

6-10 September 2009

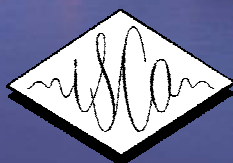


CONFERENCE PROGRAMME & ABSTRACT BOOK

Speech and Intelligence

Interspeech 2009 Brighton UK

10th Annual Conference of the
International Speech
Communication Association



www.interspeech2009.org

Map of Venue & Central Brighton

Conference venue highlighted with red arrow.



Map courtesy of





6-10 September 2009
The Brighton Centre
Brighton, United Kingdom

Copyright © 2009 International Speech Communication Association
<http://www.isca-speech.org>
secretariat@isca-speech.org
All rights reserved.
ISSN 1990-9772 Proceedings CD-ROM
Editors: Maria Uther, Roger Moore, Stephen Cox

Acknowledgements:

Photos of Brighton and map of Brighton courtesy of Visit Brighton (copyright holder), Brighton & Hove City Council. Thanks also to our colleagues on the organising committee, Meeting Makers and Brighton & Hove City Council for contributing relevant information for this booklet. Thanks to Dr. James Uther for assistance on various elements of the booklet.

Abstracts and Proceedings CD-ROM Production by
Causal Productions Pty Ltd
<http://www.causalproductions.com>
info@causalproductions.com

Welcome to Interspeech 2009

Message from the ISCA President

On behalf of the International Speech Communication Association (ISCA), welcome to INTERSPEECH 2009 in Brighton. Ten years ago, in Budapest, we took the first steps towards creating what later became a unified Association and a unified conference, joining the best of the former EUROSPEECH and ICSLP conferences. This conference is the 10th in the long cycle that bears the INTERSPEECH label and that already includes conferences in such wonderful venues as Beijing, Aalborg, Denver, Geneva, Jeju, Lisbon, Pittsburgh, Antwerp and Brisbane.

I would like to start this long message of thanks and commendation by honouring the ISCA Medalist for 2009, Prof. Sadaoki Furui, for his outstanding contributions to speech processing, and leadership as one of the first Presidents of the *International* Speech Communication Association.

At this meeting we shall also recognize several other ISCA members who, for their technical merit and service to the Association, were recently elected as ISCA Fellows: Anne Cutler, Wolfgang Hess, Joseph Mariani, Hermann Ney, Roberto Pieraccini, and Elizabeth Shriberg.

Next in my list of ISCA volunteers to thank are the members who are promoting research activities on speech science and technology by giving lectures in different parts of the world, as ISCA Distinguished Lecturers: Rich Stern, Abeer Alwan, and James Glass. ISCA's growing efforts to promote speech science and technology research are reflected in the work of Special Interest Groups, and International Sub-Committees as well, in the many workshops spanning multidisciplinary areas that continuously enlarge our electronic archive, in the increasing number of grants to students, in our very long newsletter ISCApad, and in many other activities. Of particular relevance are the ones undertaken by our very active Student Advisory Committee, who recently launched a resume posting service. Thank you all for your continuing help and support!

The coordination of all these activities is the responsibility of the ISCA Board, and I would like to particularly thank the two members who have completed their terms in 2009 for all their efforts for the community: Eva Hajičova and Lin-shan Lee. The past year has been a year of expansion, but also of consolidation of many new different activities. This was the prime motivation for enlarging the Board to 14 members. I take this opportunity to welcome the new members: Nick Campbell, Keikichi Hirose, Haizhou Li, Douglas O'Shaughnessy, and Yannis Stylianou.

For INTERSPEECH 2009, over thirteen hundred papers were submitted, and approximately 57% of the regular papers were selected for presentation after the peer review process, with at least 3 reviews per paper. Many members of the international community participated in this review process, and I hope to have the chance to thank you all personally in Brighton.

The large number of submissions marks INTERSPEECH conferences as the major annual conferences in speech science and technology, a role that we would like to further enhance by getting these conferences included in major citation indexes. We hope that this number keeps increasing in the next events in Makuhari (2010) and Florence (2011), and Portland (2012).

This conference is chaired by the first ISCA President, Prof. Roger Moore. We are particularly appreciative of his organizing team for testing a new model for INTERSPEECH organization, independent of any university sponsorship. For this bold step, and for all the work and devotion that you have put into organizing this conference, thank you all so much!

Having heard you discuss your plans for almost 4 years, we are looking forward to a superbly organized conference, with excellent keynotes and tutorials, interesting sessions, a very lively social programme, and innovations such as the Loebner award.

I invite you all to join us in celebrating the 10th anniversary of ISCA and wish you a very successful conference.

Isabel Trancoso, ISCA President

Message from the General Chair

Dear Colleague,

On behalf of the organising team, it is with great personal pleasure that I welcome you to Brighton and to *INTER_SPEECH-2009*: the 10th Annual Conference of the International Speech Communication Association (ISCA).

It has been almost four years since we put together the bid to host INTER_SPEECH in the UK, and it has been a truly momentous experience for everyone involved. A large number of people have worked very hard to bring this event to fruition, so I sincerely hope that everything will run as smoothly as possible, and that you have an enjoyable and productive time with us here in Brighton.

The theme of this year's conference is *Speech and Intelligence*, and we have arranged a number of special events in line with this. For example, on Sunday we are hosting the 19th annual *Loebner Prize for Artificial Intelligence* (a text-based instantiation of a Turing test) run by Hugh Loebner himself. Also on the Sunday, and inspired by the Loebner Prize, for the first time we will be attempting to run a real-time speech-based version of the Turing test. Although we only have a couple of contestants, we hope that it will be an informative and entertaining aspect of the day's activities. Another conference event related to *Speech and Intelligence* is a special semi-plenary discussion that is scheduled to take place on Tuesday at 16:00 (Tue-Ses3-S1). This will involve a number of distinguished panellists who have agreed to engage in a lively Q&A interaction with the audience. Please come along and join in with the debate.

As well as the two competitions, other events taking place on Sunday include eight excellent and varied Tutorials presented by fourteen top quality lecturers. Tutorials have become a popular feature of INTER_SPEECH conferences, and several hundred attendees take advantage of the opportunity to learn at first-hand some of the core scientific principles underpinning different aspects of our developing field. Another first for INTER_SPEECH-2009 is Sunday's public exhibition of speech-related activities. Public engagement with science is an important issue in modern society, and we are very grateful for the help and support that we have received to mount this event from the UK Engineering and Physical Sciences Research Council funded science outreach project *Walking with Robots* (<http://www.walkingwithrobots.org/>). It will be interesting to see what the people of Brighton make of our particular brand of science and technology – let's hope for some positive feedback!

Monday sees the beginning of the main conference programme and, after the opening ceremony (during which we will pay tribute to our recently departed senior colleague and ESCA Medallist, Gunnar Fant), we are honoured to welcome this year's *ISCA Medallist*, Prof. Sadaoki Furui (Tokyo Institute of Technology), who will be presenting the first Keynote talk of the conference. Sadaoki's subject is "*Selected Topics from 40 Years of Research on Speech and Speaker Recognition*", and I'm sure that he will provide us with a wealth of interesting insights into the progress that he has seen at the forefront of these areas of research.

The main technical programme of oral and poster sessions starts after lunch on Monday and runs through until Thursday afternoon. This year we received an almost record number of submissions: 1303 by the published April deadline. These were assessed by 645 reviewers, and the resulting ~4000 reviews were organised by the 24 Area Coordinators so that the final accept/reject decisions could be made at the Technical Programme Committee meeting held in London at the beginning of June. This careful selection process resulted in the acceptance of 762 papers (707 in the main programme and 55 in special sessions), all of which means that we have a total of 38 oral sessions, 39 poster sessions and 10 special sessions at this year's conference.

In addition to the main programme, each day starts with a prestigious Keynote talk from a distinguished scientist of international standing. On Tuesday, Tom Griffiths (UC Berkley) will present his talk entitled "*Connecting Human and Machine Learning via Probabilistic Models of Cognition*"; on Wednesday, Deb Roy (MIT Media Lab) promises to lead us towards "*New Horizons in the Study of*



Language Development"; and on Thursday, Mari Ostendorf (University of Washington) will address her topic of "*Transcribing Speech for Spoken Language Processing*". Keynote presentations are often the scientific highlight of any conference, so I hope that, like me, you are looking forward to some stimulating early morning talks.

Alongside the regular sessions, we also have a number of special sessions, each of which is devoted to a 'hot' topic in spoken language processing. Daniel Hirst has organised a session on '*Measuring the rhythm of speech*'; Oliver Lemon and Olivier Pietquin have put together a session on '*Machine learning for adaptivity in spoken dialogue systems*'; Carol Espy-Wilson, Jennifer Cole, Abeer Alwan, Louis Goldstein, Mary Harper, Elliot Saltzman and Mark Hasegawa-Johnson have arranged a session on '*New approaches to modeling variability for automatic speech recognition*'; Bruce Denby and Tanja Schultz have put together a session on '*Silent speech interfaces*'; Bjoern Schuller, Stefan Steidl and Anton Batliner have organised the '*INTERSPEECH 2009 emotion challenge*'; Anna Barney and Mette Pedersen have gathered together people interested in '*Advanced voice function assessment*'; Nick Campbell, Anton Nijholt, Joakim Gustafson and Carl Vogel are responsible for a session on '*Active listening and synchrony*'; and Mike Cohen, Johan Schalkwyk and Mike Phillips have organised a session on '*Lessons and challenges deploying voice search*'.

As well as the scientific programme, we have also arranged a series of social events and activities. Unfortunately, due to the sheer number of attendees, we had to abandon the idea of holding a *Party on the Pier*. Instead, we are very pleased to have found an excellent alternative, *Revelry at the Racecourse*, which is taking place high above the town with stunning seaward views for an evening of food and fun. Other events include the Welcome Reception at the Brighton Museum and Art Gallery, the Students' Reception at the stylish Italian Al Duomo restaurant, and the Reviewers' Reception at the amazing Royal Pavilion.

These organised social events are just a few of the opportunities that you will have to discover the delights of the local area. Brighton is a vibrant British seaside town, so I hope that you will enjoy exploring its many attractions and perhaps (weather permitting) the beach.

As I mentioned above, an event the size of INTERSPEECH simply cannot take place without the help and support of large a number of individuals, many of whom give their services freely despite the many other calls on their time. I would particularly like to thank Stephen Cox (University of East Anglia) for taking on the extremely time-consuming role of Technical Programme Chair and Valerie Hazan (University College London) for diligently looking after the most crucial aspect of the whole operation – the financial budget. In fact, due to the lack of underwriting by any particular institution, we were obliged to adopt a very different financial model for running INTERSPEECH this year. So I would also like to thank Valerie and Stephen for agreeing to join me in taking on these additional responsibilities, and ISCA for providing extra help with managing our cash flow.

I would also like to thank the rest of the organising committee: Anna Barney (University of Southampton) for putting together a fun social programme; Andy Breen (Nuance UK) for doing a tremendous job raising sponsorship in a very difficult financial climate; Shona D'Arcy (Trinity College, Dublin) for liaising with the students and organising the student helpers; Thomas Hain (University of Sheffield) for organising an impressive array of tutorials; Mark Huckvale (University College London) for doing a superb job as web master and for significantly upgrading the submission system; Philip Jackson (University of Surrey) for liaising with Hugh Loebner and organising the Loebner Prize competition; Peter Jancovic (University of Birmingham) for coordinating the different meeting room requirements; Denis Johnston (Haughgate Innovations) for helping with the sponsorship drive; Simon King (University of Edinburgh) for providing a contact point for the satellite workshops; Mike McTear (University of Ulster) for helping to smooth the registration process; Ben Milner (University of East Anglia) for organising the exhibition; Ji Ming (Queen's University Belfast) for coordinating the special sessions; Steve Renals (University of Edinburgh) for arranging the plenary sessions; Martin Russell (University of Birmingham) for looking after publicity and for producing a terrific poster; Maria Uther (Brunel University) for preparing the abstract book and conference proceedings; Simon Worgan (University of Sheffield) for organising the public outreach event and the speech-based competition; and Steve Young (University of Cambridge) for assisting with obtaining an opening speaker.

