifier are enriched with confidence measures that are useful to represent uncertain situations concerning the label to be assigned. The main aim of the paper is to show that there exists a correspondence between the uncertain situations that are identified during an inter-transcriber experiment and the uncertain situations that the fuzzy classifier detects. Labeling time reduction encourages the use of this strategy.

2.4.17 p.462

Marie-Catherine Michaux, Sandrine Brognaux, George Christodoulides,
The production and perception of L1 and L2 Dutch stress.
This study aims at exploring the production and perception of Dutch word stress by Francophone learners of (Belgian) Dutch. For this purpose a production experiment was first carried out. In line with other studies, it was hypothesized that participants would show a tendency to stress the final syllable. Even though this hypothesis was confirmed, there was also a substantial lack of agreement between the five labellers who perceptually annotated the data for stress position. To further investigate this matter, acoustic measures were extracted. The data suggest that both groups of speakers do not use acoustic

correlates to signal prominence in the same way, the Dutch group using intensity, vocalic nucleus duration and pitch movement more, while the French group prefers duration and pitch movement. This study also led us to develop tools to phonetise, syllabify and facilitate the acoustic analysis of Dutch speech.

2.4.18 p.467

Takayuki Kagomiya, Seiji Nakagawa,

Evaluation of bone-conducted ultrasonic hearing-aid regarding transmission of speaker gender and age information

Human listeners can perceive speech signals in a voicemodulated ultrasonic carrier from a bone-conduction stimulator, even if the listeners are patients with sensorineural hearing loss. Considering this fact, we have been developing a bone-conducted ultrasonic hearing aid (BCUHA). The purpose of this study was to assess the usefulness of the BCUHA in transmission of speakers' physical attributes: gender and age. The evaluation used gender and age-identification experiments. The experiments were also conduced under air-conduction (AC) and cochlear implant simulator (CIsim) conditions. The results showed that: the BCUHA can well transmit speakers' gender information; the BCUHA can transmit speaker age information better than CIsim.

2.4.19 p.472

Mara Breen, Sarah Weidman, Katharine Guarino,

Rhythm and Expression in The Cat in the Hat

In recent years, there has been increasing interest in whether rhythmic interventions support young children's literacy development [1]. To begin to explore this connection, we assessed several aspects of rhythmicity and expressivity of productions of the notably rhythmic and rhyming children's book, The Cat in The Hat by Dr. Seuss. Participants subjectively rated either the rhythmicity or expressivity of speech taken from recordings of the book read aloud. These perceptual ratings were correlated with acoustic measures of rhythmicity and expressivity. Moreover, we observed a surprising lack of consistency between perceptual ratings of rhythmicity and expressivity. However, we observed a consistent relationship between the perceptual ratings of the first couplet of verses and the second. These findings can inform our investigation of the role of rhythm in literacy development.

2.4.20 p.477

Laurence White, Sven Mattys, Linda Stefansdottir, Victoria Jones,

Lengthened Consonants are Interpreted as Word-Initial

Prosody facilitates listeners' segmentation of the speech stream into a sequence of words and phrases. With regard to speech timing, vowel lengthening is interpreted as a cue to an upcoming boundary, in accordance with the iambic-trochaic law. However, the impact of consonant lengthening on segmentation, in the absence of other boundary cues, has not been tested. In a series of artificial language learning experiments, we examined how durational variation affects listeners' extraction of novel trisyllables defined by transition probabilities. In line with previous research, syllables containing lengthened vowels were interpreted by listeners as word-final. However, syllables with lengthened onset consonants were interpreted as word-initial. Thus, the structural interpretation of durational variation depends upon localization: longer vowels cue a following boundary; longer consonants cue a preceding boundary.

2.4.21 p.482

Alexandra Markó, Mária Gósy, Tilda Neuberger,

Prosody patterns of feedback expressions in Hungarian spontaneous speech

Speech communication incorporates non-verbal signals and semi-lexical vocal phenomena as well as words used as the listener's responses to the speaker's message. They are most common in conversation with various functions regardless of language. A specific subcategory is feedback expressions (FEs) that can be found in the listener's production as well as in the current speaker's speech production when reacting to the former speaker's message. This paper reports on the temporal and intonational characteristics of four types of FEs identified in 20 interviews and conversations from the BEA Hungarian database. Altogether 262 occurrences were categorized into four discourse functions signaling 'attention', 'comprehension', 'agreement' and 'other attitude'. Durations showed statistically significant differences across

discourse functions. They were significantly longer in females than in males in all functions. The pitch range data revealed a statistically significant difference depending on discourse function and gender only in the case of the 'attention' function. The dominant frequency contour was a rise in the functions of 'attention' and 'agreement' (90%). The same contour was observed only in 75.5% of the 'comprehension' function. An integrated approach is proposed to analyze these phenomena in spontaneous speech.

2.4.22 p.487

Eduardo Patricio Velázquez Patiño,
Intonation Patterns of Morelos Nahuatl
There are still relatively few studies on the phonetics and phonology of the indigenous languages of Mexico, and just a minority of them deals with less explored areas like prosody or, specifically, intonation. This study reports a preliminary analysis of Nahuatl intonation, taking into account its phonological characteristics: a) trochaic binary rhythm; b) generation of secondary stress inside rhythmic structures; c) generation of rhythmic groups according to clause structures; d) phonetic syllable lengthening at the end of sentences; e) laryngealization or voicelessness at the end of utterances, and f) vowel lengthening. Data collected by means of different methods, developed in order to obtain authentic and spontaneous utterances, show that different sentence types tend to have specific intonation patterns with many typologically common features and some original characteristics.

2.4.23 p.497

Amelia Kimball, Jennifer Cole,
Avoidance of Stress Clash in Perception of American English
We examine stress clash in perception, asking to what degree listeners perceive speech as metrically regular. We assess metrical regularity through a stress perception task carried out by untrained listeners annotating transcripts of spontaneous conversation and sentences designed to be metrically regular. Results show listeners report perceiving fewer stress clashes than predicted by random placement of stresses or by concatenating the citation form stress patterns of each individual word in a given sentence. These results suggest that listeners perceive spontaneous conversational English as metrically regular.

2.4.24 p.502

Lenka Weingartová, Eliška Churaňová, Pavel Šturm,
Transitions, pauses and overlaps: Temporal characteristics of turn-taking in Czech
This study aims to describe temporal characteristics of pausing and turn-taking phenomena in conversation. The material comes from the VASST corpus of contemporary Czech and uses four spontaneous dialogues in the form of an informal interview. We describe both general and idiosyncratic effects found in our data and compare them with results from other languages. In our material, transitions with a silent gap, overlaps and back-channels all display notably similar durational distributions with the median around 360 ms and a marked skewing. The four dialogues did not differ in the proportion of turns belonging to the interviewer (58%) vs. interviewee (42%), which is hypothesized to characterize the experimental task. Despite a number of general tendencies, individual differences in pausing and turn-taking behaviour of the speakers were found as well. For instance, the ratio of pauses and gap transitions proved to be highly dialogue-specific. We also gathered evidence for a substantial change in the speech behaviour of the interviewer resulting from a change of her communication partner.

2.4.25 p.507

Riikka Ullakonoja, Mikko Kuronen, Pertti Hurme, Hannele Dufva,
Segment Duration in Finnish as Imitated by Russians
The paper reports findings of a study in which Russian speakers without any prior knowledge of the language imitated Finnish utterances, and, in particular, how they succeeded in imitating segmental duration. The data was analyzed using acoustic measurements of segment duration as well as auditory analysis by four judges. The results show that while Russian speakers faced difficulties in imitating some aspects of the Finnish quantity, many imitated words were judged as comprehensible.

2.4.26 p.512

Candide Simard, Claudia Wegener, Albert Lee, Faith Chiu, Connor Youngberg,
Savosavo word stress: a quantitative analysis
This paper presents a quantative analysis of stress in Savosavo (unclassified), an endangered language spoken on Savo Island, (Solomon Islands). Acoustic analyses comprise the measurements of F0, duration, and intensity for each syllable in a dataset carefully selected from elicited speech from one speaker only, aiming to test the effect of increasing morphological complexity on stress realization in a system that displays some variation. Statistically significant variation is found in all correlates between stressed and unstressed syllables, thus fitting with widely attested manifestations of stress cross-linguistically. Findings were further tested with a re-synthesis tool, to confirm our initial hypotheses. Our results demonstrate that the current annotation scheme is a reliable representation of the data, and that the qTA component embedded in PENTAtrainer is effective in modelling F0 contours, even with less controlled data as input. We will argue for the usefulness of instrumental phonetic investigations in describing lesser-known languages, to enhance our understanding of the characterization of the prosodic systems of the world's languages.

2.4.27 p.515

Mortaza Taheri-Ardali, Hamed Rahmani, Yi Xu,
The Perception of Prosodic Focus in Persian
In a previous production experiment, post-focus compression (PFC) of F0 and intensity were found to be present in Persian. It was also shown that F0 and duration were the main correlates of prosodic focus in Persian. However, the perceptual relevance of PFC in Persian was not yet clear. The present paper reports the findings of an experiment on focus perception in Persian. Native speakers of Persian listened to sentences produced with focus in different positions as well as the neutral-focus sentence, and judged the presence and location of focus. Results show that final focus is identified much less well than other types of focus, and most of its confusion is with neutral focus. This shows that the presence of PFC is a main factor in recognizing prosodic focus in Persian.

2.4.28 p.520

Catherine Lai,
Final Rises in Task-oriented and Conversational Dialogue
This paper examines the distribution of utterance final pitch rises in dialogues with different task structures. More specifically, we examine map-task and topical conversation dialogues of Southern Standard British English speakers in the IViE corpus. Overall, we find that the map-task dialogues contain more rising features, where these mainly arise from instructions and affirmatives. While rise features were somewhat predictive of turn-changes, these effects were swamped by task and role effects. Final rises were not predictive of affirmative responses. These findings indicate that while rises can be interpreted as indicating some sort of contingency, it is with respect to the higher level discourse structure rather than the specific utterance bearing the rise. We explore the relationship between rises and the need for co-ordination in dialogue, and hypothesize that the more speakers have to co-ordinate in a dialogue, the more rising features we will see on non-question utterances. In general, these sorts of contextual conditions need to be taken into account when we collect and analyze intonational data, and when we link them to speaker states such as uncertainty or submissiveness.

2.4.29 p.525

Philippe Martin,
Spontaneous speech corpus data validates prosodic constraints
In the Autosegmental-Metrical model, the prosodic structure is defined as a hierarchy of Accent Phrases (AP). Groups of AP form intermediate prosodic phrases ip, which in turn are grouped into Intonation Phrases IP, and finally sequences of IP form the sentence intonation unit. In this hierarchy several constraints affect the prosodic structure, such as the AP 7 syllables rule, the stress clash conditions, eurhythmicity and syntactic clash. These constraints have been established essentially from read sentences data. They lead to an experimental justification in the observed synchronization of AP's syllabic chunking by Delta brain waves. This paper investigates the validity of the prosodic structure constraints on spontaneous speech data in French, as well as the adequacy of the Delta waves characteristics to

synchronize AP data.

2.4.30 p.530

Michael Phelan,
Hearing the Structure of Math: Use and Limits of Prosodic Disambiguation for Mathematical Stimuli
Listeners use the prosodic cues of an utterance to help determine its syntactic structure, but how does this process happen in the specialized domain of mathematics? Mathematical expressions can contain deeply embedded structures, and listeners encounter read mathematical expressions (RMEs) far less frequently than other potentially ambiguous utterances. How does experience with listening to math affect our ability to hear the structure of an RME via its prosody? Are there limits to the amount of structure we can pull out of the prosody of an utterance? A perception experiment was conducted with subjects aged 7-59 to help answer these questions. Participants heard recordings of RMEs and attempted to determine which of two or more mathematical structures the reader intended. When subjects chose between two options for phrases like nine times A minus two, they chose the mathematical expression that had bracketing matching the prosody of the utterance. However, for more complex phrases like the square root of sixteen over A plus twelve, results were at chance. Age played a surprising role: subjects' performance increased dramatically from age 7 to 16, but adults' performance varied widely. This is attributed to variation in exposure to read mathematics.

2.4.31 p.534

Sujan Kumar Roy, Md. Khademul Islam Molla, Keikichi Hirose,
Robust Pitch Estimation using Ensemble Empirical Mode Decomposition
This paper presents an efficient pitch estimation algorithm for noisy speech signal using ensemble empirical mode decomposition (EEMD) based time domain filtering. The dominant harmonic of noisy speech is enhanced to make pitch period more prominent. The normalized autocorrelation function (NACF) of the modified signal is then decomposed into time varying subband signals using EEMD. In contrast to the ordinary EMD, it does not introduce any mode mixing during decomposition. The subbands containing pitch component are selected and separated yielding partially reconstructed signal. The pitch period is determined from thus separated signals. The experimental results show that the proposed algorithm performs better compared to other recently reported algorithms in noisy environment.

2.4.32 p.539

Mónica Domínguez, Mireia Farrús, Alicia Burga, Leo Wanner,
The Information StructureProsody Language Interface Revisited
Several grammar theories relate information structure and prosody, highlighting a major correspondence between theme and rheme, and intonation patterns. Although these theories have been successfully exploited in some specific speech synthesis applications, they are mainly based on short default-order sentences, which limits their expressiveness for real discourse with longer sentences and complex structures. This paper revises these theories, identifying cases in which they are valid, and providing a new proposal for cases in which a more complex model is needed. Specifically, our experiments performed on real discourse from the Wall Street Journal corpus show that we need a model that: (1) foresees a hierarchical theme/rheme structure, and (2) introduces, apart from the traditional theme and rheme, a new element—the specifier.

2.4.33 p.544

Bahia Guellai, Alan Langus, Marina Nespor,
Prosody is perceived in the gestures of the speaker
It has been suggested that speech and hand gestures could form a single system of communication that facilitates the interaction between the speaker and the listener. What kind of information do gestures carry? In the present study, two experiments test the possibility that spontaneous gestures accompanying speech carry prosodic information. Experiment 1 shows that gestures provide prosodic information as adults are able to perceive the congruency between a low-pass filtered thus unintelligible - speech stream and the gestures of the speaker. These results show that prosody is not a modality specific phenomenon

and can be perceived in spontaneous gestures that accompany speech.

2.4.34 p.548

Joseph Tyler,
Rising pitch and quoted speech in everyday American English
Phonetic variation in rising pitch has been analyzed for how it correlates with contextual factors like speaker gender, utterance type (questions vs. statements) and turn position (turn-medial vs. turn-final). This paper analyzes variation in terminal rising pitch between quoted and non-quoted speech, using data from the Santa Barbara Corpus of Spoken American English. Results show rises in quoted speech start and end higher, rise more overall, but are no different in duration. These results are gender-dependent, however, for while women produce 65% of all rises in the corpus sample, they produce 100% (n=23) of the quoted speech rises.

2.4.35 p.553

Daniel Aalto, Stina Ojala,
Fine temporal structure of Finnish sign language
Signs can be divided to syllables and further into transitions and nuclei based on the signing flow of the handshapes. Here, a mixed effects linear regression model is used to describe the variation in the duration of the syllable nuclei in a data set of 341 signs (474 syllables) produced by five native FinSL signers during a map task. The phonetic fixed variables are the duration of the adjacent transients and syllable nuclei; phonological fixed variables are the syllabic length of the sign, the syllable position within the sign, and the sign type (functional or content bearing). Both preceding and following nucleus had a significant effect on the nucleus duration, while an asymmetric effect was found for the transitions: only the postnuclear transition had a significant effect. The syllable structure had no effect. However, the nuclei were shorter in function signs. These results suggest that signs are produced in two stages where the first stage, preparatory transition, is merged with the production of the previous syllable, and the second stage consists of executing the sign.

2.4.36 p.558

Hiroko Muto, Yusuke Ijima, Noboru Miyazaki, Hideyuki Mizuno,
Pause insertion prediction using evaluation model of perceptual pause insertion naturalness
This paper describes a pause insertion prediction technique for generating more natural synthesized speech for text-to-speech (TTS) synthesis systems. A novel point of the proposed technique is the use of an evaluation model of perceptual pause insertion naturalness in addition to a prediction model based on machine learning. The evaluation model represents the relationship between several features related to pause insertion and the perceptual pause insertion naturalness obtained in a subjective evaluation. First, using a prediction model based on machine learning, we obtain the N-best sequences that indicate whether or not a pause is present at each phrase boundary. We then estimate pause insertion naturalness scores for each N-best sequence using the evaluation model and select the sequence with the highest naturalness score. Objective and subjective evaluation results show that the proposed technique gives better results than a conventional technique.

