Joakim Gustafson CV

1.0 INTRODUKTION

1.1 Name:
Joakim Gustafson

1.2 Date of birth
August 17, 1966

1.3 Gender
Male

1.4 Home address and telephone number
Trillans väg 37, 131 49, Nacka. Mobile: 070 206 08 17

1.5 Contact address, telephone number, email
Address: Instutitionen för Tal, Musik och Hörsel, KTH CSC, Lindstedsvägen 24, 10044 Stockholm.
Phone: 08-790 89 65
E-mail jocke@speech.kth.se
Web: http://www.speech.kth.se/~jocke

1.6 Current Position
Professor at TMH, since 2013-03-01, Subject area: Speech Technology

1.7 Previous employments

2011-2013	Associate Professor (Lektor) at TMH/KTH
2007-2011	Assistant Professor (future faculty Biträdande Lektor) at TMH/KTH
2000-2007	Industrial researcher in speech technology at the Telia R&D department
1996-2000	Doctoral student at TMH/KTH
1993-1996	Research Engineer at TMH/KTH

2.0 DEGREES AND EVALUATIONS

2.1 University degrees

Civ. Ing.	Electrical engineering (with computer science orientation) at KTH, 1995.
Tech. Lic.	Technical licentiate degree in speech communication 1996 (title: *A Swedish Name Pronunciation System*)
Tech. Doc.	Technical doctorate degree in speech communication 2002 (Title: *Developing Multimodal Spoken Dialogue Systems - Empirical Studies of Spoken Human-Computer Interactions)*

2.2 Conferment of the title Docent

Docent Docent in speech technology 2009
(Title on seminar: *A keyboard..how quaint! – Building the Next Generation intuitive speech interfaces*)

3.0 SCIENTIFIC ACHIEVEMENTS

3.1 Brief account of own research profile

Over the past 20 years, I have gathered experience from working on all aspects of spoken language technology, in research as well as commercial settings. The dialogue system projects I have been involved in have both used and developed core speech and dialogue technology components (i.e. speech recognition, speech synthesis, multimodal output generation, spoken language processing and dialogue management). I have presented papers at more than 25 international workshops and conferences, and participated actively in six EU projects and in two national multi-site projects. I have co-written EU-proposal on 22 occasions (8 times as coordinator) and proposals for national funding agencies on 10 occasions (VR, KK, FAS and RJ). At Telia Research, I participated in four development projects involving close collaboration with external companies. Taken together, these activities have given me an extensive network, both nationally and internationally, at research institutes as well as companies in my field. My core research interests are in the overall design of complete spoken dialogue systems, including empirical studies and evaluation of these systems through usability studies, as well as analysis of human-human conversational behavior. I have been able to pursue these research interests on account of my ability to collaborate with and inspire others, in order for us to do things we could not have done on our own. I have special interests in building publically available spoken dialogue systems and system that makes use of several modalities. I have conducted studies on how users are influenced by system behaviors, with the purpose of improving the system's understanding capabilities. However, my goal is to build conversational systems that do not force its users to interact in a special computer-directed manner, but rather ones that allow users to interact with a human metaphor in mind.

I did my undergraduate studies at the electrical engineering program at KTH 1987-1992. In addition, I took courses in linguistics at the Stockholm University. In 1992, I completed my undergraduate thesis work at TMH. In 1993 got a position ar research engineer at TMH, working in the EU project Onomastica, in which I developed a Swedish name pronunciation system. This resulted in my licentiate thesis in 1996. In 1995 I started working with dialogue analysis, multimodal output and dialogue design in the Waxholm project (Carlson et al., 1995), and I decided to do a PhD in dialogue systems at TMH. In 1997, I participated in the Olga project (multi-site project: TMH, SICS, NADA and SU), where I developed the speech recognition grammars and worked on multimodal output generation together with Jonas Beskow. In 1997 I started the development of the *Gulan dialogue laboratory toolkit* together with Kåre Sjölander, and continued its development as part of the multi-site project *Swedish Dialogue Systems* (Gustafson et al. 1998). In 1998, TMH participated with an installation for the "*Cultural Capital of Europe*" at Stockholm Akademiska Forum in Kulturhuset. In conjunction with this, I initiated the *August project* where I led the development work (Gustafson et al., 1999). The aim was to build a synthetic August Strindberg that visitors could talk with. Apart from the system design, I built the dialogue and understanding parts of the system in collaboration with Nikolaj Lindberg. The system was used to collect and analyze more than 10.000 utterances from the general public, dialogue data which I analyzed in collaboration with Linda Bell. (Gustafson and Bell, 2000). In 1999, I started my collaboration with Johan Boye and Mats Wirén, at Telia Research who at the time participated in the Center for Speech Technology (CTT). Together with other colleagues we developed *Adapt*, which was a multimodal spoken dialogue system used to find apartments for sale in Stockholm (Gustafson et al., 2000).