I would also like to thank the team at Meeting Makers – our Professional Conference Organisers – who have brought our vision into reality by providing valuable help and advice along the way.

I would particularly like to thank our generous sponsors, without whose support it would have been very difficult to mount the event. When we started this process in 2005, who could have envisaged the dire financial situation faced by the world's economies and banking systems in 2009? It is a great relief to us that our sponsors have found the means to provide us with support in these difficult times. We are especially grateful to Brighton & Hove City Council for their subvention towards the costs of the conference centre and to Visit Brighton for their encouragement in bringing a large conference such as INTERSPEECH to the UK.

Finally, I would like to thank everyone who submitted a paper to this year's conference (whether they were successful or not), the Reviewers for diligently evaluating them, the Area Chairs for putting together a varied and high-quality programme, and the Session Chairs and Student Helpers for ensuring a smooth running of the event itself.

I do hope that you will have an enjoyable and productive time here in Brighton and that you will leave with fond memories of INTERSPEECH-2009 – the place, the people, and the scientific exchanges you engaged in while you were here.

With best wishes for a successful conference.

Roger Moore, Conference Chair, Interspeech 2009

Interspeech 2009 Information

Venue

Interspeech 2009 will take place in the Brighton Centre. The Brighton Centre is located on the King's Road, about a 10 minute walk from Brighton station.

All keynote talks are in the Main Hall. See conference programme for venues of other sessions.

Registration

The Conference Registration Desks are located in the main foyer. For registration or any administrative issues please enquire at the Conference Registration Desks. These desks will be open at the following times:

Tutorial registration will be open on Sunday 6 September from 0830 hours to 1430 hours.

Conference registration will be open at the following times:

Sunday 6 September	1400 – 1800 hours
Monday 7 September	0900 – 1700 hours
Tuesday 8 September	0800 – 1700 hours
Wednesday 9 September	0800 – 1700 hours
Thursday 10 September	0800 – 1700 hours

The full registration package includes:

- Entry to all conference sessions (excluding satellite workshops and tutorial sessions)
- Conference bag containing:
 - Abstract book and conference programme
 - CD-ROM of Conference Proceedings
 - Promotional material
- Welcome Reception at Brighton Museum (Monday 7th September)
- Revelry at the Racecourse (Wednesday 9th September)
- Coffee breaks as per programme
- Badge

Badge

Your name badge, issued to you when you register, must be worn to all conference sessions and social events for identification and security purposes.

Non-Smoking event

Smoking is not permitted anywhere inside the Conference Centre.

Language

The official language of Interspeech 2009 is English.

Internet access

Wi-fi access is available throughout the Brighton Centre. Internet access is also provided on allocated PCs in the Rainbow Room on the Ground Floor and will be open from Monday to Thursday during conference open hours. The following website provides details of Wireless access points in Brighton http://www.brighton.co.uk/Wireless_Hotspots/.

Speaker Preparation Room

The Speaker Preparation Room is located in the Sunrise Room. If you are presenting an oral paper, you must load your presentation onto the central fileserver and check that it displays correctly well before your talk.

We recommend you do this well in advance of your presentation and certainly no later than two hours beforehand.

The room will be open during the following times:

Sunday 6th September 1400-1800 hours

Monday 7th September 0900-1700 hours

Tuesday 8th September 0830-1700 hours

Wednesday 9th September 0830-1700 hours

Thursday 10th September 0830-1700 hours

Coffee Breaks

These are scheduled according to the programme and will be served in the Hewison Hall and foyers. All those with special dietary requirements should make themselves known to Centre Staff who will provide alternative catering.

Lunch Breaks

Lunch is not included as part of your registration. There are several cafes and restaurants in Brighton from which you can purchase lunch.

Insurance

Registration fees do not include personal, travel or medical insurance of any kind. Delegates are advised when registering for the conference and booking travel that a travel insurance policy should be taken out to cover risks including (but not limited to) loss, cancellation, medical costs and injury. Interspeech 2009 and/or the conference organisers will not accept any responsibility for any delegate failing to insure.

General Information and the City of Brighton

Getting There

Rail

Brighton is under an hour by rail from London Victoria station, with 2 services every hour. There are also regular services from many other points, including a direct service from St Pancras station, connecting with Eurostar, and a link to Gatwick and Luton airports. More information can be found at www.nationalrail.co.uk or www.firstcapitalconnect.co.uk

Road

Brighton is about 45 minutes from the M25 London orbital down the M23 motorway, and 30 minutes from Gatwick airport.

Coach

Regular services to Brighton depart from London, Heathrow and Gatwick airports, and many other locations in the U.K. See www.nationalexpress.com for further information.

Air

Brighton is just 30 minutes by road or rail from London Gatwick International Airport and 90 minutes by road from London Heathrow. There are fast coach links between Heathrow, Gatwick and Brighton.

Sea

Brighton is 20 minutes by road or 25 minutes by rail from the port of Newhaven where a ferry service operates to Dieppe. See www.aferry.com/visitbrighton/ for more details.

Car Rental

Major car rental companies have offices located at the major airports. To drive in the U.K. you must have a current driver's license. Note that cars travel on the left side of the road in the U.K.

Brighton

Brighton is one of the most colourful, vibrant and creative cities in Europe. It has a very cosmopolitan flavour that is compact, energetic, unique, fun, lively, historic and free-spirited. Nestling between the South Downs & the sea on the south coast, Brighton offers everything from Regency heritage to beachfront cafes and a lively nightlife. It is a fantastic mix of iconic attractions, award winning restaurants, funky arts, cultural festivals and events.

Time Zone

Brighton and the U.K. in general are in British Summer Time (BST) at the time of the conference, BST = UTC (Greenwich) + 1.

Money and Credit Cards

Currency is the British Pound (GBP). Major international credit and charge cards such as Visa, American Express and MasterCard are widely accepted at retail outlets. Travellers' cheques are also widely accepted and can be cashed at banks, airports and major hotels.

Transportation in Brighton

Brighton and Hove is so compact, that once you're here, you might find it easiest to explore the city on foot. A frequent bus service also runs, costing £1.80 for a standard ticket across the city or various day tickets exist for £3.60 or £4.50 depending on the zone covered. Brighton is also one of five nationally selected 'cycling demonstration towns', and bicycles may be hired in a number of shops.

Accommodation

The Brighton area has an abundance of different kinds of accommodation from budget to luxury hotels. We recommend the Visit Brighton website to find accommodation to suit your needs:

<http://www.visitbrighton.com/site/accommodation>

Electrical Voltage

The electrical supply in the U.K. is 230-240 volts, AC 50Hz. The U.K. three-pin power outlet is different from that in many other countries, so you will need an adaptor socket. If your appliances are 110-130 volts you will need a voltage converter. Universal outlet adaptors for both 240V and 110 V appliances are sometimes available in leading hotels.

Tipping

There are no hard and fast rules for tipping in the U.K. If you are happy with the service, a 10-15% tip is customary, particularly in a restaurant or café with table service. Tipping in bars is not expected. For taxi fares, it is usual to round up to the nearest pound (£).

Non-Smoking Policy

The UK smoking laws prevent smoking in enclosed public spaces. It is therefore not acceptable to smoke in restaurants, bars and other public venues. There may be designated smoking areas.

Shop Opening Hours

Shopping hours tend to be from 1000 – 1800 hours with late opening until 2000 hours on a Thursday.

Emergencies

Please dial 999 for Fire, Ambulance or Police emergency services. The European emergency number 112 may also be used.

Medical Assistance

In the case of medical emergencies, there is a 24 hour Accident & Emergency department at the Royal Sussex County Hospital, Eastern Road, BN2 5BF - 01273 696955 (For ambulances, telephone 999).

The Doctor's Surgery for temporary residents and visitors is:

Chilvers McCrea Medical Centre, 1st Floor Boots the Chemist, 129 North Street, Brighton, 01273 328080. Open: Mon - Friday 08.00 - 18.00 Saturday 09.00 - 13.00 Closed on Sundays.

For emergency dental treatment (out of hours) call the Brighton & Hove health authority on 01273 486444 (lines open weekdays 06.30 - 21.30, weekends and bank holidays 09.00 - 12.30)

Pharmacies

The following pharmacies open later than normal working hours:

Ashtons: 98 Dyke Road, Seven Dials, Brighton, BN1 3JD, 01273 325020

Open: Mon - Sunday 09.00 - 22.00 except for 25 December

Asda: Brighton Marina, BN2 5UT, 01273 688019

Open: Mon - Thursday and Saturday: 09.00 - 20.00 Friday: 09.00 - 21.00 Sunday: 10.00 - 16.00
(times vary on Bank Holidays)

Asda: Crowhurst Road, Brighton, BN1 8AS, 01273 542314

Open: Mon - Thursday & Saturday: 09.00 - 20.00 Friday: 09.00 - 21.00 Sunday: 11.00 - 17.00
(times vary on Bank Holidays)

Westons: 6-7 Coombe Terrace, Lewes Road, Brighton, BN2 4AD, 01273 605354

Open: Mon - Sunday 09.00 - 22.00

Places of Interest

- The Royal Pavilion – the seaside palace of the Prince Regent (George IV) – www.royalpavilion.org.uk.
- Brighton Walking Tours – Hear about Brighton's history and discover interesting facts by downloading www.coolcitywalks.com/brighton/index.php onto your mp3 player.
- Brighton Pier – enjoy a typical day at the seaside on Brighton Pier. Enjoy the funfair rides, enjoy traditional seaside treats like candyfloss or a stick of rock or hire a deckchair and relax.
- Enjoy a traditional afternoon tea at the Grand hotel – www.grandbrighton.co.uk
- Take a 45 minute pleasure cruise along the coast and past the 2 piers or join a 90 minute mackerel fishing trip – www.watertours.co.uk.

Dining and Entertainment

There are numerous cafes and restaurants in the Brighton area, particularly around the 'The Lanes' and 'North Laine' areas.

Interspeech 2009 Social Programme

Monday 7th September, Welcome Reception

Everyone is invited to the *Welcome Reception* to be held at the Brighton Museum and Art Gallery, part of the historic Brighton Pavilion Estate. *Event starts at 19:30.*

Tuesday 8th September, Reviewers Reception

A reception for our hardworking *reviewers* will take place at the Royal Pavilion, the spectacular seaside palace of the Prince Regent (King George IV) transformed by John Nash between 1815 and 1822 into one of the most dazzling and exotic buildings in the British Isles. *Event starts at 19:30.* Admission by ticket enclosed with conference pack.

Tuesday 8th September, Student delegates drinks reception

Also on Tuesday evening there will be a drinks reception for *student delegates* at the stylish Italian Al Duomo Restaurant in the heart of the town. *Event starts at 19:30.* Admission by ticket enclosed with conference pack.

Wednesday 9th September, Revelry at the Racecourse

Look out for *Revelry at the Racecourse!* The Brighton Racecourse, set high on the Sussex Downs with stunning views of Brighton and Hove, will host the Conference Dinner and Party. No live horses, but expect a conference event with a difference and plenty of fun. Everyone is welcome and the *event will start at 19:30.* Transport to and from the event will be provided.