2.4.37 p.492

Sophie Herment, Nicolas Ballier, Elisabeth Delais-Roussarie, Anne Tortel,
Modelling interlanguage intonation: the case of questions
In this paper, we study the intonational patterns observed in learners' productions in order to evaluate what motivates the deviations observed: systemic differences between the learners' L1 and the L2, differences in phonetic implementation, etc. The analysis consists of a cross-comparison of the intonation of yes-no questions in French, English and English as an L2. It is based on five information-seeking yes-no questions that were extracted from the AixOx corpus, which contains a set of 40 texts that were read by 10 native French speakers, 10 Native English speakers and 20 French learners of English. The analysis of the data showed that the differences between native and non-native speakers do not affect

the form of the nuclear contour. It mostly shows that French speakers of English have a tendency to assign a rising pitch movement at the end of any prosodic words, which leads to a clear difference in rhythm.

2.4.38 p.563

Fabian Santiago, Paolo Mairano, Elisabeth Delais-Roussarie,
Non-native perception of final boundary tones in French interrogatives
In this paper, we report a perception experiment in which native and non-native listeners judged resynthesized questions varying in respect to two aspects: their morphosyntactic structure (presence/absence of an interrogative marker) and the form of their final tonal contour (falling, rising and extra rising). The goal of the experiment was to exanimate how non-native listeners of French did perceive the extra-rising final contour that was observed in learners' productions. Do they consider it as unmarked form during the acquisition process? By and large, the results of the experiment show that native listeners preferred rising contours over falling ones in all question types, whereas non-native listeners rated the extra rising contours higher than French natives in stimuli having a morphosyntactic structure that differs from the one used in their L1. These results suggest that rising contours represent a default tonal form associated with the interrogative modality at the beginning of the L2 acquisition process.

2.4.39 p.568

Katalin Mády, Ádám Szalontai,
Where do questions begin? – phrase-initial boundary tones in Hungarian polar questions
Hungarian prosody is left-headed, as suggested by the placement of the accent on the initial syllable on the level of prosodic words and the placement of the strongest pitch accent on the first accented word of the prosodic phrase. Earlier studies have pointed out that the left edge of the intonational phrase can bear a phrase-initial boundary tone that distinguishes between string-identical wh-interrogatives and wh-exclamatives. In this paper, two other string-identical sentence types, polar questions and declaratives, are investigated with respect to their prosodic features. Polar questions were characterised by a higher f0 maximum and a lower sentence-initial f0 than declaratives. The only pitch accent within the sentence was low, whereas declaratives had falling pitch accents. Sentence-final f0 and the pitch level of the accented syllable did not show a consistent pattern across speakers. It is concluded that low sentence-initial f0 together with the high tone on the penultimate syllable is a relevant marker of polar questions in Hungarian.

2.4.40 p.573

Xiaoming Jiang, Marc Pell,
Encoding and decoding confidence information in speech
This study aims to investigate the perceptual-acoustic correlates of vocal confidence. Statements with different communicative functions (e.g., stating facts, making judgments) were spoken in confident, close-to-confident, unconfident and neutral voices. Statements with preceding linguistic cues (e.g. I'm positive, Most likely, Maybe, etc.) or no linguistic cues were presented to sixty listeners in a perceptual study. The listeners were asked to judge whether statements conveyed some level of confidence, and if so, they were asked to evaluate the level of confidence of the speaker. The results demonstrated that the intended levels of confidence varied in a graded manner in the perceptual rating score; the more confident the statement intended to be, the higher the rating. In general, the neutral voice was judged to be more confident than the close-to-confident voice, but less than the confident voice. The presence of a linguistic cue tended to increase ratings of confident voices but decrease ratings of voices in the less confident voice conditions. To evaluate how specific prosodic cues are used to encode and decode confidence information, acoustic analyses were performed on the stimuli without the linguistic cue based on the mean perceptual rating of speaker confidence for each item. Results showed that statements rated as confident versus unconfident differed in the mean and the variance of fundamental frequency (f0) as well as speech rate, with confident statements exhibiting lower mean f0, smaller f0 variance, and faster speaking rate than unconfident statements. The perceived level of confidence was differentiated in the mean fundamental frequency in a parametric way, the lower the level of confidence, the higher the mean f0. Confident voices were also distinct from the other three conditions in terms of mean and range of amplitude (i.e., loudness). These findings shed light on how linguistic and paralinguistic cues reveal confidence-related information to listeners during speech.

2.4.41 p.577

Jane Kühn,
Aspects of Prosodic Phrasing in Turkish
This pilot study investigates the prosodic marking of contrastive in-situ focus in monolingual Turkish. The results of the production study are based on a phonological and phonetic analysis of information structure modified target sentences. The prosodic analyses reveal (i) feature that derive properties of prosodic phrasing which are inherent to phrase languages. It is shown that Turkish is a radical splitting language since each prosodic word (ω) forms its own phonological phrase ϕindicated by a high phrase tone (H-) aligned to ω- final syllables. The languages preference for radical splitting of simple SOV sentences is maintained in information structure modified targets by one speakers group, but modified by another group in favor of wrapping adjacent given constituents into one ϕ. The analyses reveal (ii) that prosodic cues are not crucial to mark in-situ focus in Turkish, but they may be used to contextualize information structure. If focused constituents are marked at all by prosodic means they do not show an increased pitch like most Germanic languages, but focused constituents are aligned to prosodic boundaries. The data motivate the claim that prosodic alignment is an adequate way to describe the prosodic realization of focus in Turkish.

2.4.42 p.582

Lehlohonolo Mohasi, Thomas Niesler, Hansjörg Mixdorff,
Perceptual evaluation of the effect of mismatched Fujisaki model commands and surface tone in Sesotho
Sesotho is a tonal Southern Bantu language which has so far received extremely little attention by the speech research community. We consider tone modelling for Sesotho using the Fujisaki model-based analysis with a view to the development of a text-to-speech (TTS) system. Fujisaki analysis can be used to indicate the tone associated with a syllable, but it often differs from the surface tone that would be available for TTS synthesis. We investigate instances in which the surface tone differs from the tone indicated by Fujisaki analysis, and determine the effect of these discrepancies on speech quality. The amplitude of Fujisaki tone commands is manipulated to match the surface tones, and the resulting resynthesized speech subsequently analysed by perceptual tests. We find that the effect of inserting tone commands at high surface tone syllables is more severe than matching the Fujisaki tone commands with low surface tone syllables, in terms of naturalness. Furthermore, some discrepancies can be attributed to errors in the surface tonal transcription. However, on average, all manipulations lead only to a mild degradation in speech quality. We conclude that the Fujisaki model is a feasible way to model tone in Sesotho even in the presence of limited and under-developed linguistic resources.

2.4.43 p.587

Suki Yiu,
Musical Intervals of Tones in Cantonese English
It has been shown that the relative pitch levels of Cantonese tones closely correspond to musical intervals (MIs). Given that an emerging tone language, Cantonese English, has developed tone under the substrate influence of Cantonese, this paper examines the correspondence between the newly emerged tones and MIs, and how the musical analogy relates to those established for Cantonese. The fundamental frequencies of the tones produced by six speakers of Cantonese English were extracted with Praat, then time-normalized across rhymes. The mean values of the interval points of two tones were expressed in terms of ratio, then matched with the closest MI on the musical scale. This paper demonstrates that the pitch levels of tones in Cantonese English correspond to MIs, given the converging ranges of MIs for different speakers and similar MIs of different tone pairs for different speakers. It also shows that the MIs of tones in Cantonese English are related to the corresponding tone pairs for Cantonese. The viability of MI as a means to understand the tonal system of non-tonal languages whose speakers' native language is tonal extends the link between the use of pitch in speech tones and music.

2.4.44 p.592

Wentao Gu, Keikichi Hirose,
Rhythmic Patterns in Native and Non-Native Mandarin Speech
Rhythm plays an important role in the naturalness of speech. This study compared rhythmic patterns

of Mandarin speech between native speakers and two groups of L2 speakers whose first languages were Cantonese and English, respectively. The study started from isolated words, but focused on continuous speech, for which eleven durational metrics were used as objective rhythm indicators. The results on continuous speech showed that nonnative Mandarin gave a quite similar rhythmic mode as native one in terms of rate-normalized/independent metrics, but shifted towards the stress-timed class in terms of raw metrics, regardless of the rhythmic class of the L1. This seems to conflict with the L1 transfer effect and the results for isolated words, but it coincides with auditory impression and can be explained by speech rate difference and the lengthening effects associated with the change in prosodic structure.

2.5 Wednesday Evening Session

2-5-evening - **Reviewers' Reception in Trinity College Long Room** (by invitation only)

3 Day Three

3.1 Thursday Session One

May 22nd, 9am - 10:30am : 3-1-plenary (1+3 presentations)

3.1.1 KeyNote 3 (p.598)

Jürgen Trouvain 30-min
Laughing, breathing, clicking the prosody of nonverbal vocalisations

When analysing human spoken communication the focus on the linguistic side lies on speech with its verbal message, whereas the focus on the non-linguistic side usually is on the visually transported information such as gestures and facial expression. However, speech, especially in talk-in-interaction, also features numerous nonverbal vocalisations including various forms of laughter and inhalation noises as their most frequent forms. Although nonverbal vocalisations are usually short in duration they may provide rich information on linguistic, paralinguistic and extralinguistic levels including prosodic phrasing, cognitive load, affective state or speaker identity. The talk provides an overview on the phonetic and prosodic structure and the timing of laughter and audible breathing. Special attention is put on conversational speech where we can frequently find situations in which interlocutors temporally overlap. An emphasis is given to apical click sounds that often occur with inhalation before upcoming articulation but also during word-finding difficulty.

3.1.2 p.603

Willemijn Heeren, Sarah Bibyk, Christine Gunlogson, Michael Tanenhaus,
Tuning in to whispered boundary tones
Very little is known about how listeners incorporate "intonational" information in whispered speech during online language processing. We present data showing that listeners can incorporate information about boundary tones in whispered speech rapidly, but this process is complicated by additional structural biases as well as by the fact that speakers do not produce cues to boundary tones consistently in whisper. Listeners, however, are able to adapt to these differences in order to correctly identify different boundary tones in whisper.

3.1.3 p.608

Oliver Niebuhr,
"A little more ironic" Voice quality and segmental reduction differences between sarcastic and neutral utterances
The presented production experiment analyzes the phonetic differences between neutral (i.e. sincere) and sarcastically ironic utterances in German. Results show in line with previous studies that sarcastic irony is expressed by longer utterance durations, lower and flatter F0 contours, and a lower intensity level. Moreover, extending previous findings, sarcastic irony is also characterized by a more variable (in tendency breathier) voice quality and a higher degree of segmental reduction, probably reflecting the speakers' dissociation from the wording of their utterances.

3.1.4 p.613

Amélie Rochet Capellan, Gérard Bailly, Susanne Fuchs,
Is breathing sensitive to the communication partner?
This paper investigates breathing profiles in eleven female speakers (subjects) when talking successively with the same two females (partners). Breathing kinematics of the two interlocutors was recorded synchronously by means of two Inductance Plethysmographs. In order to understand the implication of breathing in dialogue, we analyzed changes in breathing pauses according to the main dialogue events (listening, backchannels, turns start and turns continuation). Breathing and syllable rates were also compared among partners and subjects. The duration of inhalations and related pauses was reduced before a turn continuation in comparison to a turn start. The delay between speech offset in a breathing cycle and the onset of the next inhalation increased when a speaker and a listener swap roles as compared to a speaker who continued the turn. This was observed for both partners and subjects. The partners differed in their breathing and articulation rates but the two rates were not clearly correlated. In agreement with previous works, the current study shows that breathing kinematics is strongly linked to dialogue events. However, it doesn't show any clear effect of partner on speaker's breathing. This last result is discussed relative to methodological aspects.

3.2 Thursday Session Two - Poster

11am - 1pm : 3-2-poster (48 presentations)

- theoretical and linguistic prosody -

3.2.1 p.619

Carlos Gussenhoven, Lu Wang,
Yuhuan Wu tone and the role of sonorant onsets
Co-occurrence restrictions on tones and consonants in Yuhuan Wu Chinese syllables provide a powerful illustration of the phonetic basis of phonological contrasts, with sonorant contexts allowing more tone contrasts than other contexts. Interestingly, the language also reveals that the phonetic implementation of tones depends on the phonological contrast it is involved in. Such phonetic enhancement may be the opposite of what could be expected on the basis of speech ergonomics. Moreover, the language has two tone deletion rules that exempt tones in the context with the largest number of contrasts, showing a phonological version of enhancement.

3.2.2 p.623

Preethi Jyothi, Jennifer Cole, Mark Hasegawa-Johnson, Vandana Puri,
An Investigation of Prosody in Hindi Narrative Speech
This paper investigates how prosodic elements such as prominences and prosodic boundaries in Hindi are perceived. We approach this using data from three sources: (i) native speakers of Hindi without any linguistic expertise (ii) a linguistically trained expert in Hindi prosody and finally, (iii) automatic classifiers trained on English for prominence and boundary detection. We use speech from a corpus of Hindi narrative speech for our experiments. Our results indicate that non-expert transcribers do not have a consistent notion of prosodic prominences. However, they show considerable agreement regarding the placement of prosodic boundaries. Also, relative to the non-expert transcribers, there is higher agreement between the expert transcriber and the automatically derived labels for prominence (and prosodic boundaries); this suggests the possibility of using classifiers for automatic prediction of these prosodic events in Hindi.

3.2.3 p.628

Zenghui Liu, Aoju Chen, Hans Van de Velde,
Prosodic focus marking in Bai
This study investigates prosodic marking of focus in Bai, a Sino-Tibetan language spoken in the Southwest of China, by adopting a semi-spontaneous experimental approach. Our data have shown Bai speakers increase the duration of the focused constituent and reduce the duration of the post-focus constituent to encode focus. However, duration is not used in Bai to distinguish focus types differing in size and contrastivity. Further, pitch plays no role in signaling focus and differentiating focus types. The results thus

suggest that Bai uses prosody to mark focus, but to a lesser extent, compared to Mandarin Chinese, with which Bai has been in close contact for decades, and Cantonese, to which Bai is similar in the tonal system.

3.2.4 p.633

Maria Paola Bissiri, Margaret Zellers, Hongwei Ding,
Perception of Glottalization in Varying Pitch Contexts in Mandarin Chinese
Although glottalization has often been associated with low pitch, evidence from a number of sources supports the assertion that this association is not obligatory, and is likely to be language-specific. Following a previous study testing perception of glottalization by German, English, and Swedish listeners, the current research investigates the influence of pitch context on the perception of glottalization by native speakers of a tone language, Mandarin Chinese. Listeners heard AXB sets in which they were asked to match glottalized stimuli with pitch contours. We find that Mandarin listeners tend not to be influenced by the pitch context when judging the pitch of glottalized stretches of speech. These data lend support to the idea that the perception of glottalization varies in relation talinao language-specific prosodic structure.

3.2.5 p.638

Robert Bo Xu, Peggy Pik Ki Mok,
Cross-linguistic perception of Mandarin intonation
This study investigated how phonological knowledge and psychoacoustic mechanism interact in intonation perception. In the experiment, Mandarin and Cantonese listeners identified Mandarin statement and question in both unfiltered and low-pass filtered contexts. The results show that the importance of different perceptual factors varies depending on the perception materials. Language background plays an important role even in processing low-level psychoacoustic materials.