In 2000, I was offered a position at Telia Research as a senior researcher in their multimodal group. My main role was to initiate and lead new research projects at Telia Research. Initially, I finished the development of Adapt by continuing my collaboration with Jens Edlund at CTT (Gustafson et al., 2002). In 2001, I participated in an internal Telia project, *TänkOm*, where we collaborated with industrial design companies, a computer game company (Liquid Media) and the Telecom museum in Stockholm. In the project we built an apartment of the future that the museum visitors could enter. I worked on the development of an animated agent called Pixie that visitors could talk with to play a game and to control the apartment. In order to handle speech from children in the exhibition Kåre Sjölande and I developed a method for improving the speech recognition rates for young speakers, (Gustafson and Sjölander, 2002). The system was open for the public for more than two years resulting in more than 100.000 utterances from the visitors. We transcribed and analyzed about 25.000 of these in order to study how children interact with computers (Bell and Gustafson, 2003). In 2002 I defended my PhD thesis, which contained research work done both at KTH and Telia. In 2001, I was the co-author to the EU project application *NICE*. The project got funded 2002, and the consortium consisted of Liquid Media, University of Odense, LIMSI and Philips. The goal of the NICE

project was to build a computer game where children could talk to 3D animated fairy-tale characters in H-C Andersen's fairy-tale world. I was the project leader of the Telia part of the NICE project, and I collaborated with Johan Boye on the system and dialogue design (Boye and Gustafson, 2005, Boye, Gustafson and Wiren, 2005); with Kåre Sjölander to develop a Swedish unit selection synthesizer (Gustafson and Sjölander, 2004) and with Liquid Media for the multimodal output generation part of the system (Gustafson et al., 2005).

My secondary role was to act as expert in the development and roll-out of speech-based applications. In this capacity, I worked in projects with the aim of investigating the commercial aspects of spoken dialogue applications, as well as in projects were we performed usability studies on deployed applications. In 2006 I was project leader for a project where we did Swedish adaptations and performed usability tests of a system developed by the US company V-enable, in which customers used their voices to search for and download multimodal content on mobile phones. I also headed the subsequent deployment project which aimed at launching this system on a new web portal, (www.innovationworld.com), where TeliaSonera presented and distributed new products and services. The project involved 25 people, from five internal units and three external companies. This is one example of a business development process where I have been involved from beginning to end. Another example is the introduction of an automated call routing system for Telia's customer care line. While visiting AT&T in 2001, we got to know that AT&T was launching a call routing system for their customer care. I and my colleagues then spent three years convincing the management at Telia to add such a system in their customer care. This led to a successful pilot test in 2005, in which I participated as a spoken dialogue expert. I then participated in the business requirements study that lead to the decision of developing an automated call routing system in the customer care line. In 2006 the system was launched in collaboration with the world's largest speech technology company (Nuance), and I participated as a dialogue expert support in the development project. In conjunction with the data collection for the deployed call routing system, I also initiated a research project, in which we performed a small WoZ collection on the usefulness of using a human metaphor in call routing systems (Gustafson, Heldner and Edlund, 2008). Lastly, I initiated a project in 2006 that investigated the possibility to deploy a speech-based problem-solving spoken dialogue system for the broadband support line. As usual, the first part of the project was educational: convincing both the technical staff and management in the broadband department that it would be feasible and profitable to develop a speech based problem-solving system in their support line. I then participated in the negotiations with an American company (SpeechCycle) that had deployed similar systems in the US. In the end Telia decided to run an internal project that investigated the possibility to build the problem solving dialogue system in-house, using the dialogue technology developed by Johan Boye in the NICE project.