Interspeech 2009 Organisers

Conference Committee

Role	Name	Institution
General Chair	Prof. Roger Moore	University of Sheffield
Technical Programme	Prof. Stephen J. Cox	University of East Anglia
Finance	Prof. Valerie Hazan	University College London
Publications	Dr. Maria Uther	Brunel University
Web Master	Dr. Mark Huckvale	University College London
Plenary Sessions	Prof. Steve Renals	University of Edinburgh
Tutorials	Dr. Thomas Hain	University of Sheffield
Special Sessions	Dr. Ji Ming	Queen's University Belfast
Satellite Workshops	Dr. Simon King	University of Edinburgh
Sponsorship	Dr. Andrew Breen	Nuance UK
Exhibits	Dr. Ben Milner	University of East Anglia
Publicity	Prof. Martin Russell	University of Birmingham
Social Programme	Dr. Anna Barney	University of Southampton
Industrial Liaison	Mr. Denis Johnston	Haughgate Innovations
Advisor	Prof. Steve Young	University of Cambridge
Student Liaison	Dr. Shona D'Arcy	Trinity College, Dublin
Public Outreach	Dr. Simon Worgan	University of Sheffield
Registration	Prof. Michael McTear	University of Ulster
Loebner Contest	Dr. Philip Jackson	University of Surrey
Meeting Rooms	Dr. Peter Jancovic	University of Birmingham

Conference Organisers

 *Meeting Makers Ltd*
76 Southbrae Drive
Glasgow G13 1PP
Tel: +44 (0) 141 434 1500
Fax: +44 (0) 141 434 1519
interspeech@meetingmakers.co.uk

Scientific Committee

The conference organisers are indebted to the following people who made such a large contribution to the creation of the technical programme.

Technical Chair

- Stephen Cox, University of East Anglia

Area Coordinators

1. Aladdin Ariyaeinia, University of Hertfordshire
2. Nick Campbell, Trinity College Dublin
3. Mark J. F. Gales, Cambridge University
4. Yoshi Gotoh, Sheffield University
5. Phil Green, Sheffield University
6. Valerie Hazan, University College London
7. Wendy Holmes, Aurix Ltd
8. David House, KTH Stockholm
9. Kate Knill, Toshiba Research Europe Ltd
10. Bernd Möbius, Stuttgart
11. Sebastian Möller, Deutsche Telekom Laboratories
12. Satoshi Nakamura, NICT/ATR
13. Kuldip Paliwal, Griffith University
14. Gerasimos Potamianos, Athens
15. Ralf Schlüter, RWTH Aachen University
16. Tanja Schultz, Carnegie Mellon University
17. Yannis Stylianou, University of Crete
18. Marc Swerts, Tilburg University
19. Isabel Trancoso, INESC-ID Lisboa / IST
20. Saeed Vaseghi, Brunel University
21. Yi Xu, University College London
22. Bayya Yegnanarayana, International Institute of Information Technology Hyderabad
23. Kai Yu, Cambridge University
24. Heiga Zen, Toshiba Research Ltd.

Special Session Organisers

- Abeer Alwan, UCLA
- Anna Barney, ISVR, University of Southampton
- Anton Batliner, FAU Erlangen-Nuremberg
- Nick Campbell, Trinity College Dublin
- Mike Cohen, Google
- Jennifer Cole, Illinois
- Bruce Denby, Université Pierre et Marie Curie
- Carol Espy-Wilson, University of Maryland
- Louis Goldstein, University of Southern California
- Joakim Gustafson, KTH Stockholm
- Mary Harper, University of Maryland
- Mark Hasegawa-Johnson, Illinois
- Daniel Hirst, Université de Provence
- Oliver Lemon, Organiser Edinburgh University
- Anton Nijholt, Twente
- Mette Pedersen, Medical Centre, Voice Unit, Denmark
- Mike Phillips, Vlingo
- Olivier Pietquin, IMS Research Group
- Elliot Saltzman, Haskins Laboratories
- Johan Schalkwyk, Google
- Björn Schuller, Technische Universität München
- Tanja Schultz, Carnegie Mellon University
- Stefan Steidl, FAU Erlangen-Nuremberg
- Carl Vogel, Trinity College Dublin

Scientific Reviewers

Alberto Abad
Sherif Abdou
Alex Acero
Andre Gustavo Adami
Gilles Adda
Martine Adda-Decker
Masato Akagi
Masami Akamine
Murat Akbacak
Jan Alexandersson
Paavo Alku
Abeer Alwan
Eliathamby Ambikairajah
Noam Amir
Ove Andersen
Tim Anderson
Elisabeth Andre
Walter Andrews
Takayuki Arai
Masahiro Araki
Aladdin Ariyaeinia
Victoria Arranz
Bishnu Atal
Roland Auckenthaler
Cinzia Avesani
Matthew Aylett
Harald Baayen
Michiel Bacchiani
Pierre Badin
Janet Baker
Srinivas Bangalore
Plinio Barbosa
Etienne Barnard
Anna Barney
Dante Barone
William J. Barry
Anton Batliner
Frederic Bechet
Steve Beet
Homayoon Beigi
Nuno Beires
Jerome Bellegarda
Jose Miguel Benedi Ruiz
M. Carmen Benitez Ortuzar
Nicole Beringer
Kay Berkling
Laurent Besacier
Frédéric Bettens
Frederic Bimbot
Maximilian Bisani
Judith Bishop
Alan Black
Mats Blomberg
Gerrit Bloothoof
Antonio Bonafonte
Jean-Francois Bonastre
Zinny Bond
Helene Bonneau-Maynard

Herve Boulard
Lou Boves
Daniela Braga
Bettina Braun
Catherine Breslin
John Bridle
Mirjam Broersma
Niko Brummer
Udo Bub
Luis Buera
Tim Bunnell
Anthony Burkitt
Denis Burnham
Bill Byrne
Jose R. Calvo de Lara
Joseph P. Campbell
Nick Campbell
William Campbell
Antonio Cardenal Lopez
Valentin Cardenoso-Payo
Michael Carey
Rolf Carlson
Maria Jose Castro Bleda
Mauro Cettolo
Joyce Chai
Chandra Sekhar Chellu
Fang Chen
Yan-Ming Cheng
Rathi Chengalvarayan
Jo Cheolwoo
Jen-Tzung Chien
KK Chin
Gerard Chollet
Khalid Choukri
Heidi Christensen
Hyunsong Chung
Robert Clark
Mike Cohen
Luisa Coheur
Jennifer Cole
Alistair Conkie
Martin Cooke
Ricardo de Cordoba
Piero Cosi
Bob Cowan
Felicity Cox
Catia Cucchiarini
Fred Cummins
Francesco Cutugno
Joachim Dale
Paul Dalsgaard
Geraldine Damnati
Morena Danieli
Marelie Davel
Chris Davis
Amedeo De Dominicis
Angel de la Torre Vega
Renato De Mori

Carme de-la-Mota Gorriz
David Dean
Michael Deisher
Grazyna Demenko
Kris Demuynck
Bruce Denby
Matthias Denecke
Li Deng
Giuseppe Di Fabbrizio
Vassilis Digalakis
Christine Doran
Ellen Douglas-Cowie
Christoph Draxler
Jasha Droppo
Andrzej Drygajlo
Jacques Duchateau
Christophe D`Alessandro
Mariapaola D`Imperio
Kjell Elenius
Daniel Ellis
Ahmad Emami
Julien Epps
Anders Eriksson
Mirjam Ernestus
David Escudero Mancebo
Carol Espy-Wilson
Sascha Fagel
Daniele Falavigna
Mauro Falcone
Isabel Fale
Kevin Farrell
Marcos Faundez-Zanuy
Marcello Federico
Junlan Feng
Javier Ferreiros
Carlos A. Ferrer Riesgo
Tim Fingscheidt
Janet Fletcher
José A. R. Fonollosa
Kate Forbes-Riley
Eric Fosler-Lussier
Diamantino Freitas
Juergen Fritsch
Sonia Frota
Christian Fuegen
Olac Fuentes
Hiroya Fujisaki
Toshiaki Fukada
Sadaoki Furui
Mark J. F. Gales
Ascensiòn Gallardo Antolin
Yuqing Gao
Fernando Garcia Granada
Carmen Garcia Mateo
Mária Gósy
Panayiotis Georgiou
Dafydd Gibbon
Mazin Gilbert

Juan Ignacio Godino Llorente
Simon Godsill Godsill
Roland Goecke
Pedro Gomez Vilda
Joaquin Gonzalez-Rodriguez
Allen Gorin
Martijn Goudbeek
Philippe Gournay
Bjorn Granstrom
Agustin Gravano
David Grayden
Phil Green
Rodrigo C Guido
Susan Guion
Ellen Haas
Kadri Hacioglu
Reinhold Haeb-Umbach
Jari Hagqvist
Thomas Hain
John Hajek
Eva Hajicova
Dilek Hakkani-Tur
Gerhard Hanrieder
John H.L. Hansen
Mary Harper
Jonathan Harrington
John Harris
Naomi Harte
Mark Hasegawa-Johnson
Jean-Paul Haton
Valerie Hazan
Timothy J. Hazen
Harald Höge
Matthieu Hebert
Martin Heckmann
Per Hedelin
Peter Heeman
Paul Heisterkamp
John Henderson
Inmaculada Hernaez Rioja
Luis Alfonso Hernandez
Gomez
Javier Hernando
Wolfgang Hess
Lee Hetherington
Ulrich Heute
Keikichi Hirose
Hans Guenter Hirsch
Julia Hirschberg
Daniel Hirst
Wendy Holmes
W. Harvey Holmes
Chiori Hori
John-Paul Hosom
David House
Fei Huang
Qiang Huang
Isabel Hub Faria
Juan Huerta
Marijn Huijbregts
Melvyn Hunt
Lluis F. Hurtado Oliver

John Ingram
Shunichi Ishihara
Masato Ishizaki
Yoshiaki Itoh
Philip Jackson
Tae-Yeoub Jang
Esther Janse
Arne Jönsson
Alexandra Jesse
Luis Jesus
Qin Jin
Michael Johnston
Kristiina Jokinen
Caroline Jones
Szu-Chen Stan Jou
Denis Jouvét
Ho-Young Jung
Peter Kabal
Takehiko Kagoshima
Alexander Kain
Hong-Goo Kang
Stephan Kanthak
Hiroaki Kato
Tatsuya Kawahara
Hisashi Kawai
Mano Kazunori
Thomas Kemp
Patrick Kenny
Jong-mi Kim
Sunhee Kim
Byeongchang Kim
Hoirin Kim
Hong Kook Kim
Hyung Soon Kim
Jeesun Kim
Nam Soo Kim
Sanghun Kim
Simon King
Brian Kingsbury
Yuko Kinoshita
Christine Kitamura
Esther Klabbers
Dietrich Klakow
Bastiaan Kleijn
Kate Knill
Hanseok Ko
Takao Kobayashi
M. A. Kohler
George Kokkinakis
John Kominek
Myoung-Wan Koo
Sunil Kumar Kopparapu
Christian Kroos
Chul Hong Kwon
Hyuk-Chul Kwon
Oh-Wook Kwon
Francisco Lacerda
Pietro Laface
Unto K. Laine
Claude Lamblin
Lori Lamel
Kornel Laskowski