3.2.6 p.643

Wilbert Heeringa, Jörg Peters, Heike Schoormann,
Segmental and prosodic cues to vowel identification: The case of /I i i:/ and /U u u:/ in Saterland Frisian
Saterland Frisian has a complete set of closed short tense vowels. Together with the long tense vowels and the short lax vowels they constitute series of phonemes that differ by length and/or tenseness. We examined the cues that distinguish the front unrounded and the back rounded series of short lax and short and long tense vowels in triplets by eliciting 'normal speech' and 'clear speech' in a reading task from two speakers. Short and long vowels were distinguished by vowel duration, and lax and short vowels by their location in the F1-F2 space. The durational difference between short tense and long tense vowels, however, was largely restricted to the 'clear speech' condition. In 'clear speech', f0 excursion and centralization in the F1-F2 space were used as additional means to make short tense vowels more distinct from long tense vowels. These results suggest that length and tenseness are used as distinctive features, while f0 excursion and centralization in the F1-F2 space were optionally used to enhance the contrast between short and long tense vowels.

3.2.7 p.648

Marilisa Vitale, Philippe Boula de Mareüil, Anna De Meo,
An acoustic-perceptual approach to the prosody of Chinese and native speakers of Italian based on yes/no questions
The present study investigates the prosody of yes/no questions (in comparison with statements) in Chinese learners and native speakers of Italian. Acoustic analyses and a perceptual test were performed, in order to identify the main trends in non-native productions. Results show the relevance of prosody, which differentiates elementary, intermediate and advanced Chinese learners of Italian. Listening tests based on prosody transplantation also suggest that non-native segments with a native Italian prosody are rated as less accented than are native Italian segments with a non-native prosody. Similar trends were found, overall, in terms of question/assertion discrimination, confirming the relative importance of prosody. These findings could be helpful for teachers and learners of Italian as a foreign language.

3.2.8 p.653

Wei Lai, Ya Li, Hao Che, Shanfeng Liu, Jianhua Tao, Xiaoying Xu,
Final Lowering Effect in Questions and Statements of Chinese Mandarin Based on a Large-scale Natural Dialogue Corpus Analysis
To support text-to-speech with detailed prosody rules and generate natural prosody, the paper studied the pitch variation near the end of sentences based on a Chinese natural dialogue corpus. An additional lowering effect on the last prosodic word was found in both questions and statements, and proved to be independent of tone influence. Nevertheless, this effect, which is referred to as final lowering in other languages, was claimed to be absent in Chinese by some previous experimental studies. The cause of such a contradiction is very likely to be the difference between experimental speech vs. natural speech. Based on this observation, this paper proposed a combination of the two methods in intonation studies, in which experimental speech serves as an entry point to develop new topics, and natural speech serves as a necessary extension to revise and apply prosody rules.

3.2.9 p.658

Vered Silber-Varod, Tal Levy,
Intonation Unit Size in Spontaneous Hebrew: Gender and Channel Differences
In this corpus-driven research, the question of whether there is a tempo at the Intonation Unit (IU) level, and whether defined IUs differ not only with regard to their pitch contour and boundary tones but also with respect to their phonological size. For this reason, the inventory of syllable size (in terms of segments (phonemes)) and word size (in terms of syllables) was examined, and then each IU category (mainly Terminal vs. Continuous) was measured with respect to the number of syllables and words it contains. Moreover, terminal IU size was also measured with regard to the amount of embedded continuous IUs. Results showed that terminal IUs in spontaneous Israeli Hebrew (IH) do not necessarily consist of embedded continuous IUs. This can be explained due to their massive use as short feedback units in spontaneous speech. Statistical measurements for gender and channel (Face-to-Face vs. telephone conversations) variables were carried with no significance for gender, but with statistically significance for several channel aspects. Last, estimated durational measurements of the IU size are presented.

3.2.10 p.663

Allison Benner, John Esling,
Acoustic Cues to Tone and Register in Bai: Adult Baseline Data
This paper presents the results of a study of the acoustic cues associated with the tense/lax distinction in Bai, a Tibeto-Burman register tone language spoken in Yunnan, China. The purpose of the paper is to provide baseline adult data for comparison with infant speech in an acoustic study of infants' acquisition of Bai register tones in the second and third years of life. The results show that among adults, F0, F1, and spectral tilt combine to create the tense/lax contrast in Bai. While these three cues tend to be correlated, individual speakers differ in their use, particularly spectral tilt. The patterns in this study suggest that as Bai infants acquire tones in the second and third years of life, their utterances are likely to become structured around these three acoustic cues in previously unattested ways that exemplify the complex interaction between universal physiological and developmental tendencies and the ambient phonological tone system of Bai.

3.2.11 p.668

José Hualde, Tomas Riad,
Word accent and intonation in Baltic
We examine the realization of word accent contrasts in Standard Latvian and East Aukštaitian Lithuanian across intonational contexts. In our Latvian data the contrast is manifested as level vs. falling pitch in most contexts, in addition to a durational difference. In Aukštaitian Lithuanian, instead, differences in vowel quality and duration cue the lexical contrast in the nuclei that we examine. While Latvian retains a tonal contrast, in Aukštaitian Lithuanian it has been replaced with a combined segmental/quantitative contrast, where the so-called circumflex tone corresponds to relatively shorter duration and, in the case of diphthongs, centralized quality in the first half. We discuss the implications of these findings for further typological work.

3.2.12 p.673

Neville Ryant, Malcolm Slaney, Mark Liberman, Elizabeth Shriberg, Jiahong Yuan,
Highly Accurate Mandarin Tone Classification In The Absence of Pitch Information
A deep neural network (DNN) classifier based only on 40 mel-frequency cepstral coefficients (MFCCs) achieved 29.99% frame error rate (FER) and 16.86% segment error rate (SER) in recognizing five tonal categories in Mandarin Chinese broadcast news. With the addition of sub- band autocorrelation change detection (SACD) pitch-class features, the classifier scored 27.58% FER and 15.56% SER. These results are substantially better than the best previously reported results on broadcast-news tone classification, and are also better than a human listener achieved in categorizing test stimuli created by amplitude- and frequency-modulating complex tones to match the extracted F0 and amplitude parameters. The same DNN architecture scored substantially worse when trained and tested with SACD pitch-class parameters alone: 39.22% FER and 24.89% SER. RAPT F0 estimates are worse yet: 44.37% FER and 27.28% SER. The 40 MFCC parameters do not encode F0 in any obvious way and attempts to predict SACD or other pitch features from them work badly. These surprising results raise difficult questions for theories of Chinese tone.

3.2.13 p.678

Marzena Zygis, Daniel Pape, Luis Jesus, Marek Jaskula,
Intended intonation of statements and polar questions in Polish in whispered, semi-whispered and normal speech modes.
This paper provides acoustic correlates of intonation in whispered, semi-whispered and normal speech modes. In particular, it investigates correlates of utterance-final rising intonation in polar questions and falling intonation in statements. The paper does not only examine properties of vowels but also properties of the following voiceless consonant clusters. For the purpose of this study 2592 items produced by 16 native speakers of Polish were analysed. The results point to differences in spectral properties of both utterance-final vowels and consonants where falling intonation in statements contrasts with rising intonation in polar questions. Regarding the consonants, questions are produced with higher peaks, intensity, COG and STD values as well as smaller skewness and kurtosis values. Some spectral differences of consonants, including spectral slopes, are more distinguishable for questions versus statements in the whispered speech mode than in other speech modes. The more pronounced role of these cues in whispered speech suggests their compensatory function for the fundamental frequency, which is present in phonated speech. In summary, the study shows that speakers produce intended intonation patterns by varying the choice of cues as well as their magnitude in dependence on both (i) speech modes and (ii) intonation patterns.

3.2.14 p.683

Alina Lausecker, Annika Brehm, Ingo Feldhausen,
Intonational Aspects of Imperatives in Mexican Spanish
This paper sheds new light on the intonation of imperatives in Mexican Spanish. Results from a production experiment based on scripted speech show that imperative sentences have two different nuclear configurations depending on the position of the imperative verb (VI): (i) (L+)H* L% with VI in sentence-final position, and (ii) L* L% with VI in non-final position. The pitch accent on VI in non-final position is characterized by a late peak (L+>H*). However, if the sentence is uttered with some sort of emphasis, the nuclear configuration in the non-final context can also be rising. While these results partly confirm claims made concerning the nuclear configuration in De-la-Mota et al. (2010), they contradict the findings in Willis (2002), who attested strong pitch accent variation on VI.

3.2.15 p.688

Joanne Jingwen Li, Peggy P.K. Mok,
The acquisition of English lexical stress by Cantonese-English bilingual children at 2;06 and 3;0
This study investigates the acquisition of English lexical stress by Cantonese-English bilingual children at the age of 2;06 and 3;0 respectively, comparing them with the English monolingual peers. Research on early bilingual phonological acquisition often focuses on segmental level. Few studies are available when it concerns prosodic features, especially in children speaking non-Indo-European languages. This study

examines an important prosodic feature, lexical stress, in Cantonese-English bilingual children. The results showed that there is delayed acquisition of English lexical stress among the bilingual children, as reflected in less contrastive syllable duration and peak F0, possibly due to a lack of lexical stress in Cantonese, a typical syllable-timed language. This study helps to understand the bilingual interaction of two distinctive prosodic systems, and broaden our knowledge about early bilingual prosodic development.

3.2.16 p.693

Adrian Leemann, Volker Dellwo, Marie José Kolly, Stephan Schmid,
Disentangling sources of rhythmic variability between dialects
Speech rhythm is highly variable. Previous studies reported variability between languages, dialects, speakers, and labelers. Research further revealed an effect of sentence in the rhythmic characteristics of speakers of the same language. In the present study we tested whether the effect of sentence material is constant across varieties of the same language. We addressed this question by an example of analyzing rhythmic variability between eight dialects of Swiss German in three different sentences. Results showed a significant interaction for dialect*sentence for most of the tested rhythm metrics. We take this as evidence that differences between dialects are contingent upon the sentences used in the experiment. We further investigated which sources in the sentence material caused between-dialect differences in rhythm scores to vary. W e found exemplary evidence that dialect-specific phonological and morphological phenomena contained in the individual sentences are the prime suspects. Implications for future speech rhythm research are discussed.

3.2.17 p.698

Maria Del Mar Vanrell, Olga Fernández Soriano,
Dialectal variation at the Prosody-Syntax interface: Evidence from Catalan and Spanish interrogatives
In this study we investigate how prosody interacts with word order in the expression of interrogativity in different varieties of two Ibero-Romance languages, Catalan and Spanish. We analyze a corpus obtained by means of the Discourse Completion Task Methodology. The collected data were prosodically and syntactically annotated and show that the absence of syntactic marking (wh-word, subject-verb inversion or subject dislocation) for questions tends to correspond to a more salient intonational marking. Thus, wh-questions favor general falling intonational patterns. By contrast, yes-no questions can be classified depending on the nuclear tone (with preference for low tones in Catalan and high tones in Spanish) and final tone (low for language varieties with subject inversion or dislocation, but optionally high for those that do not present syntactic marking in a mandatory way.

3.2.18 p.703

Mathieu Avanzi, George Christodoulides, Elisabeth Delais-Roussarie,
Prosodic Phrasing of SVO Sentences in French
In the literature on prosody/syntax interface, syntactic information is usually considered as playing an important role in deriving the prosodic phrasing of an utterance. NP subjects, for instance, have often been claimed to phrase independently from the VP. It has nevertheless been shown that metrical factors could have an impact on phrasing, and that NPs could be phrased in the same prosodic phrase as the VP, or that the verb could be phrased with the subject. Several methods were used to measure metrical weight: number of syllables, of prosodic words, syntactic branchingness, etc. In order to determine which factors are more important, and how the all interact, we evaluate the weight that different metrical predictors have on prosodic phrasing. This is done by analyzing the phrasing of SVO structures in 200 sentences extracted from various French corpora. From the observation of the data that were semi-automatically annotated, it appears that subjects can be phrased independently or in the same PP as the VP, and that objects are rarely isolated from the verb. The analysis reveals interesting results regarding the effect of articulation rate and number of syllables, whereas syntactic-branchingness didn't show any effect.

3.2.19 p.708

Michelina Savino, Andrea Bosco, Martine Grice,
Intonational cues to item position in lists: evidence from a serial recall task
Intonation can convey information about how lists are structured into groups, as well as about specific item positions within a group. In Bari Italian, this function is expressed by three different tunes a) a rising contour, signalling that the list has not yet been completed; b) a high-rising contour, marking the penultimate item, i.e. signalling that the end of the list is approaching; c) a falling contour, marking the last item, i.e. cueing the end of the sequence. In this paper we explore the effects of such intonational information on working memory. In particular, we demonstrate that when listeners are requested to recall spoken nine-digit sequences by strictly following their serial order, their performance is significantly better when lists are characterised by tunes of the type described above, compared to sequences whose items are marked by a neutral, peak accent and/or are grouped by inserting a silent pause. We also observed that recall of items marked by specific contours at positions 3, 6 and 9 is particularly enhanced at these positions, whereas in sequences also containing intonational cues to items in penultimate position (2, 5 and 8) recall of those items is not equally improved. Therefore, it appears that in serial recall of spoken sequences, even when a large number of specific intonational cues to serial positions are available, listeners can make use of only a selection of them.

3.2.20 p.713

Anqi Yang, Aoju Chen,
Prosodic focus-marking in Chinese four- and eight-year-olds
This study investigates how Mandarin Chinese speaking children use prosody to distinguish focus from non-focus, and focus types differing in size of constituent and contrastivity. SVO sentences were elicited from four- and eight-year-olds in a game setting. Sentence-medial verbs were acoustically analysed for both duration and pitch range in different focus conditions. The children started to use duration to differentiate focus from non-focus at the age of four. But their use of pitch range varied with age and depended on non-focus conditions (pre- vs. post-focus) and the lexical tones of the verbs. Further, the children in both age groups used pitch range but not duration to differentiate narrow focus from broad focus, and they did not differentiate contrastive narrow focus from non-contrastive narrow focus using duration or pitch range. The results indicated that Chinese children acquire the prosodic means (duration and pitch range) of marking focus in stages, and their acquisition of these two means appear to be early, compared to children speaking an intonation language, for example, Dutch.

3.2.21 p.718

Simone Graetzer, Janet Fletcher, John Hajek,
Prosodic effects on vowel spectra in three Australian languages
In this paper, the spectral properties of vowels in three Australian languages are examined with the aim of determining whether prosodic prominence and domain-edge effects on formant frequencies, formant variability and vowel space dispersion can be identified. It is shown that these vowel systems are sufficiently dispersed, with an anchoring of the system by the open central vowel. It is also shown that for Burarra but not for Gupapuyngu or Warlpiri there is some evidence of prosodically-driven hyper-articulation. Finally, the data indicate pre-boundary lengthening in all three languages, which in some cases appears to be associated with changes in vowel quality.

3.2.22 p.723

Xi Chen, Peggy Pik Ki Mok,
Rhythmic Correspondence between Music and Speech in English Vocal Music
This study aims to investigate the rhythmic structures of music and speech, and to find out the possible corresponding rhythmic patterns between the two domains in English vocal music. With fifteen English songs as samples, lexical stress of multi-syllabic words is compared with three musical dimensionsmetrical stress, duration, and pitch respectively. It is found that in the chosen English songs, there is a good mapping between the metrical stress of music and the lexical stress of lyrics. In addition, the duration and the pitch patterns not only generally match lexical stress patterns most of the time, but also serve to manifest the prominence of the primary lexical stress on one hand, and to reflect the weakness of the unstressed syllables on the other. Except a general good match in rhythm, this study also shows

matching differences within the three comparisons. Matching degrees vary according to different meter patterns. Moreover, pitch takes priority over duration in their respective matching with lexical stress of the lyrics. Finally, the primarily stressed syllables match duration and pitch patterns much better than the unstressed ones do. Index Terms: Musical rhythm, Speech rhythm, English songs

3.2.23 p.728

Christoph Gabriel, Elena Kireva,
Speech rhythm and vowel raising in Bulgarian Judeo-Spanish
The study investigates selected prosodic characteristics of (Sofian) Bulgarian Judeo-Spanish, a diaspora variety of Spanish spoken by descendants of the Jews expelled from Spain, all of them bilingual speakers with Bulgarian as their dominant language. While exhibiting some few relics from Old Spanish on the segmental level, Judeo-Spanish shows a puzzling similarity with Bulgarian with respect to speech rhythm and vowel raising. It is shown that the two languages spoken by the bilinguals, Bulgarian and Judeo-Spanish, pattern alike in displaying almost the same rhythmic values (except for %V) and that raising of unstressed /a/ and /o/ as is typical of the variety of Bulgarian spoken in Sofia also regularly occurs in the Judeo-Spanish data. Our findings show that Judeo-Spanish is crucially influenced by Bulgarian, thus suggesting that it has largely converged toward the surrounding language on the phonological level.