In July 2007 I got a position as Future Faculty *Biträdande Lektor (Assistant Professor)* at TMH, in June 2009 I became *Docent,* February 2011 I was promoted to *Lektor (Associate Professor)* and in February 2013 I was promoted to *Professor*. I have taken four pedagogical courses and been a teacher myself in seven different courses for both undergraduate and graduate students. Since I got the position in 2007 I have spent a lot of time trying to get external funding to our group. I have co-authored 32 project proposals for external funding, resulting in a total funding of approximately 2.5 M€. I am currently supervising four PhD students, and I have continued my connection with the commercial side of spoken dialogue systems by supervising Master thesis students at two of the largest speech application companies in Sweden. I have worked on the development of tools for exploring and examining spoken dialogue systems with a human metaphor (Edlund, Gustafson and Heldner, 2008). The purpose is to acquire the data and knowledge necessary to build systems that allow more natural interaction. In order to examine conversational and expressive functions of prosody I have developed the EXPROS tool (Gustafson and Edlund, 2008). We have used this re-synthesis tool to produce stimuli for preliminary perceptual studies of speaking style (Strangert and Gustafson, 2008) and grounding status (Edlund and Gustafson 2010). The tool has also been used to synthesis cue phrases (Hjalmarsson, 2009). In order to acquire data for the development of more human-like spoken dialogue, I have also developed a WoZ tool that enable a human operator the control the flow and evolvement of an ongoing human-human dialogue. These tools were used by my Master Thesis student (Gustafson and Merkes, 2009). I have also done work within the VR project Simulekt, where I together with Jonas Beskow have built dialect voices using HMM synthesis (Schötz et al, 2010). I have worked in the EU project *MonAMI*, which aimed at evaluating the usefulness of ECA technology for the disabled and elderly. My contribution is in the development and evaluation of a speech application in the calendar domain, including modules for audiovisual interaction control (Skantze & Gustafson, 2009). I have presented the Reminder at international/national workshops/conferences. I have co-authored four deliverables and I presented KTH's work at the annual review meetings. I am currently working in IURO (Interactive Urban Robot, 2010-2013), an EU project where we are developing a mobile robot that asks humans for directions in the streets of Munich. In this project I have also presented KTH's work at the yearly review meetings (Munich 2011, 2012). Finally, I am currently formal project leader in the following projects:

- **SAVIR** (Situated Audio Visual Interaction with Robots, 2010-2013), a SRA/TNG project where we will collaborate with CVAP in order to build a robot that can talk with a human in order to get help in interpreting visual scenes.
- **Talsyntes för samtal** (2010-2013), a VR funded project where I am project leader and perform research work.
- **GetHomeSafe** (2012-2015), an EU project about humanlike proactive dialogue behaviors for in-car dialogue systems
- **Two EIT LABS KIC projects** (2012) - both are catalyst projects to the GetHomeSafe project: one develops a proof-of-concept social mobility dialogue system, and the other sets up a test bed for safe proactive dialogue.

Since January 2011, I am deputy Head of the Department of Speech, Music and Hearing, and since March 2013 I am head of the department's Speech Group. I was part of the writing group of the evaluation package for the Applied Computer Science Unit of Assessment in the RAE2012. In the review report the research output of our UoA assessesed to be world leading and among the best UoA at KTH.