Javier Latorre
Gary Geunbae Lee
Chin Hui Lee
Lin-shan Lee
Oliver Lemon
Qi (Peter) Li
Haizhou Li
Carlos Lima
Georges Linares
Mike Lincoln
Bjorn Lindblom
Anders Lindström
Zhen-Hua Ling
Yang Liu
Yi Liu
Karen Livescu
Eduardo Lleida Solano
Joaquim Llisterri
Deborah Loakes
Maria Teresa Lopez Soto
Ramon Lopez-Cozar Delgado
Patrick Lucey
Changxue Ma
Ning Ma
Dusan Macho Cierna
Javier Macias-Guarasa
Abdullahsain E. Mahdi
Brian Mak
Robert Malkin
Nuno Mamede
Claudia Manfredi
Lidia Mangu
José B. Mariño Acebal
Ewin Marsi
Carlos David Martínez
Hinarejos
Jean-Pierre Martens
Rainer Martin
Alvin Martin
Enrique Masgrau Gomez
Sameer Maskey
John Mason
Takashi Masuko
Ana Isabel Mata
Tomoko Matsui
Karen Mattock
Bernd Möbius
Sebastian Möller
Christian Müller
Alan McCree
Erik McDermott
Gordon McIntyre
John McKenna
Michael McTear
Hugo Meinedo
Alexsandro Meireles
Carlos Eduardo Meneses
Ribeiro
Helen Meng
Florian Metzger
Bruce Millar
Ben Milner

Nobuaki Minematsu		Jose Luis Perez Cordoba	Michelina Savino
Wolfgang Minker		Pascal Perrier	Joan Andreu Sánchez Peiró
Holger Mitterer		Hartmut R. Pfitzinger	Thomas Schaaf
Hansjoerg Mixdorff		Mike Phillips	Johan Schalkwyk
Parham Mokhtari		Michael Picheny	Odette Scharenborg
Garry Molholt		Joe Picone	Ralf Schlüter
Juan Manuel Montero	Montero	Roberto Pieraccini	Jean Schoentgen
Martinez		Olivier Pietquin	Marc Schroeder
Asuncion Moreno		Ferran Pla Santamaría	Bjoern Schuller
Pedro J. Moreno		Aristodemos Pnevmatikakis	Michael Schuster
Nelson Morgan		Louis C.W. Pols	Reva Schwartz
Ronald Mueller		Alexandros Potamianos	Antje Schweitzer
Climent Nadeu		Gerasimos Potamianos	Encarnacion Segarra Soriano
Seiichi Nakagawa		David Powers	Jose Carlos Segura
Satoshi Nakamura		Rohit Prasad	Frank Seide
Shrikanth Narayanan		Mahadeva Prassanna	Mike Seltzer
Eva Navas		Kristin Precoda	D. Sen
Jiri Navratil		Patti Price	Stephanie Seneff
Ani Nenkova		Tarun Pruthi	Cheol Jae Seong
João Neto		Mark Przybocki	Antonio Serralheiro
Ron Netsell		Yao Qian	Kiyohiro Shikano
Sergio Netto		Zhu Qifeng	Jiyoung Shin
Hermann Ney		Thomas F. Quatieri	Koichi Shinoda
Patrick Nguyen		Raja Rajasekaran	Carlos Silva
Anton Nijholt		Nitendra Rajput	Olivier Siohan
Elmar Noeth		Bhuvana Ramabhadran	Malcolm Slaney
Albino Nogueiras		V. Ramasubramanian	Raymond Slyh
Michael Norris		Preeti Rao	Connie So
Mohaddeseh Nosratighods		Andreia S. Rauber	M. Mohan Sondhi
Regine Obrecht		Mosur Ravishankar	Victor Sorokin
John Ohala		Mario Refice	Dave Stallard
Sumio Ohno		Norbert Reithinger	Mark Steedman
Luis Oliveira		Steve Renals	Stefan Steidl
Peder Olsen		Fernando Gil Vianna Resende Jr.	Andreas Stergiou
Mohamed Omar		Douglas Reynolds	Richard Stern
Roeland Ordelman		Luca Rigazio	Mary Stevens
Rosemary Orr		Michael Riley	Helmer Strik
Alfonso Ortega Gimenez		Christian Ritz	Volker Strom
Javier Ortega-Garcia		Tony Robinson	Sebastian Stueker
Beatrice Oshika		Eduardo Rodriguez Banga	Matt Stuttle
Mari Ostendorf		Luis Javier Rodriguez-Fuentes	Yannis Stylianou
Douglas O`Shaughnessy		Richard Rose	Rafid Sukkar
Fernando S. Pacheco		Olivier Rosec	Torbjorn Svendsen
Tim Paek		Antti-Veikko Rosti	Marc Swerts
Vincent Pagel		Jean-Luc Rouas	Ann Syrdal
Sira Elena Palazuelos	Palazuelos	Antonio Rubio	David Talkin
Cagigas		Martin Russell	Zheng-Hua Tan
Kuldip Paliwal		Yoshinori Sagisaka	Yun Tang
Yue Pan		Josep M. Salavedra	Jianhua Tao
P.C. Pandey		K Samudravijaya	Carlos Teixeira
Jose Manuel Pardo		Rubén San-Segundo	Antonio Teixeira
Jun Park		Victoria Eugenia Sanchez Calle	Joao Paulo Teixeira
Jong Cheol Park		Emilio Sanchis	Louis ten Bosch
Patrick Paroubek		Eric Sanders	Jacques Terken
Steffen Pauws		George Saon	Barry-John Theobald
Mette Pedersen		Shimon Sapir	William Thorpe
Antonio M. Peinado		Murat Saraclar	Jilei Tian
Carmen Pelaez-Moreno		Ruhi Sarikaya	Michael Tjalve
Jason Pelecanos		Hiroshi Saruwatari	Tomoki Toda
Bryan Pellom		Antonio Satue Villar	Roberto Togneri
Christina Pentland			Keiichi Tokuda
Fernando Perdigao			Doroteo T Toledano

Laura Tomokiyo
María Inés Torres Barañano
Arthur Toth
Dat Tran
Isabel Trancoso
David Traum
Kimiko Tsukada
Roger Tucker
Gokhan Tur
L. Alfonso Urena Lopez
Jacqueline Vaissiere
Dirk Van Compernelle
Henk van den Heuvel
Jan van Doorn
Hugo Van hamme
Arjan van Hessen
Roeland van Hout
David van Leeuwen
Jan van Santen
Rob Van Son
Peter Vary
Saeed Vaseghi
Mario Vayra
Werner Verhelst
Jo Verhoeven
Ceu Viana

Marina Vigario
Fábio Violaro
Carlos Enrique Vivaracho
Pascual
Robbie Vogt
Julie Vonwiller
Michael Wagner
Marilyn Walker
Patrick Wambacq
Hsiao-Chuan Wang
Kuansan Wang
Ye-Yi Wang
Chao Wang
Hsin-min Wang
Nigel Ward
Catherine Watson
Christian Wellekens
Stanley Wenndt
Stefan Werner
Yorick Wilks
Daniel Willett
Briony Williams
Monika Woszczyna
Johan Wouters
Stuart Wrigley
John Zhiyong Wu

Yi-Jian Wu
Bing Xiang
Yi Xu
Haitian Xu
Junichi Yamagashi
Bayya Yegnanarayana
Nestor Becerra Yoma
Chang D. Yoo
Dongsuk Yook
Steve Young
Roger (Peng) Yu
Dong Yu
Kai Yu
Young-Sun Yun
Milos Zelezny
Heiga Zen
Andrej Zgank
Tong Zhang
Yunxin Zhao
Jing Zheng
Bowen Zhou
Imed Zitouni
Udo Zoelzer
Geoffrey Zweig

Sponsors and Exhibitors

Sponsors

We gratefully acknowledge the support of:



Google, Inc.

for silver-level sponsorship of the conference.

Carstens Medizinelektronik GmbH

for bronze-level sponsorship of the conference.



Northern Digital, Inc.

for bronze-level sponsorship of the conference.

Nuance, Inc.

for bronze-level sponsorship of the conference.



Toshiba Research Europe

for bronze-level sponsorship of the conference.



Appen Pty Ltd

for bronze-level sponsorship of the conference.



Crown Industries, Inc

Crown Industries, Inc.

for supporting the Loebner Prize competition.

IBM Research Inc.

for supporting the Loebner Prize competition.

IBM Research



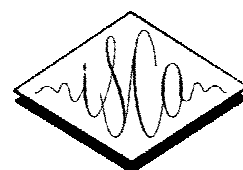
Brighton & Hove

Brighton & Hove City Council

for sponsoring the conference venue.

Exhibitors

Interspeech 2009 welcomes the following exhibitors:



Supporting Committee Institutions

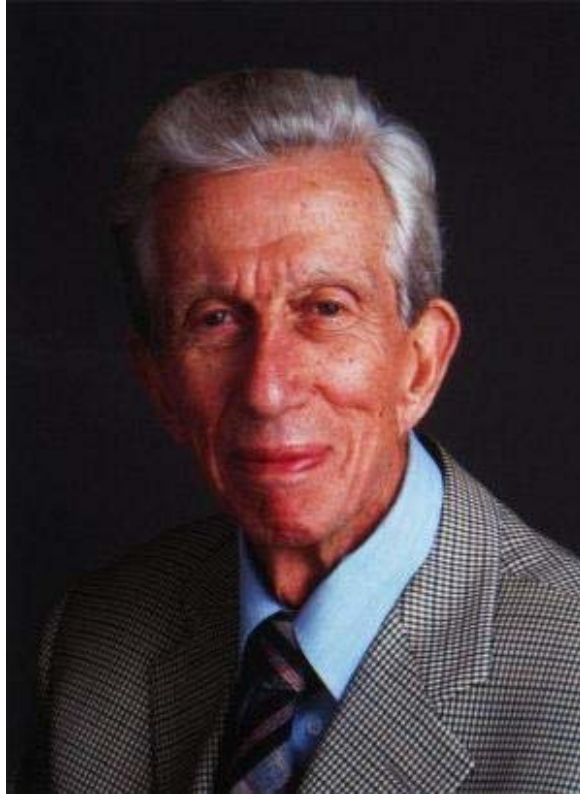
The organising committee would like to acknowledge the support of their respective institutions (institutions presented in alphabetical order).



Trinity College Dublin



In Memoriam



Gunnar Fant

8th October 1919 – 6th June 2009

Speech research pioneer and recipient of the 1989
ESCA Medal for scientific achievement

Contest for the Loebner Prize 2009

Time: 10:45am Sunday 6 September
Venue: Rainbow room, Brighton Centre

The Loebner Prize for artificial intelligence is the first formal instantiation of a Turing Test. The test is named after Alan Turing the brilliant British mathematician with many accomplishments in computing science. In 1950, in the article *Computing Machinery and Intelligence* which appeared in the philosophy journal *Mind*, Alan Turing asked the question "Can a Machine Think?" He answered in the affirmative, but a central question was: "If a computer could think, how could we tell?" Turing's suggestion was, that if the responses from a computer in an imitation game were indistinguishable from that of a human, the computer could be said to be thinking.

The Loebner prize competition seeks to find out how close we are to building a computer to pass the Turing test. In 1950 Alan Turing wrote:

"I believe that in about fifty years' time it will be possible, to programme computers, with a storage capacity of about 10^9 , to make them play the imitation game so well that an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of questioning..."

The 2009 Loebner Prize will operate in the following manner.

- Panels of judges communicate with two entities over a typewritten link. One entity is a human, one is a computer program, allocated at random.
- Each judge will begin the round by making an initial comment to the first entity and continue interacting for 5 minutes. At the conclusion of the five minutes, the judge will begin the interaction with the second entity and continue for 5 minutes.
- Entities will be expected to respond to the judges' initial comments or questions. There will be no restrictions on what names etc the entries, humans, or judges can use, nor any other restrictions on the content of the conversations.
- At the conclusion of the 10 minutes of questioning, judges will be allowed 10 minutes to review the conversations. They will then score one of the two entities as the human. Following this, there will be a 5 minute period for judges and confederates to take their places for the next round.
- The system that is most often considered to be human by the judges will win a Bronze Loebner medal and \$3000.