3.2.24 p.733

Sandra Schwab, Carla V. Jara Murillo,
The role of stress perception in the assignment of written accent in Spanish
The aim of this investigation is to examine whether the adults' difficulty in placing the written accent in Spanish words is related to their ability in perceiving stress. The following variables were also taken into account in this study: the participant's education level (academic and non-academic), the stimulus lexical status (words and non-words), accentual pattern (proparoxytone, paroxytone and oxytone words) and length (2, 3 and 4 syllables). Participants performed a stress identification task and a word spelling task. Besides the effects of lexical status, education level and accentual pattern, results show an effect of the stress perception in the assignment of the written accent: stimuli with a correctly identified stress were more likely to be correctly written (i.e. with or without written accent) than the incorrectly perceived stimuli. This finding reinforces the idea that there is a relationship between prosodic and written skills.

3.2.25 p.738

Uwe Reichel, Alexandra Markó, Katalin Mády,
Parameterization and automatic labeling of Hungarian intonation
In Hungarian intonation research a common framework developed by Varga (2002) is to categorize the intonation within the domain of accent groups by character contours. We propose a linear parameterization of a subset of these contours derived from polynomial stylization. These parameters were used to train classification trees and support vector machines for contour prediction. Parameter extraction and training was carried out on the original F0 contours of spontaneous speech data as well as on three differently normalized variants suppressing fundamental frequency level and range effects. The highest accuracies were obtained for classification trees and F0 residuals after midline subtraction, but the overall performances were rather poor. Nevertheless, a significant improvement of the results was achieved by a Hidden Markov model to predict the correct label sequence from the partly erroneous classification output.

3.2.26 p.743

Maciej Karpinski, Katarzyna Klessa, Agnieszka Czoska,
Local and global convergence in the temporal domain in Polish task-oriented dialogue
Conversational parties tend to mutually adapt their communicative behaviour in a number of dimensions, from the level of physical aspects of speech signal and gesture, utterance properties, up to the level of mental representations. In the present study, an attempt is made to track the process of convergence in the temporal domain both as a global tendency and a local phenomenon. The material under study consists of two sets of task-oriented dialogues recorded with or without eye contact (telephone conversations) between the speakers. All the recordings were segmented into syllables and analysed in terms of speech rate and nPVI for each speaker as well as for the correlations between the speakers in each pair. Global convergence tendencies were proven to be weak but some influence of dialogue settings and gender was

found. The results seem to support the hypotheses that the alignment-related processes remain under the influence of many factors related to the dialogue flow and cannot be modelled as simply incremental.

3.2.27 p.748

Beatriz Raposo de Medeiros, Fred Cummins,
Speech and song synchronization: A comparative study
Does synchronization among speakers or singers require the presence of a beat? Is an implied underlying pulse or meter relevant? We set out to explore synchronization among speakers and singers as they speak or sing a variety of texts. We compare metrically strong nursery rhymes with non-metered prose. We compare singing in genres with two very different types of rhythm (samba and rock), and we compare sung and spoken versions of texts. In each case, we ask whether the rhythmic qualities of the texts facilitate synchronization. The metrical structure of the nursery rhyme does not facilitate synchronization compared to prose, while the simple beat of rock music does help. Further comparisons are provided in the text.

3.2.28 p.752

Katalin Mády, Uwe D. Reichel, Štefan Beňuš,
Accentual phrases in Slovak and Hungarian
Languages with primarily delimitative function of word stress commonly make use of accentual phrases (APs) in their intonational phonology (e.g. Tamil or French). Slovak and Hungarian are genetically unrelated but geographically close languages with word-initial lexical stress. In this paper we compared the stylised f0 of single accent groups (AGs) with the f0 level pattern of the entire intonational phrase (IP) to test if AGs are relevant for the intonational phonology of Slovak and Hungarian. Steep f0 slopes with a recurring pattern (rising or falling) and large deviations from IP level patterns were interpreted as evidence for the autonomy of the AG in the given language. The results suggest that Hungarian is indeed a language in which accent groups form a unit on their own, however, such evidence was not found for Slovak.

3.2.29 p.757

Nicole Dehé,
Final devoicing of /l/ in Reykjavík Icelandic
Icelandic has a phonological process which devoices sonorants after voiced segments in domain-final position, but to date the category of the relevant domain and potential further factors affecting it have not been identified. The present paper studies final devoicing of /l/, by which /l/ is realized as the voiceless lateral fricative [ɬ] in domain-final position. It reports on the results of an experimental reading study designed to test the exact environments of this process and the implications for a prosodic hierarchy for Icelandic. The results suggest that devoicing of /l/ is bound by the prosodic utterance. All instances of /l/ were devoiced in utterance final position. Within the utterance, final devoicing is optional, but the frequency of its application reflects the syntactic and prosodic hierarchy such that it is most frequent at a clause/an IP-boundary, significantly less frequent at a syntactic XP-edge and it almost never occurs within a syntactic XP.

3.2.30 p.762

Scott Lee,
The Realiziation of French Rising Intonation by Native Speakers of American English
This study examines the acquisition of French intonational rises by adult native speakers of American English. Production data were gathered using a discourse completion task and a storytelling task from eight American college students beginning a semester-long study abroad program in Southern France. Results suggest that speakers struggled with two particular aspects of French intonation: the grouping of words into Accentual Phrases, and the phonetic realization of phrase-final rises. In particular, the probability distribution for the alignment of the late L elbow was bimodal for L2 speakers but unimodal for L1 speakers, suggesting the use in the learner speech of two distinct tonal patterns instead of the single French LH*. Mean values for overall pitch range and the scaling of continuative rises were significantly

lower and less variable than French L1 values as well.

3.2.31 p.767

Niamh Kelly, Rajka Smiljanic,
Monosyllabic Lexical Pitch Contrasts in Norwegian
This paper examines the lexical tonal accent contrast in monosyllabic words in the Trondersk dialect of Norwegian. The results of a production experiment in which speakers produced the unmarked accent and the circumflex accent showed that the tonal distinction is characterized by a difference in f0 maximum, f0 height at onset, f0 minimum and its timing, and height of the final Accent Phrase H tone. The presence of the tonal accent contrast on monosyllabic words is unusual among dialects of Norwegian and Swedish.

3.2.32 p.772

Yu Lun Hsieh, Ching-Ting Chuang, Feng Fan Hsieh, Yueh Chin Chang, Wen Lian Hsu,
Taiwanese Tone Recognition Using Fractionalized Curve-fitting of Prosodic Features
In this paper, we examined different methods of modeling prosodic features of tones, and their effects on a speaker-independent Taiwanese tone recognition system. Tones can be modeled either by plain or curve-fitted features. Plain features represent the original curve faithfully using pitch values, while curve-fitted features can be thought of as an approximation to the values using mathematical functions, such as a Legendre polynomial. In addition, durational information of tones was also proven effective in previous researches. Thus, we proposed a new approach of modeling Taiwanese tones using curve-fitted features extracted from fractions of the pitch curve, along with duration as an additional prosodic feature. Our experimental results showed that using these features in an SVM classifier could substantially improve the accuracy of tone recognition in Taiwanese. Besides, we provided an empirical perspective for theoretic studies on tonal neutralization.

3.2.33 p.776

Bistra Andreeva, Grazyna Demenko, Magdalena Wolska, Bernd Möbius, Frank Zimmerer, Jeanin Jügler, Magdalena Oleskowicz- Popiel, Jürgen Trouvain,
Comparison of Pitch Range and Pitch Variation in Slavic and Germanic Languages
This study presents the results of a large-scale comparison of various measures of pitch range and pitch variation in two Slavic (Bulgarian and Polish) and two Germanic(German and British English) languages. The productions of twenty two speakers per language (eleven male and eleven female) in two different tasks (read passages and number sets) are compared. Significant differences between the language groups have been found: German and English speakers use lower pitch maxima, narrower pitch span and generally less variable pitch than Bulgarian and Polish speakers. These findings support the hypothesis that particular linguistic communities tend to be characterized by particular pitch profiles.

3.2.34 p.781

Philippe Martin,
Silent reading and prosodic structure constraints
Silent reading of written texts involves necessarily a process of subvocalization, i.e. the presence of a voice reading the text in the head of the reader speaking to her/himself. This process includes not only the sequences of syllables corresponding to the written material, but also sentence intonation. Since subvocalization cannot be eliminated other than by changing the status of each word into a pictographic function (as it may be the case for a STOP road panel sign), it is argued here that sentence intonation is essential to language comprehension, and more specifically to the conversion of sequences of syllables into higher order linguistic units (corresponding to accent phrases AP in the Autosegmental-Metrical model). Consequently, reading and in particular silent reading is constrained by the same rules than the prosodic structure in general, and specifically to the minimal duration of accent phrases. This minimal value, occurring when AP's contain only one syllable, is about 250 ms, a value which corresponds to the minimal period value of Delta brain waves. Therefore this AP minimal duration limits also the maximal number of AP that could be processed in silent reading, i.e. about 240 per minute, which corresponds to the maximal number of words per minutes experts in fast reading can process while keeping a reasonable

level of comprehension, i.e. about 800 wpm.

3.2.35 p.785

Emma Valtersson, Francisco Torreira,
Rising intonation in spontaneous French: how well can continuation statements and polar questions be distinguished?
This study investigates whether a clear distinction can be made between the prosody of continuation statements and polar questions in conversational French, which are both typically produced with final rising intonation. We show that the two utterance types can be distinguished over chance level by several pitch, duration, and intensity cues. However, given the substantial amount of phonetic overlap and the nature of the observed differences between the two utterance types (i.e. overall F0 scaling, final intensity drop and degree of final lengthening), we propose that variability in the phonetic detail of intonation rises in French is due to the effects of interactional factors (e.g. turn-taking context, type of speech act) rather than to the existence of two distinct rising intonation contour types in this language.

3.2.36 p.790

Ludger Paschen,
Intonation and focus marking in Ulyap Kabardian
This paper presents a pilot study that aims at establishing a model for the intonation of Ulyap Kabardian in the ToBI framework. On the basis of data gathered during a field trip in 2012, it is suggested that four/three pitch accents and three boundary tones are needed to describe intonation in four communicative contexts. Additionally, it is shown that for focus marking in Ulyap Kabardian questions, a stress shifting rule dislocates word stress to a prosodically determined position. This shift rule is extraordinary in that it is insensitive to stress clashes. From a cross-linguistic perspective, the intonation system of Ulyap Kabardian bears a higher resemblance to the system of one of the Kabardian dialects spoken in Turkey than to Russian, the principal contact language.

3.2.37 p.795

Oliver Jokisch, Tristan Langenberg, Gabor Pinter,
Intonation-Based Classification of Language Proficiency Using FDA
State-of-the-art pronunciation tutoring (CAPT) systems are based on ASR technology. Consequently, they can provide a distinguished learning feedback which is focused on phonetic features and the positions of articulation errors. In contrast with the relative success with segmental errors, the acquisition and assessment of second language (L2) prosody is still a challenging problem. Although prosodic parameters like f0 contour or duration measures are usually displayed, the consequential evaluation components are generally missing. Considering the strong variation in speech data, functional data analysis (FDA) is a useful concept which statistically analyses interrelations between principal components (e.g., given accentuation) and their contribution to superimposed forms (e.g., resulting f0 contour). This article describes baseline processing and preliminary results of a pilot study on the intonation-based proficiency classification of German by using FDA methods. The experimental part contains the FDA-based classification results compared to a perceptual classification by German natives.

3.2.38 p.800

Kieu Phuong Ha, Martine Grice, Marc Brunelle,
Tonal allophony in Vietnamese: Evidence from task-oriented dialogues
In this paper we investigate the behaviour of the lexical rising tone (SAC) in disyllabic sequences in the Northern variety of Vietnamese. Results from task-oriented dialogues show that this rising tone (SAC), when occurring before the lexical high-level tone (NGANG), can be realised as low level or falling, resembling a different tone in the language (HUYEN). This is the case word-internally and within noun phrases. Two further observations give us an indication that a sandhi process could be developing: (a) this variation is not found in sequences across a larger juncture, and (b) the SAC tone does not undergo this change before other tones.

3.2.39 p.804

Nina Grønnum,
Laryngealization or Pitch Accent - the Case of Danish Stød
According to recent proposals Danish stød is the phonetic manifestation of a HL tonal pattern compressed within one syllable, making the stød/non-stød distinction a special case of the more general tonal word accent distinction in Swedish and Norwegian. This review of the relevant aspects of Danish stød and intonation demonstrates that (1) such a tonal representation of stød is contradicted by the phonetic reality. (2) Stød is distributed in words according to roughly the same principles across regional varieties of Danish, but tonal patterns are highly variable. (3) Word accents in Swedish and Norwegian are associated exclusively with stressed syllables, whereas stød occurs also in less than fully stressed syllables, devoid of autonomous pitch movements. (4) A word in Swedish and Norwegian can have one pitch accent only, but Danish words may have more than one stød.

3.2.40 p.809

Ann Bailey,
Intonational Phonology of Cuban Spanish: A Preliminary AM Model
The present study proposes a preliminary model of intonational phonology for Cuban Spanish in the framework of Autosegmental-Metrical phonology. Data from controlled and semi-spontaneous speech were used to establish the boundary tones and pitch accents which are contrastive in this variety of Spanish. It was found that Cuban Spanish shares various tonal categories with both the Pan Spanish ToBI (Tones and Break Indices) and other Caribbean Island Spanish dialects (Puerto Rican and Dominican), but differ from these dialects in how those pitch accents and boundary tones are used to convey meaning. Cuban Spanish shares its primary prenuclear pitch accents and nuclear contours for imperative statement and narrow focus with the Pan Sp_ToBI, but shares the nuclear contours for broad focus, vocative, and wh-questions with Puerto Rican Spanish. Similar to the other Caribbean Island Spanish varieties, the Cuban Spanish boundary tone inventory consists of a subset of the attested boundary tones found in the Pan Sp_ToBI, and all three Caribbean varieties share low boundary tones in non-wh questions, a marker of Caribbean Spanish speech.

3.2.41 p.814

Roberto Paternostro, Jean Philippe Goldman,
Modeling of a rise-fall intonation pattern in the language of young Paris speakers
Intonation seems to be one of the major cues for identifying youth language in the Paris region. As part of a large-scale corpus-based analysis, this paper attempts to model a high-low final prosodic pattern, considered to be representative of a Paris working-class suburbs accent. Comparison with the emphatic high-low prosodic pattern, well-known in general French, will provide the opportunity for sociolinguistic insights. The ethnic hypothesis is dismissed in favor of a context-bound and interaction-sensitive interpretation.

3.2.42 p.819

David Le Gac,
Topic and Focus Intonation in Argentinean Porteño
This paper investigates the intonation of topics and focus in Argentinean Porteño. We have found that whereas tonal alignment is phonetically conditioned, pitch height and duration constitute the main cues to express various types of focus in declarative and interrogative sentences; at least four intonational categories seem to be used by our speakers and the relevance of a register feature is discussed. As for topics, they are marked by special tunes that depend on the type of sentence; in particular, the topic tune in questions is the opposite of those found in declaratives.

3.2.43 p.824

Liang Zhang, Yuan Jia, Aijun Li,
Analysis of Prosodic and Rhetorical Structural Influence on Pause Duration in Chinese Reading Texts
This paper investigates factors that influence pause duration in Chinese reading texts through examining

the stress degree in pre-pausal and post-pausal positions and the rhetorical structure in discourse as a whole. The RSTTool is used in diagramming the rhetorical structures of the texts. The recordings, extracted from the ASCCD corpus, are further analyzed acoustically and statistically by applying Praat and R. The statistical analysis results show that the stress degree in both pre- and post-pausal positions has a significant impact on pause duration. Moreover, the nuclearity in both positions have also been shown to have a remarkable influence. Specifically, the nucleus in pre-pausal and satellite in post-pausal positions can significantly lengthen the pause duration.