3.2 Brief account of planned research effort

Below is an overview of topics of research I plan to pursue, with pointers to current and planned projects:

Human-like conversational systems

I have examined the usefulness of developing human-like spoken dialogue systems (Edlund et al. 2008 and Gustafson et al. 2008). The aim is to build and study spoken dialogue systems that can be understood through a human metaphor. Humans are often quite willing participants, and the human metaphor allows us to draw on this by borrowing from other areas. For example, the gaming, film and fiction industries rely heavily on willing 'suspension of disbelief' – the ability to ignore minor inconsistencies in order to enjoy a work of fiction. While the goal of human-likeness may eventually turn out to be overly ambitious, it has already been useful to me and my colleagues by guiding our research efforts towards gaps in our knowledge about human conversation. It is for example evident that state-of-the-art speech technology neither sounds like a conversational partner, nor understands fundamental aspects of human conversational behavior. A massive amount of cross-disciplinary research is needed to realize the visionary goal. Although we have a long way to go, we learn useful lessons from every step. Our group is currently involved in a half-dozen national and international research projects orchestrated towards our goal. In this area we are collecting and analyzing human-human dialogues in order to get the knowledge necessary to build or simulate more human-like dialogue systems.

Today's state-of-the-art talking computers do not sound like they are having a conversation even if they say the same thing a human in the same situation would. One problem is that many of the conversational cues are removed in speech synthesis (e.g. hesitations and filled pauses). As part of the VR funded project *Introducing interactional phenomena in speech synthesis* we aim at investigating interactional tokens that are traditionally lacking in speech synthesis voices, e.g. conversational grunts and audible breathing. We are currently looking at events that occur in the transitions between speech and silence, especially breathy and creaky voices. At the end of the project we will perform user studies using a semi-automatic dialogue system that models speaking and listening as parallel and mutually aware processes. We will use two scenarios to verify and validate our results: the attentive speaker – an interruptible virtual narrator making use of synthesized cues for hesitation and end-of contribution; and the active listener – an information gathering system, aiming to encourage the user to continue speaking (cf. Gustafson, Heldner, & Edlund, 2008).

When people talk to each other, their prosody is used for more than deciding who talks when. It also signals the speaker's attitude towards the issues under discussion, it distinguishes the new and important information from the old and referential and it places contrastive focus. If we really want to be able to understand and generate conversational speech it is crucial that we gain more knowledge about how prosody is used in human dialogues. To date, we still have incomplete knowledge about how prosody is used in human-human spoken interaction, and less still about how it can be utilized in human-computer interfaces utilizing the speech channel. I would like to research prosodic aspects important for spoken dialogue components, and I would like to evaluate the effect of doing so on users that interact with dialogue systems that make extensive use of prosody. I would also like to investigate which aspects of prosody that actually can be manipulated in a dialogue system given the state of current technology.

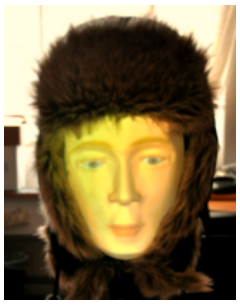
I am the KTH PI for a new EU project, GetHomeSafe, that exploits a combination of humanlike dialogue behaviors, flexible and dynamic speech recognition and speech synthesis, in combination exceeding by far state-of-the-art in in-car systems. In this project state-of-the-art technologies such as Nuance's ASR and speech synthesis are developed and combined with innovative methods such situation-aware multimodal interaction and humanlike proactive behaviors to create innovative, but safe systems which are tested realistic environments. KTH leads a work package called Human-Like Pro-Active Behavior that will focus on unobtrusive attention grabbing, user controlled dialogue pacing and situation-sensitive spoken output. In this project we will make study how cognitive load influences the speaking style and turn taking patterns in dialogue.

Situated face-to-face interaction with robots

Future human-like and situated systems that operate in real and populated situations and spaces will have to deal with unstructured, uncertain and dynamically changing environments. Here, pre-programmed models and predefined concepts are of limited use. Instead, these systems need the capability to learn and reason incrementally and in real-time. The systems need to be made aware of the environment and the situation while capturing and responding to multi-sensory input such as speech, gestures and actions. They must also achieve the short feedback loops humans are used to and expect from intelligent systems. This necessitates the development and implementation of incremental processes for detection, interpretation and generation of interaction, in particular spoken face-to-face communication and audio-visual conversational behavior.