More details at the Loebner Prize web site: <http://www.loebner.net/Prizef/loebner-prize.html>.

The Loebner Prize is made possible by funding from Crown Industries, Inc., of East Orange NJ and contributions from IBM research.

Organiser: Philip Jackson, p.jackson@surrey.ac.uk



INTERSPEECH 2010

26-30 SEPTEMBER 2010, MAKUHARI, JAPAN



"Spoken Language Processing for All"

Spoken Language Processing for All Ages, Health Conditions,
Native Languages, and Environments

INTERSPEECH 2010, the 11th conference in the annual series of Interspeech events, will be held at the International Convention Hall at the Makuhari Messe exhibition complex in Chiba, Japan. The conference venue allows easy access for international travelers: 30 minutes from Narita International Airport by bus, 30 minutes from Tokyo station by train, and within walking distance of a number of hotels with a wide variety of room rates.

INTERSPEECH 2010 returns to Japan for the first time in 16 years. Japan hosted the first and third International Conferences on Spoken Language Processing (ICSLP) in 1990 and 1994. In 2010, we seek to emphasize the interdisciplinary nature of speech research, and facilitate cross-fertilization among various branches of spoken language science and technology.

Mark 26-30 September 2010 on your calendar now!

For further details,
visit <http://www.interspeech2010.org/> or
write to office@interspeech2010.org

General Chair:

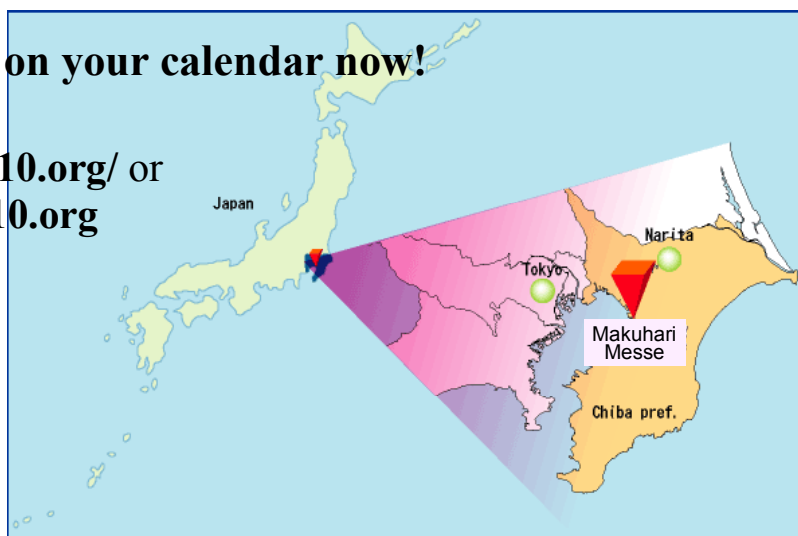
Keikichi Hirose
(The University of Tokyo)

General Co-Chair:

Yoshinori Sagisaka (Waseda University)

Technical Program Committee Chair:

Satoshi Nakamura (NICT/ATR)



Satellite Events

This is the list of satellite workshops linked to Interspeech 2009.

ACORNS Workshop on Computational Models of Language Evolution, Acquisition and Processing

11 September 2009, Brighton, UK.

The workshop brings together up to 50 scientists to discuss future research in language acquisition, processing and evolution. Deb Roy, Friedemann Pulvermüller, Rochelle Newman and Lou Boves will provide an overview of the state-of-art, a number of discussants from different disciplines will widen the perspective, and all participants can contribute to a roadmap.

AVSP 2009 - Audio-Visual Speech Processing

10-13 Sept 2009, University of East Anglia, Norwich, U.K.

The International Conference on Auditory-Visual Speech Processing (AVSP) attracts an interdisciplinary audience of psychologists, engineers, scientists and linguists, and considers a range of topics related to speech perception, production, recognition and synthesis. Recently the scope of AVSP has broadened to also include discussion on more general issues related to audiovisual communication. For example, the interplay between speech and the expressions of emotion, and the relationship between speech and manual gestures.

Blizzard Challenge Workshop

4 September, University of Edinburgh, U.K.

In order to better understand and compare research techniques in building corpus-based speech synthesizers on the same data, the Blizzard Challenge was devised. The basic challenge is to take the released speech database, build a synthetic voice from the data and synthesize a prescribed set of test sentences which are evaluated through listening tests. The results are presented at this workshop. Attendance at the 2009 workshop for the 4th Blizzard Challenge is open to all, not just participants in the challenge. Registration closes on 14th August 2009.

SIGDIAL - Special Interest Group on Dialogue

11-12 Sept 2009, Queen Mary University, London, U.K.

The SIGDIAL venue provides a regular forum for the presentation of cutting edge research in discourse and dialogue to both academic and industry researchers. The conference is sponsored by the SIGDIAL organization, which serves as the Special Interest Group in discourse and dialogue for both the Association for Computational Linguistics and the International Speech Communication Association.

SLaTE Workshop on Speech and Language Technology in Education (SLaTE)

3-5 September 2009, Wroxall, Warwickshire, U.K.

SLaTE 2009 follows SLaTE 2007, held in Farmington, Pennsylvania, USA, and the STiLL meeting organized by KTH in Marholmen, Sweden, in 1998. The workshop will address all topics which concern speech and language technology for education. Papers will discuss theories, applications, evaluation, limitations, persistent difficulties, general research tools and

techniques. Papers that critically evaluate approaches or processing strategies will be especially welcome, as will prototype demonstrations of real-world applications.

Young Researchers' Roundtable on Spoken Dialogue Systems

13-14 September 2009, Queen Mary, University of London, U.K.

The Young Researchers' Roundtable on Spoken Dialog Systems is an annual workshop designed for students, post docs, and junior researchers working in research related to spoken dialogue systems in both academia and industry. The roundtable provides an open forum where participants can discuss their research interests, current work and future plans. The workshop is meant to provide an interdisciplinary forum for creative thinking about current issues in spoken dialogue systems research, and help create a stronger international network of young researchers working in the field.

Interspeech 2009 Keynote Sessions

Keynote 1

ISCA Scientific Achievement Medallist for 2009

Sadaoki Furui,
Tokyo Institute of Technology

Selected topics from 40 years of research on speech and speaker recognition

Mon-Ses1-K: Monday 11:00, Main Hall

Chair: Isabel Trancoso

Abstract

This talk summarizes my 40 years research on speech and speaker recognition, focusing on selected topics that I have investigated at NTT Laboratories, Bell Laboratories and Tokyo Institute of Technology with my colleagues and students. These topics include: the importance of spectral dynamics in speech perception; speaker recognition methods using statistical features, cepstral features, and HMM/GMM; text-prompted speaker recognition; speech recognition by dynamic features; Japanese LVCSR; spontaneous speech corpus construction and analysis; spontaneous speech recognition; automatic speech summarization; WFST-based decoder development and its applications; and unsupervised model adaptation methods.

Presenter

Sadaoki Furui is currently a Professor at Tokyo Institute of Technology, Department of Computer Science. He is engaged in a wide range of research on speech analysis, speech recognition, speaker recognition, speech synthesis, and multimodal human-computer interaction and has authored or coauthored over 800 published articles. He is a Fellow of the IEEE, the International Speech Communication Association (ISCA), the Institute of Electronics, Information and Communication Engineers of Japan (IEICE), and the Acoustical Society of America. He has served as President of the Acoustical Society of Japan (ASJ) and the ISCA. He has served as a member of the Board of Governor of the IEEE Signal Processing (SP) Society and Editor-in-Chief of both the Transaction of the IEICE and the Journal of Speech Communication. He has received the Yonezawa Prize, the Paper Award and the Achievement Award from the IEICE (1975, 88, 93, 2003, 2003, 2008), and the Sato Paper Award from the ASJ (1985, 87). He has received the Senior Award and Society Award from the IEEE SP Society (1989, 2006), the Achievement Award from the Minister of Science and Technology and the Minister of Education, Japan (1989, 2006), and the Purple Ribbon Medal from Japanese Emperor (2006). In 1993 he served as an IEEE SPS Distinguished Lecturer.

Keynote 2

Tom Griffiths,
UC Berkeley

Connecting human and machine learning via probabilistic models of cognition

Tue-Ses0-K: Tuesday 08:30, Main Hall
Chair: Steve Renals

Abstract

Human performance defines the standard that machine learning systems aspire to in many areas, including learning language. This suggests that studying human cognition may be a good way to develop better learning algorithms, as well as providing basic insights into how the human mind works. However, in order for ideas to flow easily from cognitive science to computer science and vice versa, we need a common framework for describing human and machine learning. I will summarize recent work exploring the hypothesis that probabilistic models of cognition, which view learning as a form of statistical inference, provide such a framework, including results that illustrate how novel ideas from statistics can inform cognitive science. Specifically, I will talk about how probabilistic models can be used to identify the assumptions of learners, learn at different levels of abstraction, and link the inductive biases of individuals to cultural universals.

Presenter

Tom Griffiths is an Assistant Professor of Psychology and Cognitive Science at UC Berkeley, with courtesy appointments in Computer Science and Neuroscience. His research explores connections between human and machine learning, using ideas from statistics and artificial intelligence to try to understand how people solve the challenging computational problems they encounter in everyday life. He received his PhD in Psychology from Stanford University in 2005, and taught in the Department of Cognitive and Linguistic Sciences at Brown University before moving to Berkeley. His work and that of his students has received awards from the Neural Information Processing Systems conference and the Annual Conference of the Cognitive Science Society, and in 2006 IEEE Intelligent Systems magazine named him one of "Ten to watch in AI."

Keynote 3

Deb Roy,
MIT Media Lab

New Horizons in the Study of Language Development

Wed-Ses0-K: Wednesday 08:30, Main Hall
Chair: Roger Moore

Abstract

Emerging forms of ecologically-valid longitudinal recordings of human behavior and social interaction promise fresh perspectives on age-old questions of child development. In a pilot effort, 240,000 hours of audio and video recordings of one child's life at home are being analyzed with a focus on language development. To study a corpus of this scale and richness, current methods of developmental sciences are insufficient. New data analysis algorithms and methods for interpretation and computational modeling are under development. Preliminary speech analysis reveals surprising levels of linguistic "finetuning" by caregivers that may provide crucial support for word learning. Ongoing analysis of various other aspects of the corpus aim to model detailed aspects of the child's language development as a function of learning mechanisms combined with everyday experience. Plans to collect similar corpora from more children based on a streamlined recording system are underway.

Presenter

Deb Roy directs the Media Lab's Cognitive Machines group, is founding director of MIT's Center for Future Banking, and chairs the academic program in Media Arts and Sciences. A native of Canada, he received his bachelor of computer engineering from the University of Waterloo in 1992 and his PhD in cognitive science from MIT in 1999. He joined the MIT faculty in 2000 and was named AT&T Associate Professorship of Media Arts and Sciences in 2003.

Roy studies how children learn language, and designs machines that learn to communicate in human-like ways. To enable this work, he has developed new data-driven methods for analyzing and modeling human linguistic and social behavior. He has begun exploring applications of these methods to a range of new domains from financial behavior to autism. Roy has authored numerous scientific papers in the areas of artificial intelligence, cognitive modeling, human-machine interaction, data mining and information visualization.