3.2.44 p.829

Irina Nesterenko,
Statistical and temporal properties of prosodic phrasing in French conversational speech
Our study investigates prosodic phrasing in a corpus of French conversational speech. We looked at statistical and temporal properties of prosodic constituents, which were previously identified within laboratory phonology paradigm. Prosodic annotation of our corpus implements two-level hierarchical model distinguishing major prosodic units (Intonational Phrases, IPs) and minor prosodic units (Accentual Phrases, AP). Both temporal data and distribution of the number of APs in an IP evidence the global tendency to produce shorter units in conversation. Moreover, Intonational phrases containing no more than two Accentual phrases cover 80% of the data. We discuss the implication of these results for both phonological studies of the constraints on prosodic phrasing and oral document tagging.

3.2.45 p.833

Oyedeji Musiliyu, Miguel Oliveira,
Intonational Patterns of Telephone Numbers In Brazilian Portuguese
The main purpose of this study was to identify intonational patterns of a quite common type of numeric grouping in Brazilian Portuguese: the one associated with telephone numbers. To this aim, 30 samples of spoken telephone numbers, read aloud by 85 native speakers of Brazilian Portuguese were analysed. The description of their intonation contour was observed by using Momel/Intsint [1] and ProsodyPro [2] scripts for Praat (version 5.3.53) [3], through a semi-automatic analysis of pitch variations in numeric groupings that form the telephone numbers. The results show a pattern of intonation and numeric grouping strategy that are sufficient enough to prosodically characterize different types of spoken telephone numbers in Brazilian Portuguese.

3.2.46 p.838

Simone Falk, Elena Maslow,
Song and speech prosody influences VOT in stuttering and non-stuttering adolescents
Since a long time, it is known that singing helps persons who stutter to produce their utterances more fluently. The prosodic characteristics of spoken and sung utterances differ considerably in their rhythmic and tonal structure. Therefore, it has been proposed that song prosody helps stutterers to improve their rhythmic planning of verbal material [1]. In order to investigate this idea, we examined temporal aspects, namely Voice Onset Time (henceforth, VOT) of voiceless plosives, in sung and spoken utterances of young German stutterers and non-stuttering controls. VOT tends to be reduced in song compared to speech. We expected a more important reduction in the stuttering group as voice onset timing should be facilitated in song compared to speech. Eight stuttering adolescents and eight normal fluent peers read and sang an altered version of Happy Birthday with test words containing the three voiceless stops /p/, /t/, /k/. Results showed that stuttering as well as non-stuttering adolescents reduced VOT during singing compared to speech. In contrast, only adolescents who stutter were less variable in their VOT production in song compared to speech. Additional analyses indicated further group differences in vowel duration following the stop consonant. These findings suggest that young stutterers benefit from sung prosody in their timing abilities.

3.2.47 p.843

Stefanie Jannedy, Melanie Weirich,
Some aspects on individual speaking style features in Hood German
Multiethnic urban German (Hood German) as spoken by adolescents in Berlin differs in several significant ways from more standard varieties of Berlin German. It is characterized by a variety of morpho-syntactic

alternations and phonetic variants uncommon to the regional standard spoken in Berlin. Previous quantitative corpus analyses have shown that overall speakers of the multiethnic youth style German have a strong tendency to centralize /ɔɪ/ compared to speakers rendering the local regional standard. This paper now summarizes this centralization tendency and investigates auditory salient realizations of variation by individuals which show tendencies towards a hiatus in the diphthong /ɔɪ/, breaking the nucleus and the off glide. Moreover, there are other prosodic and segmental co-occurring features in the speech of some adolescents which are displayed since it is suspected that some of these may be(come) markers of Hood German.

3.2.48 p.848

Katarina Bartkova, Mathilde Dargnat,
Automatic extraction of prosodic patterns Cross linguistic study on laboratory data
The goal of our study is to use an automatic approach to extract the general prosodic tendency of the speech signal conveyed by the F0 pattern and the syllable duration. The speech signal is prosodically annotated by an automatic prosodic transcriber and then prosodic patterns are extracted from this annotation. The pertinence of the pattern extraction is tested here on laboratory data containing isolated sentences in French and English uttered by native and non-native speakers. An analysis of the extracted parameters allows observing how the prosody of the sentences is defined by their shared syntactic structures and to what extend are the prosodic features used by the two languages different or similar. It appears from the analyzed data that such an automatic prosodic parameter processing can yield relevant information for a cross-linguistic study of the prosody.

3.3 Thursday Session Three - Panel: Terminology in Prosody Research

2pm - 2:30pm : *Tribute to Sugito Miyoko sensei*

2:30pm - 3:30pm : Panel discussion: **Terminology in Prosody Research**
Organisers: Hiroya Fujisaki and Nick Campbell

1. Brief introduction - Hiroya Fujisaki
   Necessity of discussion and recommendation on common terminology in prosody research
2. Issues of teminology in several academic domains related to prosody research
   (a) acoustic/physical/physiological/psychological terms (Fujisaki)
   (b) linguistic/phonetic terms (Hirst)
   (c) semantic/pragmatic terms (Gibbon)
   (d) prosodic terms (Campbell)
3. lively discussion and debate from the floor

3.4 Thursday Session Four

4pm - 6pm : 3-4-oral (6 presentations)

- perception and production -

3.4.1 p.854

Sun-Ah Jun, Jason Bishop,
Implicit prosodic priming and autistic traits in relative clause attachment
Using the structural priming paradigm, the present study explores predictions made by the Implicit Prosody Hypothesis (Fodor 1998) by testing whether an implicit prosodic boundary generated from a silently read sentence influences attachment preference for a novel, subsequently read sentence. Results indicate that such priming does occur, although the patterns are highly dependent on individual differences in listeners' "autistic" traits.

3.4.2 p.859

Jennifer Cole, Tim Mahrt, José I. Hualde,
Listening for sound, listening for meaning: Task effects on prosodic transcription
The perception of prosodic structure (phrasal prominences and boundaries) may depend in part on acoustic information present in the signal and in part on meaning based on syntactic, semantic and pragmatic factors. Listeners may also be able to weigh acoustics and meaning to different degrees. We test naïve subjects' marking of prominences and boundaries in spontaneous American English under three different conditions, all of which involve listening to audio recordings and marking prominences and boundaries on a transcript. The three conditions differ in the instructions that transcribers were given. In one condition, subjects were instructed to transcribe prominence and boundaries based on meaning criteria, in a second condition they were told to transcribe based on criteria of acoustic salience. A third condition had more general instructions, without explicit reference to either meaning or acoustic perception. Our results show that subjects perform differently when focusing on meaning and on acoustics, especially for prominence marking, where many different words are selected as prominent under the two tasks. Boundary marking is more similar under the two instructions, with acoustic criteria resulting in a higher frequency of boundaries, but with boundaries marked largely on the same words in both tasks. When given non-specific instructions, performance was much more similar to that obtained under acoustic-based instructions. We report on agreement rates within and across conditions. This study has implications for models of prosody perception and the methodology of prosodic transcription.

3.4.3 p.864

Florian Hönig, Anton Batliner, Elmar Nöth, Sebastian Schnieder, Jarek Krajewski,
Acoustic-Prosodic Characteristics of Sleepy Speech - Between Performance and Interpretation
When we address speaker states like sleepiness, two partly competing interests can be observed: within both applications and engineering approaches, we aim at utmost performance in terms of classification or regression accuracy - which normally means using a very large feature vector and a brute forcing approach. The other interest is interpretation: we want to know what tells apart atypical (here: sleepy) speech from typical (here: non-sleepy) speech, i.e., their respective feature characteristics. Both interests cannot be served at the same time. In this paper, we preselect a small number of easily interpretable acoustic-prosodic features modelling spectrum and prosody, based on the literature and on the general idea of sleepiness being characterised by relaxation. Performance obtained with these single features and this small feature vector is compared with the performance obtained with a very large feature vector; moreover, we discuss to which extent the features chosen model relaxation as sleepiness characteristic.

3.4.4 p.869

Juraj Šimko, Štefan Beňuš, Martti Vainio,
Hyperarticulation in Lombard speech: A preliminary study
Over the last century researchers collected a considerable amount of data reflecting properties of the Lombard speech, i.e., speech in a loud environment. The documented phenomena include effects on intensity, fundamental frequency, spectral tilt, speech rate and articulationm. Relatively little attention has been paid to the effects on relative extent of movement of individual articulators. In an attempt to fill in this gap we present a preliminary analysis of EMA data collected in increasing levels of babble noise. We introduce HH-index as a measure of overall relative activity of articulators. Our results indicate a non-linearity of the effect of noise on articulatory movement and quantitatively different effects on the movement extent for different groups of articulators. The effects of noise are compared with those brought out by other techniques for eliciting articulatory variation. We also discuss possible application of Lombard speech as an elicitation paradigm for studies of hyperarticulation.

3.4.5 p.874

Susanne Schötz, Joost van de Weijer,
A Study of Human Perception of Intonation in Domestic Cat Meows
This study examined human listeners' ability to classify domestic cat vocalisations (meows) recorded in two different contexts; during feeding time (food related meows) and while waiting to visit a veterinarian

(vet related meows). A pitch analysis showed a tendency for food related meows to have rising F0 contours, while vet related meows tended to have more falling F0 contours. 30 listeners judged twelve meows (six of each context) in a perception test. Classification accuracy was significantly above chance, and listeners who had reported previous experience with cats performed significantly better than inexperienced listeners. Moreover, the two food related meows with the highest classification accuracy showed clear rising F0 contours, while clear falling F0 contours characterised the two vet related meows that received the highest classification accuracy. Listeners also reported that some meows were very easy to classify, while others were more difficult. Taken together, these results suggest that cats may use different intonation patterns in their vocal interaction with humans, and that humans are able to identify the vocalisations based on intonation.

3.4.6 p.879

Chunyue Zhu, Toshiyuki Sadanobu,
Observation of so-called "pursed-lip" and "curled-lip" utterances in Japanese, using video and MRI images
The Japanese language includes utterances described by the idioms "speaking with pursed lips" and "speaking with curled lips." This study employs video and MRI imaging to examine the articulatory characteristics of these utterances ("utterances P" and "utterances C", respectively) by comparing their articulation with that of "unmarked" utterances ("utterances U"). Through doing so, we arrive at the following four conclusions: (1) For the articulation of utterance P, the lips are projected outward, and rounded by expanding in the vertical direction and narrowing in the horizontal direction. (2) For the articulation of utterance C, curling the lips is not an absolute requirement. The articulation of utterance C is similar with that of utterance P in that the lips are projected outward and rounded. (3) Utterances P and C differ in two points: (a) Lips projection accompanies the lower jaw projection only in utterances P; (b) Lips in utterance P is wider than those in utterance C. (4) The shapes the lips make in utterance P, utterance C, and utterance U can be described as a circle, a horizontal rectangle, and a horizontal oval, respectively. (5) There are many facts that contradict the accepted theory that "Rounding the lips causes both lips to project outward. In reaction to this movement, the surface of the tongue is pushed toward the rear" (Koizumi 1989).

3.5 Banquet - May 22nd

7pm - 9:30pm : banquet - **Trinity College Old Dining Hall** - ALL WELCOME!

4 Day Four - May 23rd

4.1 Friday Session One

9am - 10:30am : 4-1-poster (48 presentations)

- intonation and speaking style -

4.1.1 p.885

Page Piccinini, Marc Garellek,
Prosodic Cues to Monolingual versus Code-switching Sentences in English and Spanish
Code-switching offers an interesting methodology to examine what happens when two linguistic systems come into contact. In the present study two experiments were conducted to see if (1) listeners are able to anticipate code-switches in speech-in-noise, and (2) prosodic cues are present in the signal to warn of an upcoming code-switch. A speech-in-noise perception experiment with early Spanish-English bilinguals found that listeners are able to accurately identify words in code-switching sentences with the same accuracy as in monolingual sentences, even in highly degraded listening conditions. We then analyzed the stimuli used in the perception experiment, and found that the speaker does use different prosodic contours for code-switching productions as compared to monolingual productions. We propose that listeners use these code-switching specific prosodic contours to anticipate code-switches, and thus ease processing costs in word identification.

4.1.2 p.890

Simon Ritter, Timo B. Roettger,
Speakers modulate noise-induced pitch according to intonational context
Recent studies have shown that speakers systematically modulate properties of voiceless segments according to intonational context. More specifically, in the absence of fundamental frequency (F0), speakers appear to adjust the Center of Gravity (CoG) and the intensity of voiceless fricatives to convey the impression of pitch. In line with these findings, the present production study extends earlier work and investigates noise-induced properties of fricatives, modulated by the intonation context. It is shown for German that the mean CoG and intensity of intended contours with a high boundary tone are higher than those produced for intended contours with a low boundary tone. Furthermore, looking at the development of CoG and intensity over the time course of the fricative, the trajectories corresponding to the boundary tones differ in intercept and slope, i.e. reveal a steeper fall in case of a corresponding falling tone.

4.1.3 p.895

Albert Rilliard, Donna Erickson, Takaaki Shochi, Joõ Moraes,
US English attitudinal prosody performances in L1 and L2 speakers
Expressive behavior linked to paralinguistic meanings finds grounds in codes proposed as universals, as well as in culture-specific conventions. This study observes performances in such kinds of attitudinal prosody for USA English, produced by L1 and L2 speakers. The results show that the observed variance is linked to individual competence, to the linguistic context, and to the cultural background of the speakers. They also show that the code used to express a given speech act, code learned in the L1 language by L2 speakers of English, may be used in their L2 language. For some of these expressions, L2 speakers received higher scores than L1 speakers, suggesting that expressions conventionalized in a foreign language, are adequately fulfilling not-conventionalized expressions in the L1 culture.

4.1.4 p.900

András Beke, György Szaszák, Viola Váradi,
An Automatic Hierarchical Multiple Level Phrase segmentation approach for Spontaneous speech
The present paper investigates automatic prosodic phrasing of spontaneous speech: a two-step segmentation technique is presented, based on unsupervised learning. In the first step, the Intonational Phrases (IP) are detected automatically based on speech energy, spectral centroid and a double-thresholding technique. In the second step, Phonological Phrases (PP) are identified within the IPs. As acoustic features, F0, overall energy and vowel duration are investigated. An adaptive thresholding method is used based on Kullback-Leibler divergence computed in an autocorrelative manner for the feature streams. For Hungarian spontaneous speech, a phrasing accuracy of over 80% can be reached when comparing to a hand-labelled reference phrasing. It is found that in Hungarian sponatenous speech, F0 and energy play an essential role in IP level phrasing, whereas PP level phrasing is most effective using F0 related features alone. Vowel durations are shown not to contribute to prosodic phrasing in Hungarian. Although the evaluation targets the Hungarian language, the applied method is universal and can be easily adapted for other languages. Index Terms: speech synthesis, unit selection, joint costs.

4.1.5 p.905

Katelyn Eng, Beverly Hannah, Keith Leung, Yue Wang,
Effects of auditory, visual and gestural input on the perceptual learning of tones
Research has shown that audio-visual speech information facilitates second language (L2) speech learning, yet multiple input modalities including co-speech gestures show mixed results. While L2 learners may benefit from additional channels of input for processing challenging L2 sounds, multiple resources may also be inhibitory if learners experience excessive cognitive load. The present study examines the use of metaphoric hand gestures in training English perceivers to identify Mandarin tones. Native Mandarin speakers produced tonal stimuli with simultaneous hand gestures mimicking pitch contours in space. The English participants were trained to identify Mandarin tones in one of four modalities: audio-only, (AO), audio-visual (AV, speaker voice and face), audio-gesture (AG, speaker voice and hand gestures) and audio-visual-gesture (AVG). Results show significant improvements in tone identification from pre-

to post-training tests across all four training groups, demonstrating that gestural as well as visual articulatory information may facilitate tone perception. However, further analyses with individual tones reveal some group differences. Most noticeably, the AVG group had a slower learning curve during training compared to the other trainee groups for Tone 4, the least accurately identified tone, indicating a negative effect of multiple input modalities on the perception of difficult L2 sounds. In contrast, for Tones 2 and 3, the AG group revealed slower learning effects compared to the AV group, presumably because of the similar gestural trajectories for these two tones, which made the gestural input less distinct. Overall, the results suggest a positive role of gestures in tone identification, one that may also be constrained by phonetic and cognitive demands.