Face-to-face conversation involves other sources of information than the speech channel, perhaps most notably a visual channel with gaze, nods, other gestures, posture, proxemics etc. that forms an intrinsic part of the communication. Thus, further investigations about how humans converse are motivated from basic and applied research perspectives alike. From a practical point of view the visual channel will become more important as speech interfaces are becoming standard for robots. I am currently involved in two projects that aim at developing a speech interface for robots. In the SRA project *Situated audiovisual interaction with robots*, the goal is to build a research platform which combines spoken dialogue technology with visual object recognition in a robot. The platform will enable research on: true mixed initiative, collaborative human-robot interaction, cognitive modeling of visual scenes and how robots can to learn by audiovisual interaction with humans. In the EU project IURO the aim is to build a robot that can approach pedestrians in the streets of Munich, and ask for directions to its destination.

In the coming years I and my colleagues will continue our collaboration with CVAP through the SRA/TNG project SAVIR, where current focus is to equip robotic heads with appropriate sensors (cameras, microphones etc) and audio-visual interpretation/fusion modules for understanding and facial gesture generation for output generation. We will compare our back-projected head FurHat, with a mechatronic head that was developed by ACREA within the IURO project. We will also build situated dialogue systems that can achieve mutual gaze and talk about objects in the common space.



3.3 List of publications

I have performed an informal citation evaluation using Google Scholar. It lists 90 articles that have been cited 1181 times, and the computed h-index is 20.

3.3.1 Peer-reviewed Journals

1. Mirnig, N., Weiss, A., Skantze, G., Al Moubayed, S., Gustafson, J., Beskow, J., Granström, B. and Tscheligi, M. (2013) "Face-to-face with a robot: what do we actually talk about?", International Journal of Humanoid Robotics, in press.
2. Neiberg, D. Salvi, G. and Gustafson, J. (2013) "Semi-supervised methods for exploring the acoustics of simple productive feedback", J. of *Speech Communication*.
3. Edlund, J., Gustafson, J., Heldner, M., and Hjalmarsson, A. (2008) "Towards more human-like spoken dialogue systems" J. of *Speech Communication*, Special Issue on Evaluating new methods and models for advanced speech-based interactive systems.
4. Boye, J., Gustafson, J. and Wiren, M. (2006) "Robust spoken language understanding in a computer game", J. of *Speech Communication*, special issue on spoken language understanding, Volume48, Issues 3-4, March-April 2006, Pages 335-353.
5. Gustafson, J. and Bell, L. (2000) "Speech Technology on Trial: Experiences from the August System", Journal *of Natural Language Engineering*: Special issue on Best Practice in Spoken Dialogue Systems.*

3.3.2 Other publications, including books

Peer reviewed conference contributions

1. Boye, J., Fredriksson, M., Götze, J., Gustafson, J. & Königsmann, J. (2012) "Walk this way: Spatial grounding for city exploration", in Proceedings of IWSDS2012.
2. Al Moubayed, S., Beskow, J., Granström, B., Gustafson, J., Mirning, N., Skantze, G., & Tscheligi, M. (2012) "Furhat goes to Robotville: a large-scale multiparty human-robot interaction data collection in a public space", In Proceedings of LREC Workshop on Multimodal Corpora. Istanbul, Turkey.
3. Blomberg, M., Skantze, G., Al Moubayed, S., Gustafson, J., Beskow J. and Granström, B. (2012) "Children and adults in dialogue with the robot head Furhat - corpus collection and initial analysis", in Proceedings of WOCCI 2012 (Workshop on Child, Computer and Interaction), Portland, USA.
4. Edlund, J., Heldner, M., & Gustafson, J. (2012) "Who am I speaking at? - perceiving the head orientation of speakers from acoustic cues alone". In proceedings of LREC Workshop on Multimodal Corpora 2012. Istanbul, Turkey.