Keynote 4

Mari Ostendorf,
University of Washington

Transcribing Speech for Spoken Language Processing

Thu-Ses0-K: Thursday 08:30, Main Hall

Chair: Martin Russell

Abstract

As storage costs drop and bandwidth increases, there has been a rapid growth of spoken information available via the web or in online archives -- including radio and TV broadcasts, oral histories, legislative proceedings, call center recordings, etc. -- raising problems of document retrieval, information extraction, summarization and translation for spoken language. While there is a long tradition of research in these technologies for text, new challenges arise when moving from written to spoken language. In this talk, we look at differences between speech and text, and how we can leverage the information in the speech signal beyond the words to provide structural information in a rich, automatically generated transcript that better serves language processing applications. In particular, we look at three interrelated types of structure (segmentation, prominence and syntax), methods for automatic detection, the benefit of optimizing rich transcription for the target language processing task, and the impact of this structural information in tasks such as parsing, topic detection, information extraction and translation.

Presenter

Mari Ostendorf received the Ph.D. in electrical engineering from Stanford University. After working at BBN Laboratories and Boston University, she joined the University of Washington (UW) in 1999. She has also been a visiting researcher at the ATR Interpreting Telecommunications Laboratory and at the University of Karlsruhe. At UW, she is currently an Endowed Professor of System Design Methodologies in Electrical Engineering and an Adjunct Professor in Computer Science and Engineering and in Linguistics. Currently, she is the Associate Dean for Research and Graduate Studies in the UW College of Engineering. She teaches undergraduate and graduate courses in signal processing and statistical learning, including a design-oriented freshman course that introduces students to signal processing and communications.

Prof. Ostendorf's research interests are in dynamic and linguistically-motivated statistical models for speech and language processing. Her work has resulted in over 200 publications and 2 paper awards. Prof. Ostendorf has served as co-Editor of *Computer Speech and Language*, as the Editor-in-Chief of the *IEEE Transactions on Audio, Speech and Language Processing*, and is currently on the IEEE Signal Processing Society Board of Governors and the ISCA Advisory Council. She is a Fellow of IEEE and ISCA.

Interspeech 2009 Special Sessions

The Interspeech 2009 Organisation Committee is pleased to announce acceptance of the following Special Sessions at Interspeech 2009.

INTERSPEECH 2009 Emotion Challenge

Mon-Ses2-S1: Monday 13:30

Place: Ainsworth (East Wing 4)

The INTERSPEECH 2009 Emotion Challenge aims to help bridge the gap between the excellent research on human emotion recognition from speech and the low compatibility of results. The FAU Aibo Emotion Corpus of spontaneous, emotionally coloured speech, and benchmark results of the two most popular approaches will be provided by the organisers. This consists of nine hours of speech from 51 children, recorded at two different schools. This corpus allows for distinct definition of test and training partitions incorporating speaker independence as needed in most real-life settings. The corpus further provides a uniquely detailed transcription of spoken content with word boundaries, non-linguistic vocalisations, emotion labels, units of analysis, etc. The results of the Challenge will be presented at the Special Session and prizes will be awarded to the sub-challenge winners and a best paper.

Organisers: Bjoern Schuller (schuller@tum.de), Technische Universitaet Muenchen, Germany, Stefan Steidl (steidl@informatik.uni-erlangen.de), FAU Erlangen-Nuremberg, Germany, Anton Batliner (batliner@informatik.uni-erlangen.de), FAU Erlangen-Nuremberg, Germany.

Silent Speech Interfaces

Mon-Ses3-S1: Monday 16:00

Place: Ainsworth (East Wing 4)

A Silent Speech Interface (SSI) is an electronic system enabling speech communication to take place without the necessity of emitting an audible acoustic signal. By acquiring sensor data from elements of the human speech production process – from the articulators, their neural pathways, or the brain itself – an SSI produces a digital representation of speech which can be synthesized directly, interpreted as data, or routed into a communications network. Due to this novel approach Silent Speech Interfaces have the potential to overcome the major limitations of traditional speech interfaces today, i.e. (a) limited robustness in the presence of ambient noise; (b) lack of secure transmission of private and confidential information; and (c) disturbance of bystanders created by audibly spoken speech in quiet environments; while at the same time retaining speech as the most natural human communication modality. The special session intends to bring together researchers in the areas of human articulation, speech and language technologies, data acquisition and signal processing, as well as in human interface design, software engineering and systems integration. Its goal is to promote the exchange of ideas on current SSI challenges and to discuss solutions, by highlighting, for each of the technological approaches presented, its range of applications, key advantages, potential drawbacks, and current state of development.

Organisers: Bruce Denby (denby@ieee.org), Université Pierre et Marie Curie, France, Tanja Schultz (tanja@ira.uka.de), Cognitive Systems Lab, University of Karlsruhe, Germany.

Advanced Voice Function Assessment

Tue-Ses1-S1: Tuesday 10:00

Place: Ainsworth (East Wing 4)

In order to advance the field of voice function assessment in a clinical setting, cooperation between clinicians and technologists is essential. The aim of this special session is to showcase

work that crosses the borders between basic, applied and clinical research and highlights the development of partnership between technologists and healthcare professionals in advancing the protocols and technologies for the assessment of voice function.

Organisers: Anna Barney (ab3@soton.ac.uk), Institute of Sound and Vibration Research, UK, Mette Pedersen (m.f.pedersen@dadlnet.dk), Medical Centre, Voice Unit, Denmark.

Measuring the Rhythm of Speech

Tue-Ses3-S2: Tuesday 16:00

Place: Ainsworth (East Wing 4)

There has been considerable interest in the last decade in the modelling of rhythm both from a typological perspective (e.g. establishing objective criteria for classifying languages or dialect as stress timed, syllable timed or mora timed) and from the perspective of establishing evaluation metrics of non standard or deviant varieties of speech such as that obtained from non-native speakers, from speakers with pathological disabilities or from automatic speech synthesis. The aim of this special session will be to bring together a number of researchers who have contributed to this debate and to assess and discuss the current status of our understanding of the relative value of different metrics for different tasks.

Organiser: Daniel Hirst (daniel.hirst@lpl-aix.fr), Laboratoire Parole et Langage, Université de Provence, France.

Lessons and Challenges Deploying Voice Search

Wed-Ses1-S1: Wednesday 10:00

Place: Ainsworth (East Wing 4)

In the past year, a number of companies have deployed multimodal search applications for mobile phones. These applications enable spoken input for search, as an alternative to typing. There are many technical challenges associated with deploying such applications, including: High perplexity: A language model for general search must accommodate a very large vocabulary and tremendous range of possible inputs; Challenging acoustic environments: Mobile phones are often used when "on the go" - which can often be in noisy environments; Challenging usage scenarios: Mobile search may be used in challenging situations such as information access while driving a car. This session will focus on early lessons learned from usage data, challenges posed, and technical and design solutions to these challenges, as well as a look towards the future.

Organisers: Mike Cohen (mcohen@google.com), Google, Johan Schalkwyk (johans@google.com), Google, Mike Phillips (phillips@vlingo.com), Vlingo.

Active Listening & Synchrony

Wed-Ses2-S1: Wednesday 13:30

Place: Ainsworth (East Wing 4)

Traditional approaches to Multimodal Interface design have tended to assume a "ping-pong" or "push-to-talk" approach to speech interaction wherein either the system or the interlocuting human is active at any one time. This is contrary to many recent findings in conversation and discourse analysis, where the definition of a "turn", or even an "utterance" is found to be very complex; people don't "take turns" to talk in a typical conversational interaction, but they each contribute actively and interactively to the joint emergence of a "common understanding". . The aim of this special session, marking the 70th anniversary of synchrony research, is to bring together researchers from the various different fields, who have special interest in novel techniques that are aimed at overcoming weaknesses of the "push-to-talk" approach in interface

technology, or who have knowledge of the history of this field from which the research community could benefit.

Organisers: Nick Campbell (nick@tcd.ie), Trinity College Dublin, Ireland, Anton Nijholt (anijholt@cs.utwente.nl), University of Twente, The Netherlands, Joakim Gustafson (jocke@speech.kth.se), KTH, Sweden, Carl Vogel (vogel@tcd.ie), Trinity College Dublin, Ireland.

Machine Learning for Adaptivity in Spoken Dialogue Systems

Wed-Ses3-S1: Wednesday 16:00

Place: Ainsworth (East Wing 4)

In the past decade, research in the field of Spoken Dialogue Systems (SDS) has experienced increasing growth, and new applications include interactive mobile search, tutoring, and troubleshooting systems. The design and optimization of robust SDS for such tasks requires the development of dialogue strategies which can automatically adapt to different types of users and noise conditions. New statistical learning techniques are emerging for training and optimizing adaptive speech recognition, spoken language understanding, dialogue management, natural language generation, and speech synthesis in spoken dialogue systems. Among machine learning techniques for spoken dialogue strategy optimization, reinforcement learning using Markov Decision Processes (MDPs) and Partially Observable MDP (POMDPs) has become a particular focus. The purpose of this special session is to provide an opportunity for the international research community to share ideas on these topics and to have constructive discussions in a single, focussed, special conference session.

Organisers: Oliver Lemon (olemon@inf.ed.ac.uk), Edinburgh University, UK, Olivier Pietquin (olivier.pietquin@supelec.fr), Supélec - IMS Research Group, France.

New Approaches to Modeling Variability for Automatic Speech Recognition

Thu-Ses1-S1: Thursday 10:00

Place: Ainsworth (East Wing 4)

Despite great strides in the development of automatic speech recognition (ASR) technology, our community is still far from achieving its holy grail: an ASR system with performance comparable to humans in automatically transcribing unrestricted conversational speech, spoken by many speakers and in adverse acoustic environments. Many of the difficulties faced by ASR models are due to the high degree of variation in the acoustic waveforms associated with a given phonetic unit measured across different segmental and prosodic contexts. Such variation has both deterministic origins (intersegmental coarticulation; prosodic juncture and accent) and stochastic origins (token-to-token variability for utterances with the same segmental and prosodic structure). Current ASR systems successfully model acoustic variation that is due to adjacent phone context, but variation due to other sources, including prosodic context, speech rate, and speaker, is not adequately treated. The goal of this special session is to bring together researchers who are exploring alternative approaches to state-of-the-art ASR methodologies. Of special interest are new approaches that model variation in the speech signal at multiple levels, from both linguistic and extra-linguistic sources. In particular, we encourage the participation of those who are attempting to incorporate the insights that the field has gained over the past several decades from acoustic phonetics, speech production, speech perception, prosody, lexical access, natural language processing and pattern recognition to the problem of developing models of speech recognition that are robust to the full variability of speech.

Organisers: Carol Espy-Wilson (espy@umd.edu), University of Maryland, Jennifer Cole (jscole@illinois.edu), Illinois, Abeer Alwan, UCLA, Louis Goldstein, University of Southern California, Mary Harper, University of Maryland, Elliot Saltzman, Haskins Laboratories, Mark Hasegawa-Johnson, Illinois.

Interspeech 2009 Tutorials - Sunday 6 September 2009

T-1: Analysis by synthesis of speech prosody, from data to models

Sunday 9:15

Place: Jones (East Wing 1)

The study of speech prosody today has become a research area which has attracted interest from researchers in a great number of different related fields including academic linguistics and phonetics, conversation analysis, semantics and pragmatics, sociolinguistics, acoustics, speech synthesis and recognition, cognitive psychology, neuroscience, speech therapy, language teaching... and no doubt many more. So much so, that it is particularly difficult for any one person to keep up to date on research in all relevant areas. This is particularly true for new researchers coming into the field. This tutorial will propose an overview of a variety of current ideas on the methodology and tools for the automatic and semi-automatic analysis and synthesis of speech prosody, consisting in particular of lexical prosody, rhythm, accentuation and intonation. The tools presented will include but not be restricted to those developed by the presenter himself. The emphasis will be on the importance of data analysis for the testing of linguistic models and the relevance of these models to the analysis itself. The target audience will be researchers who are aware of the importance of the analysis and synthesis of prosody for their own research interests and who wish to update their knowledge of background and current work in the field.