4.1.6 p.910

Céline De Looze, Daniel Hirst,
The OMe (Octave-Median) scale: a natural scale for speech melody.
Fundamental frequency, the primary acoustic correlate of speech melody, is generally analysed and displayed using a linear scale (Hertz) or a logarithmic one, generally in semitones and usually offset to an arbitrary reference level such as 100 Hz. In this paper we argue that a more natural scale for analysing speech is the OME (Octave-MEdian) scale, using the octave (o) as the basic unit, offset to the median value of the speaker's range. We present results showing that a reasonable estimate of a speaker's pitch range can be obtained directly from the median.

4.1.7 p.915

Nigel Ward,
Automatic Discovery of Simply-Composable Prosodic Element
As a way to discover the elements of prosody, Principal Component Analysis was applied to several dozen contextual prosodic features computed at 600,000 timepoints in dialog data. The results suggest that English has at least several dozen prosodic patterns, each with its own communicative function.

4.1.8 p.920

Jan Volín, Eliška Churaňová, Pavel Šturm,
P-centre Position in Natural Two-Syllable Czech Words
The ability to lock motor activity oscillator with external acoustic events is typical of various forms of human behaviour. Previous research showed that the beginning of an action is not necessarily the beginning of the rhythmic phase and led to the concept of p-centres. We present an experiment with 18 natural two-syllable Czech words spoken in synchrony with metronome beats by 18 subjects. Complexity of the consonantal onset and the type of coda together with distinctive phonological vowel length were carefully controlled to reveal a complex but comprehensible relationship between the word structure and phase locking.

4.1.9 p.925

Hyun Kyung Hwang, Satoshi Ito,
Correlation between prosody and epistemic bias of negative polar interrogatives in Japanese
The study investigates the correlation between prosodic patterns and epistemic bias observed in Japanese negative polar interrogatives, with special attention given to the perceptual and functional aspects of the correlation. The result of a naturalness rating test and a comprehension test demonstrate that listeners perceive the matching interrogative-answer pairs more natural, compared to the conflicting pairs. Also, it is revealed that the prosodic patterns successfully guide listeners to identify the epistemic bias of negative polar interrogatives.

4.1.10 p.929

Einar Meister, Lya Meister,
L2 production of Estonian quantity degrees
The Estonian quantity system involves three contrastive patterns referred to as short (Q1), long (Q2)

and overlong (Q3) quantity degrees. Our previous studies have shown that for L2 learners the distinction between Q2 and Q3 is a difficult task in both production and perception. While Q1 and Q2 structures are always distinguished in the orthography, this is not the case in most Q2 and Q3 words excluding the words with plosives between first and second syllable vowels. Thus, the orthography might be the reason for the use of the same production pattern for both Q2 and Q3. The current paper studies the role of L2 orthographic input on the L2 production of Estonian quantity degrees by two groups of subjects with different language backgrounds: Finnish and Russian. The material used in the study involves word structures with and without orthographic manifestation of quantity contrasts. The results confirm the role of Estonian orthography on the L2 pronunciation, however, the two L2 subject groups show different prosodic patterns.

4.1.11 p.934

Rachel Steindel Burdin,
Variation in list intonation in American Jewish English
Yiddish-influenced intonation has been previously noted as a potential defining characteristic of American Jewish English, and list intonation was identified as a possible area of differentiation. However, apart from remarks in general descriptions of Standard American English (SAE) prosody, a systematic study of list intonation has not been conducted in SAE. In this study, lists were defined, and extracted from sociolinguistic interviews with Jewish women with varying degrees of exposure to Yiddish. The lists were then ToBI annotated. Speakers from different language backgrounds differed significantly in their use of contours, boundary tones and pitch accents on list items, with speakers with less exposure to Yiddish using more of the standard English contour (H* H-L%) than speakers with more exposure to Yiddish. Yiddish bilinguals were more likely to use a rise fall contour (L+H* L-L%), fewer H-L% boundary tones and H* pitch accents, and more rising pitch accents (L+H* and L*+H) than non-bilinguals. In addition, speakers of all language backgrounds used a variety of list intonations, showing the need for more systematic study into the uses and meanings, social and otherwise, of list intonations.

4.1.12 p.939

Bogdan Ludusan, Stefan Ziegler, Guillaume Gravier,
Is Syllable Stress Information Robust for ASR in Adverse Conditions?
This paper presents a study on the robustness of stress information for automatic speech recognition in the presence of noise. The syllable stress, extracted from the speech signal, was integrated in the recognition process by means of a previously proposed decoding method. Experiments were conducted for several signal-to-noise ratio conditions and the results show that stress information is robust in the presence of medium to low noise. This was found to be true both when syllable boundary information was used for stress detection and when this information was not available. Furthermore, the obtained relative improvement increased with a decrease in signal quality, indicating that the stressed parts of the signal can be considered islands of reliability.

4.1.13 p.944

John Dalton, John Kane, Irena Yanushevskaya, Ailbhe Ní Chasaide, Christer Gobl,
GlóRí - the Glottal Research Instrument
This papers presents GlóRí - the glottal research instrument. GlóRí is a speech analysis interface which offers a exibility and multiplicity of approaches to voice analysis. The system allows for fully automatic processing, for instance for analysis of large corpora. However, for more ne-grained studies, which may require precise voice source measurements, the systems facilitates manual optimisation of parameter settings. The present paper highlights the main features of the GlóRí system and provides illustrations of the usefulness of this approach.

4.1.14 p.949

Bettina Braun, Muna Pohl, Katharina Zahner,
Speech segmentation is modulated by peak alignment: Evidence from German 10-month-olds
In two headturn preference experiments, we tested whether German infants' segmentation strategies are

sensitive to the position of a pitch peak relative to the stressed syllable. Specifically, we compared target words with early-peak accents (where the pitch peak is early with respect to the stressed syllable, i.e. H+L* accents) and medial-peak accent (where the pitch peak is aligned with the stressed syllable, i.e. H* accents). Such differences in accent type signal mostly pragmatic distinctions, such as the difference between new and recoverable information. We familiarized infants with target words with one of the two intonation conditions that were embedded in sentences. We measured looking times to lists of trochaic part-words that were embedded in target words or were novel to them. Results showed an effect of familiarity only in the medial-peak condition, suggesting that infants at 10 months of age are very sensitive to pitch information for segmenting running speech.

4.1.15 p.954

Hae-Sung Jeon,
The Perception of Korean Boundary Tones by First and Second Language Speakers
This paper reports an experiment which investigated the perception of prosody in Korean or non-word utterances by native Korean speakers and English learners of Korean. Listeners rated the degrees of positivity and excitement of resynthesized utterances with different pitch ranges and durations. The results revealed no significant differences between the two groups of listeners. The variations in pitch range and duration had systematic effects on the ratings. However, the interactions between various factors suggest that the mapping between prosodic shapes and their paralinguistic meaning is not straightforward.

4.1.16 p.959

Irena Yanushevskaya, John Kane, Céline De Looze, Ailbhe Ní Chasaide,
The distribution of pitch patterns and communicative types in speech-chunks preceding pauses and gaps
This paper describes the distribution of pitch patterns and communicative types in the interpausal units (IPUs) preceding pause or gap silences extracted from a corpus of spontaneous speech as part of our work towards automatic prediction of turn-taking in dialogue interaction. IPUs preceding speaker change ('Gaps') and IPUs preceding silence where the same speaker continues talking ('Pauses') were selected in the course of automatic extraction of pause/gap silences in dyadic dialogue interactions. A listening test was conducted to estab-lish 'human predictable' pause/gap data sets which were sub-sequently manually annotated in terms of pitch patterns and communicative types. Overall, the Gaps and Pauses subsets show differentiation in terms of both their communicative types and pitch tunes. Declaratives and Questions are mainly found in Gaps, whereas in Pauses we mainly find Hesitations and Incomplete Declaratives. Gaps are generally characterised by falling or rising pitch patterns, whereas in Pauses a large proportion of speech samples are realised with level pitch. Classification experiments reveal strong discrimination of pauses and gaps for both prosodic and functional annotation labels.

4.1.17 p.964

Holly S.H. Fung, Peggy P.K. Mok,
Realization of Narrow Focus in Hong Kong English declarativesa Pilot Study
Narrow focus, i.e., focus on one word, is realized differently in native English and Cantonese. While it is signaled primarily by on-focus F0 changes such as F0 range expansion in English, it is marked essentially by lengthening of duration in Cantonese. Another difference is the pitch of the post-focus elements. While native English demonstrates post-focus F0 compression, Cantonese shows no significant post-focus pitch change. To investigate how narrow focus is realized in Hong Kong English (HKE), an emergent variety of English spoken by native speakers of Cantonese in Hong Kong, a controlled production experiment was conducted with 8 HKE speakers. Results showed that while the HKE speakers did realize foci with significant on-focus F0 range expansion, they exhibited no post-focus compression.

4.1.18 p.969

David Abelman, Robert Clark,
Altering speech synthesis prosody through real time natural gestural control
This paper investigates the usage of natural gestural controls to alter synthesised speech prosody in

real time (for example, recognising a one-handed beat as a cue to emphasise a certain word in a synthesised sentence). A user's gestures are recognised using a Microsoft Kinect sensor, and synthesised speech prosody is altered through a series of hand-crafted rules running through a modified HTS engine (pHTS, developed at Université de Mons). Two sets of preliminary experiments are carried out. Firstly, it is shown that users can control the device to a moderate level of accuracy, though this is projected to improve further as the system is refined. Secondly, it is shown that the prosody of the altered output is significantly preferred to that of the baseline pHTS synthesis. Future work is recommended to focus on learning gestural and prosodic rules from data, and in using an updated version of the underlying pHTS engine. The reader is encouraged to watch a short video demonstration of the work at http://tinyurl.com/gesture-prosody.

4.1.19 p.974

Xiaoluan Liu, Yi Xu,

Body size projection by voice quality in emotional speechEvidence from Mandarin Chinese

This study attempts to extend the line of research on using body size projection theory to account for emotional speech. It is predicted by the theory that anger is expressed by projecting a large body size with low pitch, rough voice and long vocal tract; happiness is expressed by projecting a small body size with high pitch, breathy voice and short vocal tract. Ten native speakers of Mandarin with drama training background recorded sentences in happy, angry, disgust and neutral emotions. We used multiple measurements to assess voice quality, formant dispersion (as an indicator of vocal tract length) and pitch. The results show clear support for the body size projection theory in voice quality, with anger and disgust associated with pressed and rough voice while happiness with breathy voice. But the results of formant dispersion and pitch demonstrate no clear directions. While the study is the first to show clear speech production support for the body size projection theory with voice quality data, the equivocal results of formant and pitch call for improvement in method of emotion elicitation in the laboratory.

4.1.20 p.978

Jan Michalsky,

Scaling of Final Rises in German Questions and Statements

Although certain intonation contours occur more frequently with German questions than with German statements, there is evidence that the semantics of intonational phonology operates on a more abstract level [1][2][3][4]. Hence, it is unlikely that there are pitch patterns in German that are exclusively used in interrogatives. Rather, intonational signaling of interrogativity can be regarded as resulting from the interaction between tonal and phonetic features. The tonal structure provides abstract semantic features, which are modified by paralinguistic features through phonetic realization [5]. This paper deals with the question which phonetic features may serve as cues to interrogativity in German. We report a reading task that was designed to elicit utterances that have phonologically identical nuclear rising pitch contours but differed by pragmatic function, serving either as a question or a statement. The observed absolute and relative scaling of nuclear and prenuclear tonal targets suggests that questions differ from statements by larger f0 excursions of nuclear rising contours, whereas the scaling of prenuclear accents does not substantially contribute to the expression of interrogativity. We conclude that phonetic cues to interrogativity in German are mainly realized through scaling and are restricted to the nuclear part of the intonational phrase.

4.1.21 p.983

Grace Kuo,

Processing Prosodic Boundaries in Natural and Filtered Speech

The prosody of an utterance can carry information that is critically important to understand the meaning of a sentence. In addition, previous studies have shown that listeners are able to detect major prosodic boundaries in their native language in stimuli whose segmental information has been removed, such as low-pass filtered [1][2] and hummed speech [2][3][5]. The present boundary strength rating study is conducted on native and non-native speakers to Taiwanese and Swedish, in an attempt to observe native and non-native speakers' accuracy in judging the upcoming boundary size in natural and filtered speech. 36 Taiwanese and American English speakers were recruited for the rating task whose stimuli consisted of Taiwanese and Swedish utterances from three prosodic boundary types (word boundary, phrase/tone sandhi group boundary, and Intonation Phrase boundary). In Experiment 1, participants rated the

upcoming boundary strength on a slider for filtered speech stimuli. In Experiment 2, they rated the boundary strength for natural speech stimuli. The results show that both non-native speakers could accurately predict the upcoming prosodic boundary type in both natural and filtered speech. The acoustic analyses of duration, f0 range, f0 median, spectral tilt, and harmonics-to-noise ratio reveal that non-native speakers use these prosodic cues to make their judgment, however, they put different emphasis on different cues when they were presented with stimuli of different qualities (natural vs. filtered) and lengths.

4.1.22 p.987

Malin Svensson Lundmark,
Constant Tonal Alignment in Swedish Word Accent II
Studies on accentual tonal alignment of intonation languages suggest that L in rising (LH) pre-nuclear accents anchors with a specific point in the segmental string, while the timing of H varies. This study investigates if lexical accents, too, exhibit a constant alignment by testing the South Swedish word Accent II. When under the strain of tempo variability the L-target was found not to be anchored with syllable onset. The results were not fully conclusive regarding H, but no clear evidence was found against anchoring of H, which could mean that H is an important phonological event in Accent II, while L is not.

4.1.23 p.992

Antonio Origlia, Francesco Cutugno,
A simplified version of the OpS algorithm for pitch stylization
In this work we present a new version of our previously published Optimal Stylization (OpS) algorithm for pitch stylization. Here we give a better perceptual representation of the pitch curve for linguistics research. While the OpS algorithm produced good stylizations for naive listeners, when deployed in a prosodic analysis tool, we observed that, under specific conditions, important details were missed in the stylized curve to an expert's ear. Changes introduced in the dynamic tonal perception model to solve these problems resulted in a simpler and more robust model. We show how the new version of the OpS algorithm is able to recover these situations while not significantly altering the original OpS curves.

4.1.24 p.997

Hiroaki Hatano, Carlos Ishi, Miyako Kiso,
Interpersonal factors affecting tones of question-type utterances in Japanese
The purpose of this paper is to clarify the interpersonal factors affecting phrase final tones of question-type utterances in Japanese daily conversations. We extracted question-type utterances ending with final particles from our dialogue speech database and classified them into two categories according to the degree of information request. Prosodic features were then analyzed by focusing on phrase final F0 movement and pitch reset. Analysis results indicated that F0 rising and falling degrees increase when the speaker express a close attitude to the dialogue partner, such as in conversations among family members and infant-directed speech. In addition, the presence of pitch reset in the phrase final was found to have functions of relieving the speaker's tension, when the dialogue partners have distant relationship.

4.1.25 p.1002

George Christodoulides, Cédric Lenglet,
Prosodic correlates of perceived quality and fluency in simultaneous interpreting
This study explores the relationship between prosodic features specific to simultaneous interpreting and listeners' perception of the fluency and accuracy of interpreting, as well as their comprehension of the source speech. Two groups of participants (47 subject experts and 40 non-experts) listened to a 20-minute lecture in German, along with its interpretation into French under two conditions (the actual interpretation, or a read-aloud rendition of the same text by the same interpreter) and answered comprehension and rating questions. The prosodic features of the two conditions were analysed, confirming differences regarding the temporal organisation of speech, disfluencies, pitch register and the interface between prosody and syntax. Our results suggest that interpreting-specific prosodic features affect the perception of fluency, which in turn affects the perception of accuracy; however the impact on listeners who enjoy relevant contextual knowledge is less pronounced.

4.1.26 p.1007

Ghania Droua-Hamdani, Sid-Ahmed Selouani, Yousef A. Alotaibi,
Rhythm analysis in Arabic L2 speech
This paper investigates rhythm speech metrics in Modern Standard Arabic. The corpus -West Point- includes recordings of native and non-native (L2) speakers. The experiment examines the rhythm metric tendencies of L2 speech using PVI and IM models. The study describes also the application of CCI (Control/Compensation Index) to the corpus. Variation in rhythm metrics by focusing on between-speaker differences such as gender of speakers is also studied.