5. Edlund, J., Oertel, C., & Gustafson, J. (2012) "Investigating negotiation for load-time in the GetHomeSafe project", In Proceedings of Workshop on Innovation and Applications in Speech Technology (IAST). Dublin, Ireland.
6. Edlund, J., Heldner, M., & Gustafson, J. (2012). "On the effect of the acoustic environment on the accuracy of perception of speaker orientation from auditory cues alone" To be published in Proc. of Interspeech 2012. Portland, Oregon, US
7. Meena, R., Skantze, G, , & Gustafson, J. (2012). "A Data-driven Approach to Understanding Spoken Route Directions in Human-Robot Dialogue" To be published in Proc. of Interspeech 2012. Portland, Oregon, US
8. Neiberg, D. and Gustafson, J. (2012) "Exploring the implications for feedback of a neurocognitive theory of overlapped speech" in Proceedings of The Interdisciplinary Workshop on Feedback Behaviors, Oregon, USA.
9. Neiberg, D. and Gustafson, J. (2012) "Cues to perceived functions of acted and spontaneous feedback expressions" in Proceedings of The Interdisciplinary Workshop on Feedback Behaviors, Oregon, USA.*
10. Oertel, C., Wlodarczak, M., Edlund, J., Wagner, P., & Gustafson, J. (2012). "Gaze Patterns in Turn-Taking", in Proceedings of Interspeech 2012. Portland, Oregon, US
11. Skantze, G., Al Moubayed, S., Gustafson, J., Beskow, J., & Granström, B. (2012). Furhat at Robotville: A Robot Head Harvesting the Thoughts of the Public through Multi-party Dialogue. in Proceedings of IVA-RCVA. Santa Cruz.*
12. Johansson, M., Skantze, G., & Gustafson, J. (2011) "Understanding route directions in human-robot dialogue", in Proceedings of SemDial. Los Angeles, CA.
13. Johnson-Roberson, M., Bohg, J., Skantze, G., Gustafson, J., Carlson, R., Rasolzadeh, B., & Kragic, D. (2011) "Enhanced Visual Scene Understanding through Human-Robot Dialog", in proc of IEEE/RSJ Int. Conference on Intelligent Robots and Systems.*
14. Neiberg, D., & Gustafson, J. (2011) "A Dual Channel Coupled Decoder for Fillers and Feedback", In proc. of Interspeech 2011.
15. Neiberg, D., Ananthakrishnan, G., & Gustafson, J. (2011) "Tracking pitch contours using minimum jerk trajectories", In proceedings of Interspeech 2011.
16. Neiberg, D., & Gustafson, J. (2011) "Predicting Speaker Changes and Listener Responses With And Without Eye-contact", In proceedings of Interspeech 2011.
17. Neiberg, D., and Gustafson, J. (2010) "The Prosody of Swedish Conversational Grunts", In Proc Interspeech 2010, Special Session on Social Signals in Speech.
18. Gustafson, J., & Neiberg, D. (2010) "Prosodic cues to engagement in non-lexical response tokens in Swedish", In proc of DiSS-LPSS joint workshop.
19. Neiberg, D., and Gustafson, J. (2010). "Modeling Conversational Interaction Using Coupled Markov Chains", In proc of DiSS-LPSS joint workshop.
20. Schötz, S., Beskow, J., Bruce, G., Granström, B. and Gustafson, J. (2010) "Simulating Intonation in Regional Varieties of Swedish", Speech Prosody 2010, Chicago, USA.
21. Gustafson, J. and Merkes, M. (2009) "Eliciting interactional phenomena in human-human dialogues" In Proceedings of SigDial 2009. London, UK. *