Presenter: Daniel Hirst (daniel.hirst@lpl-aix.fr), Laboratoire Parole et Langage, Université de Provence, France.

T-2: Dealing with High Dimensional Data with Dimensionality Reduction

Sunday 9:15

Place: Fallside (East Wing 2)

Dimensionality reduction is a standard component of the toolkit in any area of data modelling. Over the last decade algorithmic development in the area of dimensionality reduction has been rapid. Approaches such as Isomap, LLE, and maximum variance unfolding have extended the methodologies available to the practitioner. More recently, probabilistic dimensionality reduction techniques have been used with great success in modelling of human motion. How are all these approaches related? What are they useful for? In this tutorial our aim is to develop an understanding of high dimensional data and what the problems are with dealing with it. We will motivate the use of nonlinear dimensionality reduction as a solution for these problems. The keystone to unify the various approaches to non-linear dimensionality reduction is principal component analysis. We will show how it underpins spectral methods and attempt to cast spectral approaches within the same unifying framework. We will further build on principal component analysis to introduce probabilistic approaches to non-linear dimensionality reduction. These approaches have become increasingly popular in graphics and vision through the Gaussian Process Latent Variable Model. We will review the GP-LVM and also consider earlier approaches such as the Generative Topographic Mapping and Latent Density Networks.

Presenter: Neil Lawrence (Neil.Lawrence@manchester.ac.uk), School of Computer Science, Univ. of Manchester, UK ; Jon Barker (J.Barker@dcs.shef.ac.uk), Department of Computer Science, Univ. of Sheffield, UK

T-3: Language and Dialect Recognition

Sunday 9:15

Place: Holmes (East Wing 3)

Spoken language recognition (a.k.a Language ID or LID) is a task of recognizing the language from a sample spoken by an unknown speaker. Language ID finds applications in multi-lingual dialog systems, distillation, diarization and indexing systems, speaker detection and speech recognition. Often, LID represents one of the first and necessary processing steps in many

speech processing systems. Furthermore, language, dialect, and accent are of interest in diarization, indexing/search, and may play an important auxiliary role in identifying speakers.

LID has seen almost four decades of active research. Benefitting from the development of public multi-lingual corpora in the 90's, the progress in LID technology has accelerated in the 00's tremendously. While the availability of large corpora served as an enabling medium, establishing a series of NIST-administered Language Recognition Evaluations (LRE) provided the research community with a common ground of comparison and proved to be a strong catalyst. In another positive way, the LRE series gave rise to a "cross-pollination effect" by effectively fusing the speaker and language recognition communities thus sharing and spreading their respective methods and techniques. In the past five years or so, a considerable success was achieved by focusing on and developing techniques to deal with channel and session variability, to improve acoustic language modeling by means of discriminative methods, and to further refine the basic phonotactic approaches.

The goal of this tutorial is to survey the LID area from a historical perspective as well as in its most modern state. Several important milestones contributing to the growth of the LID area will be identified. In a second, larger part, most successful state-of-the-art probabilistic approaches and modeling techniques will be described more in detail. Among these belong various phonotactic architectures, UBM-GMMs, discriminative techniques, and subspace modeling tricks. The closely related problem of detecting dialects will be discussed in the final part.

Presenter: Jiri Navratil (jiri@us.ibm.com), Multilingual Analytics and User Technologies, IBM T.J. Watson Research Center, USA.

T-4: Emerging Technologies for Silent Speech Interfaces

Sunday 9:15

Place: Ainsworth (East Wing 4)

In the past decade, the performance of automatic speech processing systems, including speech recognition, text and speech translation, and speech synthesis, has improved dramatically. This has resulted in an increasingly widespread use of speech and language technologies in a wide variety of applications, such as commercial information retrieval systems, call centre services, voice-operated cell phones or car navigation systems, personal dictation and translation assistance, as well as applications in military and security domains. However, speech-driven interfaces based on conventional acoustic speech signals still suffer from several limitations. Firstly, the acoustic signals are transmitted through the air and are thus prone to ambient noise. Despite tremendous efforts, robust speech processing systems, which perform reliably in crowded restaurants, airports, or other public places, are still not in sight. Secondly, conventional interfaces rely on audibly uttered speech, which has two major drawbacks: it jeopardizes confidential communications in public and it disturbs any bystanders. Services, which require the access, retrieval, and transmission of private or confidential information, such as PINS, passwords, and security or safety information, are particularly vulnerable.

Recently, Silent Speech Interfaces have been proposed which allow its users to communicate by speaking silently, i.e. without producing any sound. This is realized by capturing the speech signal at the early stage of human articulation, namely before the signal becomes airborne, and then transfer these articulation-related signals for further processing and interpretation. Due to this novel approach Silent Speech Interfaces have the potential to overcome the major limitations of traditional speech interfaces today, i.e. (a) limited robustness in the presence of ambient noise; (b) lack of secure transmission of private and confidential information; and (c) disturbance of bystanders created by audibly spoken speech in quiet environments; while at the same time retaining speech as the most natural human communication modality. The SSI furthermore could provide an alternative for persons with speech disabilities such as laryngectomy, as well as the elderly or weak who may not be healthy or strong enough to speak aloud effectively.

Presenter: Tanja Schultz (tanja@cs.cmu.edu), Computer Science Department, Karlsruhe University, Germany; Bruce Denby (bruce.denby@gmail.com), Université Pierre et Marie Curie (Paris-VI).

T-5: In-Vehicle Speech Processing & Analysis

Sunday 14:15

Place: Jones (East Wing 1)

In this tutorial, we will focus on speech technology for in-vehicle use by discussing the cutting-edge developments in these two applications:

1. Speech as interface: Robust speech recognition system development under vehicle-noise conditions (i.e. engine, open windows, A/C operation). This field of study includes application of microphone-arrays for in-vehicle use to reduce the effect of the noise on speech recognition employing beam-forming algorithms. The resultant system can be employed as a driver-vehicle interface for entering prompts and commands for music search, control of in-vehicle systems such as cell-phone, A/C, windows etc. instead of manual operation which engages the driver visually as well.

2. Speech as monitoring system: Speech can be used to design a sub-module for driver-monitoring systems. For the last two decades speech under stress studies has contributed to improve the performance of ASR systems. Detecting stress in speech can also help improving the performance of driver monitoring systems which conventionally relies on computer vision applications of driver head and eye tracking. On the other hand, the effects of introducing speech technologies as an interface can be assessed via driver behaviour modeling studies.

Presenter: John H.L. Hansen(john.hansen@utdallas.edu), Pinar Boyraz (pxb083000@utdallas.edu), Erik Jonsson School of Engineering and Computer Science, University of Texas, Dallas, USA

T-6: Emotion Recognition in the Next Generation: an Overview and Recent Development

Sunday 14:15

Place: Fallside (East Wing 2)

Emotional aspects have recently attracted considerable attention as being the "next big thing" for dialog systems and robotic product's market success, and practically any intelligent Human-Machine Interface. Having matured over the last decade of research, recognition technology is now becoming ready for usage in such systems, and many further applications as Multimedia Retrieval and Surveillance. At the same time systems have evolved considerably more complex: in addition to a variety of definitions and theoretical approaches, today's engines demand subject independency, coping with spontaneous and non prototypical emotions, robustness against noise, transmission, and optimal system integration.

In this respect this tutorial will present an introduction to the recognition of emotion with a particular focus on recent developments in audio-based analysis. A general introduction to researchers working in related fields will be followed by current issues and impulses for acoustic, linguistic, and multi-stream and -modal analyses. A summary of the main recognition techniques will be presented, as well as an overview on current challenges, datasets, studies and performances in view of optimal future application design. Also, the first open source Emotion Recognition Engine "openSMILE" developed in the European Union's Seventh Framework Programme Project SEMAINE will be introduced to the participants in order for them to be directly able to experiment with emotion recognition from speech or test latest technology on their datasets.

Presenter: Björn Schuller(schuller@tum.de), Munich University of Technology, Germany.

T-7: Fundamentals and recent advances in HMM-based speech synthesis

Sunday 14:15

Place: Holmes (East Wing 3)

Over the last ten years, the quality of speech synthesis has dramatically improved with the rise of general corpus-based speech synthesis. Especially, state-of-the-art unit selection speech synthesis can generate natural-sounding high quality speech. However, for constructing human-like talking machines, speech synthesis systems are required to have an ability to generate speech with arbitrary speaker's voice characteristics, various speaking styles including native

and non-native speaking styles in different languages, varying emphasis and focus, and/or emotional expressions; it is still difficult to have such flexibility with unit-selection synthesizers, since they need a large-scale speech corpus for each voice.

In recent years, a kind of statistical parametric speech synthesis based on hidden Markov models (HMMs) has been developed. The system has the following features:

1. Original speaker's voice characteristics can easily be reproduced because all speech features including spectral, excitation, and duration parameters are modeled in a unified framework of HMM, and then generated from the trained HMMs themselves.
2. Using a very small amount of adaptation speech data, voice characteristics can easily be modified by transforming HMM parameters by a speaker adaptation technique used in speech recognition systems.

From these features, the HMM-based speech synthesis approach is expected to be useful for constructing speech synthesizers which can give us the flexibility we have in human voices.

In this tutorial, the system architecture is outlined, and then basic techniques used in the system, including algorithms for speech parameter generation from HMM, are described with simple examples. Relation to the unit selection approach, trajectory modeling, recent improvements, and evaluation methodologies are summarized. Techniques developed for increasing the flexibility and improving the speech quality are also reviewed.

Presenters: Keiichi Tokuda (tokuda@nitech.ac.jp), Department of Computer Science and Engineering, Nagoya Institute of Technology; Heiga Zen (heiga.zen@crl.toshiba.co.uk), Toshiba Research Europe Ltd.

T-8: Statistical approaches to dialogue systems

Sunday 14:15

Place: Ainsworth (East Wing 4)

The objective of this tutorial is to provide a comprehensive, cohesive overview of statistical techniques in dialog management for the newcomer. Specifically we will start by motivating the research area by showing how traditional techniques fail and intuitively why statistical techniques would be expected to do better. Then, in classroom style presentation, we will explain the core algorithms and how they have been applied to spoken dialogue systems. Our intention is to provide a cohesive treatment of the techniques using a unified, common notation in order to give the audience a clear understanding of how the techniques interrelate. Finally we will report results from the literature to provide an indication of the impact in practice. Through the tutorial we will draw on both our own work and the literature (with citations throughout), and wherever possible we will use audio/video recordings of interactions to illustrate operation. We will provide lecture notes and a comprehensive bibliography. Our aim is that attendees to this course should be able to readily read papers in this area, comment on them meaningfully, and (we hope!) suggest avenues for future work in this area rich in open challenges and begin research enquiries of their own.

Presenters: Jason Williams (jdw@research.att.com), AT&T Labs – Research, USA; Steve Young (sjy@eng.cam.ac.uk) , Blaise Thomson(brmt2@cam.ac.uk) , Information Engineering Division, Cambridge University, UK.

Public Engagement Events

The first Interspeech conversational systems challenge

Sunday 14:15

Place: Rainbow Room

The first interspeech conversational systems challenge is based around the original Loebner competition but due to the unique challenges of speech we have changed things slightly. We have devised a scenario that presents an urgent and direct task full of 'full-blown' emotion. As a result competitors systems will have to convey urgency and emotion through speech, while any speech recognition system will have to function successfully in a conversational context with little time for training.