4.1.27 p.1012

Rasmus Dall, Junichi Yamagishi, Simon King,
Rating Naturalness in Speech Synthesis: The Effect of Style and Expectation
In this paper we present evidence that speech produced spontaneously in a conversation is considered more natural than read prompts. We also explore the relationship between participant's expectations of the speech style under evaluation and their actual ratings. In successive listening tests subjects are presented with either spontaneously produced, read aloud or written sentences, and are asked to rate the naturalness of each sentence with either instructions toward conversational, reading or general naturalness. It was found that, when presented with spontaneous or read aloud speech, participants consistently rated spontaneous speech more natural - even when asked to rate naturalness in the reading case. Presented with only text, participants generally preferred transcriptions of spontaneous utterances, except when asked to evaluate naturalness in terms of reading aloud. This has implications for the application of MOS-scale naturalness ratings in Speech Synthesis, and potentially on the type of data suitable for use both in general TTS, dialogue systems and specifically in Conversational TTS, in which the goal is to reproduce speech as it is produced in a spontaneous conversational setting.

4.1.28 p.1017

Hao Liu, Yi Xu,
A Simplified Method of Learning Underlying Articulatory Pitch Target
Previous research has shown that parameters of the quantitative Target Approximation model (qTA) proposed by Prom-on and Xu can be directly extracted from natural speech with high accuracy through analysis-by-synthesis implemented in PENTAtrainers. While this may raise the possibility that PENTAtrainers actually simulate natural acquisition of prosody production, it is questionable that the human brain actually replicates the full articulatory mechanics represented by qTA in order to learn and control prosody production. In this paper we explore if a much simpler function can be used to extract at least some of the qTA parameters. We first managed to reduce the number of qTA parameters from three to two by evaluating their relative sensitivity. We then tested a pursuit function that learns only pitch target height and slope. Using a corpus of Mandarin utterances varying in lexical tone and focus, we show that parameters learned by the pursuit function can be used in qTA synthesis to generate F0 contours closely resembling those generated with parameters learned with qTA-based analysis-by-synthesis, with the advantage of having a much simpler learning algorithm. These results suggest that it is possible to learn articulatory control parameters for prosody without fully replicating the mechanical process itself.

4.1.29 p.1022

Maria Del Mar Vanrell, Meghan E. Armstrong, Pilar Prieto,
The role of prosody in the encoding of evidentiality
The overarching goal of this paper is to advance on the understanding of how evidential meanings are expressed in natural languages. Specifically, we aimed to investigate what type of meaning was encoded in yes-no questions through the combination of the question particle (QP) que 'that' and the nuclear intonational pattern L+H* L% in Majorcan Catalan yes-no questions (i.e., Que és un llibre? L+H* L% 'QP-It's a book?'), and to understand any temporal information that might be encoded through this construction. Several complementary research methods were used to address our question: the Discourse Completion Task, an acceptability task and a multiple-choice questionnaire. The results show that three types of information are encoded in QP que_L+H* L% questions: sentence modality, inference based on direct evidence and immediacy of the evidence.

4.1.30 p.1027

Pierre-Edouard Honnet, Alexandros Lazaridis, Jean-Philippe Goldman, Philip N. Garner,
Prosody in Swiss French Accents: Investigation using Analysis by Synthesis
It is very common for a language to have different dialects or accents. The different pronunciations of the same words is one of the reasons for the different accents, in the same language. Swiss French accents have similar pronunciation to standard French, but noticeable differences in prosody. In this paper we investigate the use of standard French synthetic acoustic parameters combined with Swiss French prosody in order to evaluate the importance of prosody in modelling Swiss French accents. We use speech synthesis techniques to produce standard French pronunciation with Swiss French duration and intonation. Subjective evaluation to rate the degree of Swiss accent was conducted and showed that prosody modification alone reduces perceived difference between original Swiss accented speech and standard French coupled with original duration and intonation by 29%.

4.1.31 p.1032

Ya Li, Jianhua Tao, Keikichi Hirose, Wei Lai, Xiaoying Xu,
Hierarchical stress generation with Fujisaki model in expressive speech synthesis
This paper introduces a hierarchical stress generation for expressive speech synthesis. In the previous study, we proposed a novel hierarchical Mandarin stress modeling method, and the text-based stress prediction experiments demonstrates a reliable stress assignment can be obtained from textual features. However, the stress model should be further verified to be an effective and efficient prosody model in a Text-to-Speech system. In this work, Fujisaki model known as an ideal global representation of prosody is adopted to construct the pitch contours. To illustrate the effect of stress model, the Fujisaki model parameters are automatically predicted by the textural feature with and without stress information. The synthetic speech sounds more natural than that without stress modeling. The RMSE of the pitch contour and the feature importance analysis also show stress information can improve the pitch modeling. This work offers a promising method to accurate pitch modeling for Mandarin expressive speech synthesis.

4.1.32 p.1037

Frank Zimmerer, Jeanin Jügler, Bistra Andreeva, Bernd Möbius, Jürgen Trouvain,
Too cautious to vary more? A comparison of pitch variation in native and non-native productions of French and German speakers.
This article presents preliminary results indicating that speakers have a different pitch range when they speak a foreign language compared to the pitch variation that occurs when they speak their native language. To this end, a learner corpus with French and German speakers was analyzed. Results suggest that speakers indeed produce a smaller pitch range in the respective L2. This is true for both groups of native speakers. A possible explanation for this finding is that speakers are less confident in their productions, therefore, they concentrate more on segments and words and subsequently refrain from realizing pitch range more native-like. For language teaching, the results suggest that learners should be trained extensively on the more pronounced use of pitch in the foreign language.

4.1.33 p.1042

Tomoyuki Mizukami, Hiroya Hashimoto, Keikichi Hirose, Daisuke Saito, Nobuaki Minematsu,
Selection of Training Data for HMM-based Speech Synthesis from Prosodic Features - Use of Generation Process Model of Fundamental Frequency Contours
-Generation process model of fundamental frequency (F0) contours is ideal to represent global movements of F0's keeping a clear relation with back-grounding linguistic information of utterances. Using the model, improvements of HMM-based speech synthesis are expected. A new method is developed to cope with erroneous F0's of utterances included in HMM training corpus. F0 extraction errors not only cause wrong F0's, but also degrade segmental features of synthetic speech, since they affect the over-all accuracy of speech analysis. The method is to exclude speech segments from HMM training, where extracted F0's are largely different from those generated by the generation process model. Experiments on speech synthesis showed a clear improvement in synthetic speech quality when phoneme-base exclusion is conducted with a properly selected threshold.

4.1.34 p.1047

Alexandros Lazaridis, Pierre-Edouard Honnet, Philip N. Garner,
SVR vs MLP for Phone Duration Modelling in HMM-based Speech Synthesis
In this paper we investigate external phone duration models (PDMs) for improving the quality of synthetic speech in hidden Markov model (HMM)-based speech synthesis. Support Vector Regression (SVR) and Multilayer Perceptron (MLP) were used for this task. SVR and MLP PDMs were compared with the explicit duration modelling of hidden semi-Markov models (HSMMs). Experiments done on an American English database showed the SVR outperforming the MLP and HSMM duration modelling on objective and subjective evaluation. In the objective test, SVR managed to outperform MLP and HSMM models achieving 15.3% and 25.09% relative improvement in terms of root mean square error (RMSE) respectively. Moreover, in the subjective evaluation test, on synthesized speech, the SVR model was preferred over the MLP and HSMMmodels, achieving a preference score of 35.93% and 56.30%, respectively.

4.1.35 p.1057

Decha Moungsri, Tomoki Koriyama, Takashi Nose, Takao Kobayashi,
Tone Modeling Using Stress Information for HMM-Based Thai Speech Synthesis
This paper describes a modeling technique of Thai tones for HMM-based speech synthesis. Tones are important prosodic features for tonal language including Thai because the phonetically same words but with different tones give different meanings. Although there have been several approaches to improving tone correctness of synthetic speech by considering tone types, another significant factor, stress, was not used explicitly for prosody modeling. We incorporate stress/unstress information into the framework of the HMM-based speech synthesis. Objective and subjective evaluation results show that the use of stress information improves the performance in Thai tone modeling.

4.1.36 p.1062

D. Gomathi, P. Gangamohan, B. Yegnanarayana,
Understanding the significance of different components of mimicry speech
Voice conversion systems aim at finding a transformation function using statistical models. Mimicry/ Voice imitation is a natural voice transformation technique which sounds convincing to the listeners. It thus seems advisable to study the transformation used by human beings who perform mimicry. The objective of this study is to examine the various components of speech that are modified during voice imitation. To transform a given speech utterance to sound like that of a target utterance, the process needs to be understood at both production and perception level. In this paper the importance of source and system parameters and also the significance of different components of speech that contribute to the perception of imitation are studied. A flexible analysis-synthesis tool is used to modify the features of natural utterance and convert it to imitated utterance. Perceptual studies are carried out to understand if the modified features contribute to imitation. The results show that a combination of features is varied by the imitator to achieve imitation and they vary depending on the target speaker.

4.1.37 p.1067

Hansjörg Mixdorff, Angelika Hönemann, Jeesun Kim, Chris Davis, Grégory Zelic,
The Cartoon Task Exploring Auditory-Visual Prosody in Dialogs
This paper introduces and evaluates a collaborative task designed to elicit auditory-visual dialogs. The task was based on the viewing of two versions of the same cartoon film that was edited so that in order to reconstruct the story information from two incomplete versions must be shared in a consecutive fashion. The aim of this design was to elicit a relatively balanced dialog between the two participants as the story is pieced together from the beginning to the end. The current paper describes the production of a corpus consisting of audio, video and motion capture data from 22 pairs of Australian English speaking participants, and presents results on turn-distribution and raw prosodic features. Our analysis showed that the task could produce relatively balanced dialogs although this was not the case for all pairs. Analysis of raw prosodic features did not suggest that convergence occurred over the conversation, but replicated earlier findings of similarity between partners as compared to others.

4.1.38 p.1072

Peggy P.K. Mok, Holly S.H. Fung, Jingwen Li,
A preliminary study on the prosody of broadcast news in Hong Kong Cantonese
Broadcast news is a distinctive register. Previous studies only provided some general descriptions of the prosodic features in broadcast news but with few concrete data. Most of them were also on English news. This study investigated the prosodic features of Cantonese TV broadcast news using acoustic data. Speech using the same materials from two groups was compared: eight Hong Kong professional TV news anchors, and a control group consisting of eight university students. The results show clear differences between the two groups in terms of speech rate, pitch range and variability of syllable duration (speech rhythm). It was found that the news anchors spoke significantly faster than the control group, also with an enlarged pitch range. They also produced more variability in syllable duration. There is clearly more prosodic variation in the news register than ordinary speech. Finally, we provide some possible reasons for these features, as well as directions for future studies.

4.1.39 p.1052

Elisabeth Delais-Roussarie, Ingo Feldhausen,
Variation in Prosodic Boundary Strength: a study on dislocated XPs in French
Three independently motivated types of information are usually assumed to influence prosodic boundary placement and to play a role in their relative strength: the morpho-syntactic structure, the information structure and the metrical complexity. The phonetic realization associated with the different boundary types (in particular IP and ip) is also assumed to vary. Based on data of clitic left-dislocations in French, we argue here that differences in the relative strength of the prosodic boundary occurring at the end of the dislocated XP (i.e. an intermediate (ip) or an intonational phrase (IP) boundary) cannot be derived in a straightforward manner from these three types of information. In a production experiment, where the syntactic and information structure were controlled, while the metrical complexity was varied, it appeared that the strength of the boundary occurring at the right edge of the dislocated object NP displayed a high degree of variability. In addition, the results indicate a lack of correlation between metrical complexity and boundary strength. The results lead us to argue that a sort of phonological neutralization occurs in certain textual contexts. This neutralization does not allow for distinguishing between intermediate and intonational phrase boundaries in all cases.

4.1.40 p.1076

Sébastien Le Maguer, Elisabeth Delais-Roussarie, Nelly Barbot, Mathieu Avanzi, Olivier Rosec, Damien Lolive,
Prosodic chunking algorithm for dictation with the use of speech synthesis
The aim of this paper is to present an algorithm that automatically segment a text in prosodic chunks for a dictation by conforming to the rules and procedures used in real settings to dictate a text to primary school children. A better understanding and modeling of these rules and procedures is crucial to develop robust automatic tools that could be used in autonomy by children to improve their spelling skills through dictation with the use of speech synthesis. The different steps used to derive the prosodic chunks from a given text will be explained through concrete examples. The proposal made here relies on the analysis of a corpus of 10 dictations given to children in French and French Canadian elementary schools, and more precisely during their first three years in elementary school (i.e., cycle 2 in the French school system). The phrasing observed in the data is described. It is thus simplified in order to develop an algorithm that automatically generates prosodic chunks from texts.

4.1.41 p.1081

Jitka Vaňková, Radek Skarnitzl,
Within- and Between-Speaker Variability of Parameters Expressing Short-Term Voice Quality
This study focuses on short-term acoustic correlates of voice quality. It assesses the within-speaker stability (across different speaking styles) and between-speaker variability of measurements which compare the amplitudes of various spectral events H1*-H2*, H2*-H4*, H1*-A1*, H1*-A2* and H1*-A3*. Although speakers do differ with regard to the compactness of the parameters in read and spontaneous speaking styles, the parameters H1*-H2*, H1*-A1* and H1*-A2* appear both considerably stable for one speaker

in different speaking styles and efficient in between-speaker comparisons. Though not directly applicable in forensic settings, these glottal parameters outperformed vowel formants in classification using LDA.

4.1.42 p.1086

Bénédicte Grandon, Hiyon Yoo,
Do Korean L2 learners have a "foreign accent" when they speak French? Production and perception experiments on rhythm and intonation
French and Korean are two languages with similar prosodic characteristics as far as rhythm and intonation are concerned. In this paper, we present the results of production and perception tests where we describe the prosodic characteristics of Korean L2 learners of French. The aim is to analyze the impression of "foreign accent" for two prosodic components (intonation and rhythm) of speech produced by Korean L2 learners of French and the perception of this "accent" by native listeners of French (L1). We show that the productions of Korean learners and French native speakers present minor differences but that they do no translate into cues for determining clearly the presence of a "foreign accent".

4.1.43 p.1091

Linda Garami, Anett Ragó, Ferenc Honbolygó, Valéria Csépe,
Prosodic processing in the first year of life: an ERP study
From early months of life prosody has a prominent contribution to segmentation: prosodic boundaries overlap with syntactic ones and facilitate the extraction of syntactic regularities both at word and at phrase level. Therefore, the long-term representation of rhythmic features of the native language, especially the stress templates derived from regularities are assumed to play a particular role in pre-lexical processing. We examined the nature of early stress representation in a language with a fixed stress pattern in an electrophysiological experiment (acoustic passive odd-ball paradigm, 10 month-olds: 28 infants; 6 month-olds: 21 infants, 400 items, deviant: p=20%) using bi-syllabic Hungarian pseudo-words to follow how prosodic features contribute to processing saliency and how word stress templates based on regularities may emerge. We used legally and illegally stressed stimulus both in standard and deviant positions in separate conditions. In the legal standard condition two mismatch responses (MMRs) temporally synchronized to each syllable could be recorded. On the contrary, in the illegal standard condition no significant response was found. It seems that language environment influences the processing of speech prosody and the MMR correlates of word stress processing are related both to saliency and to stress templates emerging during the first year of life.

4.1.44 p.1095

Lei He,
The Inadequacy of Rhythm Metrics to Quantify L2 Suprasegmental Characteristics
The study investigated the L2 speech rhythm of Chinese English speakers (L1 = Mandarin) using the metrics of deltaV, deltaC, %V, VarcoV, VarcoC, rPVI-C and nPVI-V. Five native speakers of American English and Mandarin were recruited to record five sentences in English. In addition, the Chinese speakers also recorded five Mandarin sentences. One-way ANOVAs were conducted to see if significant differences exist on each of the metrics among L1 English, L2 English and L1 Mandarin. Results show that the two L1's are categorically distinct on all metrics, conforming to the perceptually distinct rhythmicities of English and Mandarin. However, no significant differences were found between L1 and L2 English which have different intuitive rhythmicities, suggesting that the metrics are inadequate to capture the suprasegmental details that give the final make-up of speech rhythm.Finally, new directions of speech rhythm research and new applications of the rhythm metrics are sketched.