22. Skantze, G. and Gustafson, J. (2009) "Attention and Interaction Control in a Human-Human-Computer Dialogue Setting" In Proceedings of SigDial 2009. London, UK.
23. Beskow, J., Edlund, J., Granström, B., Gustafson, J., Skantze, G., and Tobiasson, H. (2009). "The MonAMI Reminder: a spoken dialogue system for face-to-face interaction". In proceedings of Interspeech 2009. Brighton, U.K.
24. Beskow, J., Edlund, J., Granström, B., Gustafson, J., and Skantze, G. (2008) "Innovative interfaces in MonAMI: the KTH Reminder" In Proc. of the 4th IEEE Workshop on Perception and Interactive Technologies for Speech-Based Systems. Kloster Irsee, Germany.
25. Gustafson, J., and Edlund, J. (2008) "EXPROS: a toolkit for exploratory experimentation with prosody in customized diphone voices" In Proc. of the 4th IEEE Workshop on Perception and Interactive Technologies for Speech-Based Systems. Kloster Irsee, Germany.
26. Gustafson, J., Heldner, M., and Edlund, J. (2008) "Potential benefits of human-like dialogue behaviour in the call routing domain" Proc. of the 4th IEEE Workshop on Perception and Interactive Technologies for Speech-Based Systems. Kloster Irsee, Germany.*
27. Strangert, E. and Gustafson, J. (2008) "Subject ratings, acoustic measurements and synthesis of good-speaker characteristics" In Proceedings of Interspeech 2008. Brisbane, Australia.
28. Bell, L. and Gustafson, J. (2007) "Children's convergence in referring expressions to graphical objects in a speech-enabled computer game", In Proceedings of Interspeech, Antwerp, Belgium.
29. Edlund, J., Heldner, M., and Gustafson, J. (2007) "Two faces of spoken dialogue systems" In M. F. McTear, K. Jokinen, J. Larson, R. López-Cózar and Z. Callejas (Eds.), Interspeech 2006 - ICSLP Satellite Workshop Dialogue on Dialogues: Multidisciplinary Evaluation of Advanced Speech-based Interactive Systems (pp. 51-54). Pittsburgh PA, USA.
30. Boye, J., and Gustafson, J. (2005) "How to do dialogue in a fairy-tale world", Proceedings of the sixth SIGdial Workshop on Discourse and Dialogue, Lisabon, 2005
31. Bell, L., Boye, J., Gustafson, J., Heldner, M., Lindström, A. and Wiren, M. (2005) "The Swedish NICE Corpus – Spoken dialogues between children and embodied characters in a computer game scenario", proceedings of Interspeech05, Lisabon.
32. Gustafson, J., Boye, J., Fredriksson, M., Johannesson, L., and Königsmann, J. (2005) "Providing computer game characters with conversational abilities," in Proc.of Intelligent Virtual Agent (IVA05). Kos, Greece.*
33. Boye, J., Mats Wiren, M., and Gustafson, J. (2004) "Contextual Reasoning in Multimodal Dialogue Systems: Two Case Studies", Proceedings of The 8th Workshop on the Semantics and Pragmatics of Dialogue Catalogue'04 , Barcelona, July 19-21, 2004.
34. Gustafson, J. and Sjölander, K. (2004) "Voice creation for conversational fairy-tale characters", Proceedings of the 5th ISCA Speech Synthesis Workshop, Carnegie Mellon University, 2004.*
35. Gustafson, J., Bell, L., Boye, J., Lindström, A. and Wiren, M. (2004) "The NICE Fairy-tale Game System", In proceedings of SIGdial 04, Boston.