Each judge will be given the following briefing:

"You're a captain of the one of the fleets finest starships, suddenly your sensors detect a badly damaged ship heading straight for you, the intercom crackles into life: there's lots of interference but they're requesting to dock. The ship is about to crash into you, do you push the button blowing the artificial infiltrator out of the sky or do you open the landing bay and guide the human refugee to safety, you have 3 minutes to decide."

The artificial system that fools our judges for the longest period of time will be declared the winner.

Competitors: Marc Schroeder and Jens Edlund; Organiser: Simon Worgan (s.worgan@dcs.shef.ac.uk)

Interspeech 2009 public exhibition

Sunday

10:00

Place: Public Foyer

On Sunday 6th September a number of exhibitors will be demonstrating aspects of speech and language technology to the general public. Hosted in the public foyer, exhibits will include emotive talking heads, agents that attempt to elicit rapport from human speakers and customized text-to-speech systems.

DAY 0: Sunday September 6th

	Jones (East Wing 1)	Fallside (East Wing 2)	Holmes (East Wing 3)	Ainsworth (East Wing 4)	Rainbow Room
	TUTORIAL				
08:30	Registration for tutorials opens (closes at 14:30)				
09:00	ISCA Board Meeting 1 (finish at 17:00) - BCS Room 3				
09:15	T-1: Analysis by Synthesis of Speech Prosody, from Data to Models	T-2: Dealing with High Dimensional Data with Dimensionality Reduction	T-3: Language and Dialect Recognition	T-4: Emerging Technologies for Silent Speech Interfaces	
10:45	Coffee break				
11:15	T-1: Analysis by Synthesis of Speech Prosody, from Data to Models	T-2: Dealing with High Dimensional Data with Dimensionality Reduction	T-3: Language and Dialect Recognition	T-4: Emerging Technologies for Silent Speech Interfaces	Loebner Competition
12:45	Lunch				
14:00	General registration opens (closes at 18:00)				
14:15	T-5: In-Vehicle Speech Processing & Analysis	T-6: Emotion Recognition in the Next Generation: an Overview and Recent Development	T-7: Fundamentals and Recent Advances in HMM-based Speech Synthesis	T-8: Statistical Approaches to Dialogue Systems	The first Interspeech conversational systems challenge
15:45	Tea break				
16:15	T-5: In-Vehicle Speech Processing & Analysis	T-6: Emotion Recognition in the Next Generation: an Overview and Recent Development	T-7: Fundamentals and Recent Advances in HMM-based Speech Synthesis	T-8: Statistical Approaches to Dialogue Systems	
17:45					
18:00	Elsevier Thank You Reception for Former Computer Speech and Language Editors (finish at 19:30) - BCS Room 1				

DAY 1: Monday September 7th

	Main Hall (East Wing 1)	Jones (East Wing 1)	Fallside (East Wing 2)	Holmes (East Wing 3)	Hewison Hall (East Wing 4)	Ainsworth (East Wing 4)
	ORAL				POSTER	SPECIAL
09:00	Arrival and Registration (closes at 17:00)					
10:00	Opening Ceremony in Main Hall					
11:00	Mon-Ses1-K, Plenary Session in Main Hall. Keynote Speaker: Sadaoki Furui , ISCA Medallist, Department of Computer Science, Tokyo Institute of Technology <i>Selected topics from 40 years of research on speech and speaker recognition</i>					
12:00	Lunch; IAC (Advisory Council) Meeting - BCS Room 3					
13:30	Mon-Ses2-01 ASR: Features for Noise Robustness	Mon-Ses2-02 Production: Articulatory modelling	Mon-Ses2-03 Systems for LVCSR and Rich Transcription	Mon-Ses2-04 Speech Analysis and Processing I	Mon-Ses2-P1 Speech Perception I	Mon-Ses2-P2 Accent and Language Recognition
15:30	Tea Break					
16:00	Mon-Ses3-01 ASR: Language Models I	Mon-Ses3-02 Phoneme-level Perception	Mon-Ses3-03 Statistical Parametric Synthesis I	Mon-Ses3-04 Systems for Spoken Language Translation	Mon-Ses3-P1 Human Speech Production I	Mon-Ses3-P2 Prosody, Text Analysis, and Multilingual Models
18:00					Mon-Ses2-P3 ASR: Acoustic Model Training and Combination	Mon-Ses2-P4 Spoken Dialogue Systems
19:30	Welcome Reception - Brighton Dome				Mon-Ses3-P3 ASR: Adaptation I	Mon-Ses3-P4 Applications in Learning and Other Areas
						Mon-Ses2-S1 Special Session: INTERSPEECH 2009 Emotion Challenge
						Mon-Ses3-S1 Special Session: Silent Speech Interfaces

Interspeech 2009 - Main Conference Session Codes

DAYS: Mon = Monday
TIMES: Ses1 = 10:00 - 12:00
TYPE: O = Oral
 Tue = Tuesday
 Ses2 = 13:30 - 15:30
 P = Poster
 Wed = Wednesday
 Ses3 = 16:00 - 18:00
 S = Special
 Thu = Thursday
 K = Keynote

Day 2: Tuesday September 8th

	Main Hall	Jones (East Wing 1)	Fallside (East Wing 2)	Holmes (East Wing 3)	Hewison Hall	Ainsworth (East Wing 4)
	ORAL				POSTER	SPECIAL
08:00	Registration (closes at 17:00)					
08:30	Tues-Ses1-K, Plenary Session in Main Hall. Keynote Speaker: Tom Griffiths , University of California, Berkeley, USA <i>Connecting human and machine learning via probabilistic models of cognition</i>					
09:30	Coffee Break					
10:00	Tue-Ses1-O1 ASR: Discriminative Training	Tue-Ses1-O2 Language Acquisition	Tue-Ses1-O3 ASR: Lexical and Prosodic Models	Tue-Ses1-O4 Unit-Selection Synthesis	Tue-Ses1-P1 Human Speech Production II	Tue-Ses1-P2 Speech perception II
					Tue-Ses1-P3 Speech and Audio Segmentation and Classification	Tue-Ses1-P4 Speaker Recognition and Diarisation
						Tue-Ses1-S1 Special Session: Advanced Voice Function Assessment
12:00	Lunch; Elsevier Editorial Board Meeting for Computer Speech and Language - BCS Room 1; Special Interest Group Meeting - BCS Room 3					
13:30	Standardising assessment for voice and speech pathology (finish at 14:30) - BCS Room 3					
13:30	Tue-Ses2-O1 Automotive and Mobile Applications	Tue-Ses2-O2 Prosody: Production I	Tue-Ses2-O3 ASR: Spoken Language Understanding	Tue-Ses2-O4 Speaker Diarisation	Tue-Ses2-P1 Speech Analysis and Processing II	Tue-Ses2-P2 Speech Processing with Audio or Audiovisual Input
					Tue-Ses2-P3 ASR: Decoding and Confidence Measures	Tue-Ses2-P4 Robust Automatic Speech Recognition I
						ISCA Student Advisory Committee
15:30	Tea Break					
16:00	Tue-Ses3-S1 Panel: Speech & Intelligence		Tue-Ses3-O3 Speaker Verification & Identification I	Tue-Ses3-O4 Text Processing for Spoken Language Generation	Tue-Ses3-P1 Single- and Multi-Channel Speech Enhancement	Tue-Ses3-P2 ASR: Acoustic Modelling
					Tue-Ses3-P3 Assistive Speech Technology	Tue-Ses3-P4 Topics in Spoken Language Processing
						Tue-Ses3-S2 Special Session: Measuring the Rhythm of Speech
18:00						
18:15	ISCA General Assembly - Main Hall					
19:30	Reviewers' Reception - Brighton Pavilion; Student Reception - AI Duomo Restaurant					

DAY 3: Wednesday September 9th

	Main Hall	Jones (East Wing 1)	Fallside (East Wing 2)	Holmes (East Wing 3)	Hewison Hall	Ainsworth (East Wing 4)
08:00	Registration (closes at 17:00)					
08:30	Wed-Ses0-K, Plenary Session in Main Hall. Keynote Speaker: Deb Roy , MIT Media Laboratory <i>New Horizons in the Study of Language Development</i>					
09:30	Coffee Break					
10:00	Wed-Ses1-01 Speaker Verification & Identification II	Wed-Ses1-02 Emotion and Expression I	Wed-Ses1-03 ASR: Adaptation II	Wed-Ses1-04 Voice Transformation I	Wed-Ses1-P1 Phonetics, Phonology, Cross-Language Comparisons, Pathology	Wed-Ses1-P2 Prosody Perception and Language Acquisition
12:00	Lunch; Interspeech Steering Committee - BCS Room 1; Elsevier Editorial Board Meeting for Speech Communication - BCS Room 3					
13:30	Wed-Ses2-01 Word-level Perception	Wed-Ses2-02 Applications in Education and Learning	Wed-Ses2-03 ASR: New Paradigms I	Wed-Ses2-04 Single-Channel Speech Enhancement	Wed-Ses2-P1 Emotion and Expression II	Wed-Ses2-P2 Expression, Emotion and Personality Recognition
15:30	Tea Break					
16:00	Wed-Ses3-01 Language Recognition	Wed-Ses3-02 Phonetics & Phonology	Wed-Ses3-03 Speech Activity Detection	Wed-Ses3-04 Multimodal Speech (e.g. Audiovisual Speech, Gesture)	Wed-Ses3-P1 Phonetics	Wed-Ses3-P2 Speaker Verification & Identification III
18:00	Registration (closes at 17:00)					
19:30	Revelry at the Racecourse					

DAY 4: Thursday September 10th

	Main Hall	Jones (East Wing 1)	Fallside (East Wing 2)	Holmes (East Wing 3)	Hewison Hall	Ainsworth (East Wing 4)
	ORAL				POSTER	SPECIAL
08:00	Registration (closes at 17:00)					
08:30	Thu-Ses0-K, Plenary Session in Main Hall. Keynote Speaker: Mari Ostendorf , University of Washington <i>Transcribing Speech for Spoken Language Processing</i>					
09:30	Coffee Break					
10:00	Thu-Ses1-01 <i>Robust Automatic Speech Recognition III</i>	Thu-Ses1-02 <i>Prosody: Perception</i>	Thu-Ses1-03 <i>Segmentation and Classification</i>	Thu-Ses1-04 <i>Evaluation & Standardisation of SL Technology and Systems</i>	Thu-Ses1-P1 <i>Speech Coding</i>	Thu-Ses1-P2 <i>Voice Transformation II</i>
					Thu-Ses1-P3 <i>ASR: Language Models II</i>	Thu-Ses1-P4 <i>Systems for Spoken Language Understanding</i>
						Thu-Ses1-S1 <i>Special Session: New Approaches to Modeling Variability for ASR</i>
12:00	Lunch; Industrial Lunch - BCS Room 1					
13:30	Thu-Ses2-01 <i>User Interactions in Spoken Dialog Systems</i>	Thu-Ses2-02 <i>Production: Articulation and Acoustics</i>	Thu-Ses2-03 <i>Features for Speech and Speaker Recognition</i>	Thu-Ses2-04 <i>Speech and Multimodal Resources & Annotation</i>	Thu-Ses2-P1 <i>Speaker and Speech Variability, Paralinguistic and Nonlinguistic Cues</i>	Thu-Ses2-P2 <i>ASR: Acoustic Model Features</i>
					Thu-Ses2-P3 <i>ASR: Tonal Language, Cross-Lingual and Multilingual ASR</i>	Thu-Ses2-P4 <i>ASR: New Paradigms II</i>
						Thu-Ses2-O5 <i>Speech Analysis and Processing III</i>
15:30	Tea Break					
16:00	Closing Ceremony - Main Hall					