4.1.45 p.1100

Céline De Looze, Irena Yanushevskaya, John Kane, Ailbhe Ní Chasaide,
Pitch range declination and reset in turn-taking organisation
In this paper, we investigate how pitch range declination and reset contribute to turn-taking organisation. We (i) investigate the effect of the unit position in a turn on its pitch range as well as (ii) compare the difference in pitch range between consecutive units that are separated by a gap vs. a pause. We also (iii) test the effect of the number of speech units in a turn as well as the turn duration on the peak height

at the beginning of the turn. Our results suggest a pitch range declination trend between the Initial and Median speech units of a turn but a violation of this declination for the Final units of the turn. Consequently, the difference in two consecutive units' pitch range is found larger at pauses than at gaps. Our results also show that the higher the number of speech units in a turn or the longer the turn, the higher the peak height. Our findings particularly reveal that the distance between the pitch range level and its upper limit may be a salient cue in projecting the end of a turn. We discuss our findings along the debate on Projection and Reaction theories and that of Hard vs. Soft pre-planning of speech production, and address how these findings may be useful for human-machine interactions.

4.1.46 p.1105

Monica Dominguez, Mireia Farrús, Alicia Burga, Leo Wanner,
Towards Automatic Extraction of Prosodic Patterns for Speech Synthesis
This paper deals with the adaptation of AuToBI annotation for speech synthesis purposes. AuToBI is a tool that automatically detects and classifies the standard ToBI labels for American English. AuToBI annotation is performed word-by-word. However, a labeling of intonation patterns at the intonational phrase level is essential for the detection of the correlation between theme/rheme (thematicity) and prosody and also much more appropriate for speech synthesis applications that use various layers of linguistic annotation (syntax, semantic, information, and prosody structures), such that if used in speech synthesis applications, AuToBI's output would require a post-processing stage of the extracted labels. We present a procedure that includes an initial AuToBI annotation and the adaptation of the AuToBI output to a phrase-based annotation, following a set of determined rules. A further analysis of the correspondence between prosodic patterns and themacity structures is used to validate the results.

4.1.47 p.1110

Vasilisa Verkhodanova, Vladimir Shapranov,
Automatic Detection of Filled Pauses and Lengthenings in the Spontaneous Russian Speech
During automatic speech processing a number of problems appear, and among them there are such as speech variation and different kinds of speech disfluences. In this article an algorithm for automatic detection of the most frequent of them (filled pauses and sound lengthenings) based on the analysis of their acoustical parameters is presented. The method of formant analysis was used to detect voiced hesitation phenomena and a method of band-filtering was used to detect unvoiced hesitation phenomena. For the experiments on filled pauses and lengthenings detection a specially collected corpus of spontaneous Russian map-task and appointment-task dialogs was used. The accuracy of voiced filled pauses and lengthening detection was 82%. And accuracy of detection of unvoiced fricative lengthening was 66%.

4.1.48 p.1115

Aline Pessoa-Almeida, Alexsandro Meireles, Sandra Madureira, Zuleica Camargo,
Prosodic analysis of the speech of a child with cochlear implant
According to previous studies [1,6], acoustic and perceptual analysis can be considered useful clinical tools to investigate the speech characteristics of hearing impaired children (HIC). This research aimed at describing the perceptual and acoustic correlates of the speech samples from a HI and user of CI child (within the chronological age range of 5 years and 1 month and 7 years and 1 month), through the vocal quality and voice dynamics descriptions in two different moments. The speech samples were collected during speech therapy sessions. The perceptual analysis of the vocal quality was based on the Vocal Profile Analysis Scheme for Brazilian Portuguese (BP-VPAS - Camargo &Madureira, 2008). The recorded corpus was analyzed through the ExpressionEvaluator script (Barbosa, 2009) ran by Praat software v5.2.10. The measures, which were automatically extracted, comprised the fundamental frequency f0, first f0 derivative, intensity, spectral slope and long-term mean spectrum. The correlations found between the acoustic and perceptual data proved relevant for rehabilitation processes.

4.1.49 p.1119

Tatiana Luchkina, Jennifer Cole,
Structural and Prosodic Correlates of Prominence in Free Word Order Language Discourse
Production and perception experiments with native speakers of Russian, a free word order language, show

that prosody and change in word order are used to mark discourse-prominent constituents. Concurrent application of these cues to prominence is possible, as evident from distinctively higher f0 and intensity maxima, and duration values associated with ex- situ words, as well as their higher visibility in discourse. Distinctive acoustic-prosodic realization of ex-situ words may cue their relatively high informational load and discourse prominence, as well as (redundantly) signal that the word is left- or right-dislocated.

4.2 Friday Session Two

11am - 1pm : 4-2-oral (6 presentations)

- intonation -

4.2.1 p.1125

Jonathan Barnes, Alejna Brugos, Nanette Veilleux, Stefanie Shattuck Hufnagel
Segmental Influences on the Perception of Pitch Accent Scaling in English
In both tone and intonation systems, segmental context is known to influence production and perception of target F0 contours in various ways. Many languages, for example, prefer to realize critical F0 events during maximally sonorous intervals, either by varying the timing of pitch movements, or by virtue of distributional limitations on certain contour types. Current analytic practice, by contrast, routinely ignores segmental backdrop when estimating the perceptual efficacy of putative cues, such as F0 turning points, to tone scaling and timing patterns. Results of the perception study presented here argue that pitch accent scaling is best modeled using a weighted average of F0 sampled over a defined region of interest, and that individual sample weights are determined in part by the sonority of the segments from which they are taken. That is, samples from lower sonority segments contribute less to integrated scaling percepts than those from higher sonority segments. This model, called TCoG-F(requency), accounts for crosslinguistic tonal timing and distribution patterns in the literature, and underscores the danger of analyzing tonal phenomena completely apart from the segments that express them.

4.2.2 p.1130

Kristine M. Yu, Sameer Ud Dowla Khan, Megha Sundara,
Intonational phonology in Bengali and English infant-directed speech
We examined the phonetics and phonology of intonation of infant-directed speech (IDS) and non-IDS in story-reading in two typologically-divergent languages, English and Bengali. In addition to finding an increase in f0 range and variability in IDS, replicating previous work on IDS prosody, we found novel evidence that f0 manipulations in IDS are constrained by intonational phonology. Speakers in both languages used an increased proportion of tonal elements with higher tonal targets and more turning points in IDS, within the language-specific intonational grammar. The tonal elements showing increased use in IDS also were associated with marking topic and focus. Thus, phonetic changes in IDS may in part be induced by speakers' choices of phonological tonal elements, which in turn may be connected with choices about marking discourse structure.

4.2.3 p.1135

Eszter Varga, Zsuzsanna Schnell, Gabor Perlaki, Gergely Orsi, Mihály Aradi, Tibor Auer, Flora John, Tamás Dóczi, Samuel Komoly, Norbert Kovács, Attila Schwarz, Tamás Tényi, Róbert Herold, József Janszky, Réka Horváth,
Hemispheric lateralization of sentence intonation in left handed subjects with typical and atypical language lateralization: an fMRI study
Introduction: Prosody (as the melody of speech) is an important component of human social interactions. More specifically, linguistic prosody conveys meaning of speech through syllable, word, or sentence level stress and intonation. In the modern neuroimaging era the hemispheric representation of sentence intonation is widely investigated. Most of these studies suggest bilateral activations predominantly in the perisylvian language areas and in the subdominant homologues. However, there are some inconsistencies about the hemispheric representation and lateralization of linguistic prosody. These inconsistencies could be due to the lack of attention on the language lateralization of the subjects. Aims: The present study aims to investigate the hemispheric representation and lateralization of linguistic prosody with a

sentence intonation task in two groups of left handed subjects with typical and atypical language lateralization. Functional MRI was used to test the assumption that - according to the functional lateralization hypothesis - the representation of sentence intonation is predominantly lateralized within the language dominant hemisphere and the lateralization of sentence intonation is associated with language lateralization in both groups. Methods: Left handers were examined to create two groups of subjects with typical and atypical language lateralization. In all, 32 healthy subjects were evaluated with a standard verbal fluency task with fMRI in order to assess functional hemispheric language lateralization. In our final investigation the atypical group consisted of 8 subjects with right hemispheric language dominance (LI<-0.2) and the typical group also consisted of 8 subjects with left hemispheric language dominance (LI>0.2). Sentence intonation task was utilized to test linguistic prosody skills with fMRI. 49 pairs of sentences (18 pairs of neutral-neutral sentences, 10 pairs of interrogative-interrogative sentences, and 1 pair of interrogative-neutral sentence) were presented with an event-related design. Sentences were matched in terms of syntactic structure, semantic complexity and length and all were affectively neutral. In the fMRI data analysis interrogative pairs were compared to neutral pairs. Results: One of the main findings of our study is that subjects with both typical and atypical language lateralization activated the middle temporal gyrus (MTG) on the right side. The activation of the MTG on the right hemisphere is classically associated with the encoding of prosodic information. Furthermore, both groups recruited the frontal language areas only in the language-dominant hemisphere. Moreover, between-group comparison showed significantly stronger activations in subjects with typical language lateralization only in left sided language areas: pars triangularis of the inferior frontal gyrus, the superior frontal gyrus and the inferior parietal lobule. Conclusion: This finding is in accordance with the functional lateralization hypothesis of prosody, and suggests a correlation between linguistic prosody lateralization and language lateralization.

4.2.4 p.1139

Meghan Armstrong, Núria Esteve-Gibert, Pilar Prieto,

The acquisition of multimodal cues to disbelief

In this study, we examine how 3-, 4-, and 5-year-old Catalan-acquiring children are able to make use of the audio (intonational) and visual (facial gesture) modalities in the comprehension of speaker disbelief, as well as the role of a child's developing Theory of Mind. Our results suggest that in this case, facial gesture provides children with scaffolding for linguistic meaning. In addition, those children that passed a false belief task tended to perform better on the comprehension task in general. We discuss the implications of these findings for the study of intonational development.

4.2.5 p.1144

Amalia Arvaniti, Mary Baltazani, Stella Gryllia,

The pragmatic interpretation of intonation in Greek wh-questions

We experimentally investigated the pragmatics of two melodies commonly used with Greek wh-questions, L*H L-!H%, described as the default, and LH* L-L% considered less frequent and polite. We tested two hypotheses: (a) the !H%-ending melody is associated with information-seeking questions, while the L%-ending melody is pragmatically more flexible and thus appropriate also for non-information-seeking wh-questions expressing bias; (b) the !H%-ending melody, being more polite, is more appropriate for female talkers, all else being equal. In Experiment 1, comprehenders rated !H-ending and L%-ending versions of the same questions for politeness and appropriateness for the context in which they were heard (which favored either information-seeking or "biased" wh-questions). In Experiment 2, comprehenders heard the same questions and chose between two follow-up responses, one providing information, the other addressing the bias of the wh-question. Comprehenders rated !H%-ending questions more appropriate than L%-ending questions and judged the !H%-ending questions of female talkers more polite. They also chose information-providing answers more frequently after !H%- than L%-ending questions, but the preference was higher for female talkers and depended on comprehender gender. The results argue in favor of a compositional view of intonational meaning which depends not only on the tune but also on context, broadly construed.

4.2.6 p.1149

Pablo Arantes, Anders Eriksson,

Temporal stability of long term measures of fundamental frequency

We investigated long-term mean, median and base value of F0 to estimate how long it takes for their variability to stabilize. Change point analysis was used to locate stabilization points. In one experiment

stabilization points were calculated in recordings of the same text spoken in 26 languages. Average stabilization points are 5 seconds for base value and 10 seconds for mean and median. Variance after the stabilization point was reduced around 40 times for mean and median and more than 100 times for the base value. In other experiment, four speakers read each two different texts. Stabilization points for the same speaker across the texts do not exactly coincide as would be ideally expected. Average point dislocation is 2.5 seconds for the base value, 3.4 for the median and 9.5 for the mean. After stabilization, individual differences in the three measures obtained from the two texts are on average 2% on average. Present results show that stabilization points in long-term measures of F0 occur earlier than suggested in the previous literature..

4.3 Friday Session Three

2pm - 3:30pm : 4-3-plenary (1+3 presentations)

4.3.1 KeyNote 4

Anne Cutler - Invited Keynote:

Aspects of the suprasegmental structure of speech are famously subject to speaker choice. There is no obligatory location for accent in a sentence such as "She didn't run home"; speakers may say "SHE didn't run home" or "She DIDN'T run home" or "She didn't RUN home" or "She didn't run HOME", with different resulting inferences in each case. But do listeners also have any degree of choice in the auditory processing of this dimension of speech? This presentation will argue that they do, and support the argument with evidence from laboratory studies of spoken-word recognition, of semantic structure computation in spoken sentences, and of the processing of delexicalised prosodic signals.

4.3.2 p.1154

Meredith Brown, Laura Dilley, Michael Tanenhaus,
Probabilistic prosody: Effects of relative speech rate on perception of (a) word(s) several syllables earlier
Speech perception depends on the ability to rapidly accommodate considerable variability in speech rate. We present results from two eye-tracking experiments indicating that listeners use context speech rate to generate, maintain, and update probabilistic hypotheses about the timing and number of constituents in upcoming speech. Participants heard utterances containing polysyllabic nouns preceded by indefinite articles and followed by [s]-initial words (e.g. ...saw a raccoon slowly...). We altered the speech rate of the indefinite article and of the [s] with respect to surrounding context, manipulating the likelihood that the item would be perceived as singular (a raccoon) vs. plural (raccoons). Shorter indefinite articles elicited higher proportions of fixations to plural target pictures than longer articles both before and after the processing of [s], demonstrating that listeners made rapid use of prosodic cues to the presence or absence of the article. Importantly, fixations were also influenced by the duration of [s] relative to context speech rate. These findings suggest that listeners maintain and update provisional speech-rate hypotheses across multiple morphophonemic units. We interpret these results with respect to probabilistic approaches to spoken language understanding.

4.3.3 p.1159

Jill C. Thorson, James L. Morgan,
The role of intonation in early word recognition and learning
The motivation for our study is to investigate how English-acquiring toddlers are guided by the mapping between intonation and information structure during on-line reference resolution and in novel word learning tasks. We ask whether specific pitch movements (deaccented, monotonal, bitonal) more systematically predict patterns of attention and subsequent novel word learning abilities depending on the referring or learning condition (new, given, contrastive). Experiment 1 examines the attentional patterns of 18-month-old toddlers when referents are either new or given in the discourse, and carry one of the three pitch accent types. Contrary to previous work, results show increased attention to the target in the deaccented condition if the referent is new to the discourse. Also, both monotonal and bitonal pitch movements direct attention to the target even when the target is given. Thus, pitch type interacts with information structure in directing toddler attention. Experiment 2 tests two-year-olds in a novel word

learning task, varying pitch type and contrastiveness during learning. Preliminary results show that learning is aided when the novel word is introduced in contrast to a previous referent. Together, these two experiments demonstrate the role of pitch type and information structure in guiding attention and aiding early word learning.

4.3.4 p.1164

Rory Turnbull, Adam J. Royer, Kiwako Ito, Shari R. Speer,
Prominence perception in and out of context
The perception of prosodic prominence is known to be influenced by several distinct factors. In this study, we investigated the role of context, both global and local, in the prominence judgements of naïve listeners. Monolingual English listeners marked where they heard prominence on pairs of two-word phrases (e.g. *blue ball*, *green drum*). Stimuli varied in whether or not the first phrase implied a contrastive focus on the second phrase. We found clear evidence of a hierarchy of prominence across pitch accent types: L+H* >H* >!H* >unaccented. Additionally, we found that contrast status only affected prominence markings when the participants were made explicitly aware of the discourse context and were instructed to imagine themselves physically present to observe the conversation. This effect of global context suggests that information structure cannot be reliably interpreted in the absence of an established discourse context. Taken together, these results suggest that naive listeners are sensitive to prominence differences at levels corresponding to categorical annotations. Perception of a word's relative prominence was consistently influenced by phonetic and phonological factors, while pragmatic factors (such as contrast-evoking context) required more elaborate plausibility manipulations in order to affect prominence perception.

4.4 Friday Closing Session

Firewall Speeches . . .