36. Bell, L. and Gustafson, J. (2003) "Child and Adult Speaker Adaptation during Error Resolution in a Publicly Available Spoken Dialogue System", In proceedings of Eurospeech 03, Geneve, Schweiz.
37. Bell, L., Gustafson, J. and Heldner, M. (2003) "Prosodic adaptation in human–computer interaction", Proceedings of ICPhS 03, Bercelona, Spain.
38. Gustafson, J. and Sjölander, K. (2002) "Voice Transformations For Improving Children's Speech Recognition In A Publicly Available Dialogue System", Proceedings of ICSLP02, Colorado USA.
39. Gustafson, J., Bell, L., Boye, J., Edlund, J. and Wiren, M. (2002) "Constraint Manipulation And Visualization In A Multimodal Dialogue System", Proceedings of the ISCA Workshop Multi-Modal Dialogue in Mobile Environments Kloster Irsee, Germany
40. Bell, L, Boye, J, and Gustafson, J. (2001) "Real-time Handling of Fragmented Utterances", Proceedings of NAACL 2001.
41. Bell, L., Boye, J., Gustafson, J., and Wiren, M. (2000) "Modality Convergence in a Multimodal Dialogue System", Proceedings of Götalog 2000, Göteborg, Sweden.
42. Gustafson, J., Bell, L., Beskow, J., Boye, J., Carlson, R., Edlund, J., Granström, B., House, D. and Wirén M. (2000) "AdApt – a multimodal conversational dialogue system in an apartment domain", Proc. of ICSLP 2000.
43. Bell, L. and Gustafson, J. (2000) "Positive and Negative User Feedback in a Spoken Dialogue Corpus", In procedings of ICSLP 2000 Beijing, China.
44. Bell, L., Eklund, R. and Gustafson, J. (2000) "A Comparison of Disfluency Distribution in a Unimodal and a Multimodal Speech Interface", Proceedings of ICSLP 2000 Beijing, China.
45. Gustafson, J., Sjölander, K., Beskow, J., Granström, B. and Carlson, R. (1999) "Creating web-based exercises for spoken language technology", Invited tutoriaÎ session in proceedings of IDS'99, Kloster Irsee, Germany
46. Gustafson, J., Lundeberg, M. and Liljencrants, J. (1999) "Experiences from the development of August - a multimodal spoken dialogue system", in proceedings of , Kloster Irsee, Germany
47. Bell, L. and Gustafson, J. (1999) "Utterance types in the August System", in proceedings of IDS'99.
48. Bell, L. and Gustafson, J. (1999) "Repetition and its phonetic realizations: investigating a Swedish database of spontaneous computer directed speech", in Proc. of ICPhS' 99, Budapest, Hungary.
49. Bell, L. and Gustafson, J. (1999) "Interaction with an animated agent in a spoken dialogue system", in proceedings of Eurospeech '99 , Budapest, Hungary.
50. Gustafson, J., Lindberg, N., and Lundeberg, M. (1999) "The August spoken dialogue system", in proceedings of Eurospeech '99, Budapest, Hungary.
51. Sjölander, K., Gustafson, J., Beskow, J., Granström, B. and Carlson, R. (1999) "Web-based educational tools for speech technology", in proceedings of Matisse 99 (ESCA/SOCRATES Workshop on Method and Tool Innovations for Speech Science Education)
52. Gustafson, J., Elmberg, P., Carlson R. and Jönsson, A. (1998) "An Educational Dialogue System With a User Controllable Dialogue Manager", ICSLP98,Sydney, Australia.
53. Sjölander, K., Beskow, J., Gustafson, J., Lewin, E., Carlson, R., and Granström, B. (1998) "Web-based Educational Tools for Speech Technology", ICSLP98,Sydney, Australia.
54. Carlson, R., Granström, B., Gustafson, J., Levin, E. and Sjölander, K. (1998) "Hands-on Speech Technology on the Web", in proceedings of the ELSNET in Wonderland conference.
55. Gustafson, J., Larsson, A., Carlson, R. and Hellman, K. (1997) "How do System Questions Influence Lexical Choices in User Answers?", In proceedings of Eurospeech '97, Rhodes, Greece.
56. Sjölander, K. and Gustafson, J. (1997) "An Integrated System for Teaching Spoken Dialogue Systems Technology", In proceedings of Eurospeech '97, Rhodes, Greece.
57. Onomastica Consortium[1] (1995) “The ONOMASTICA interlanguage pronunciation lexicon”, in Proceedings of EuroSpeech’95, Madrid, Spain
58. Gustafson, J. (1995) "Using Two-level Morphology To Transcribe Swedish Names ", in Proceedings of Eurospeech '95, Madrid, Spain
59. Bertenstam, J. Blomberg, M., Carlson, R., Elenius, K, Granström, B., Gustafson, J., Hunnicutt, S., Högberg, J., Lindell, R., Neovius, L., de Serpa-Leitao, A., Nord, L. and Ström, N. (1995) "The Waxholm Application Data-Base", in proceedings of Eurospeech '95, in Madrid, 1995
60. Gustafson, J. (1995) "Transcribing names with foreign origin in the ONOMASTICA project", in proceedings of ICPhS'95 in Stockholm, 1995.
61. Carlson, R., Hunnicutt, S. and Gustafson, J. (1995) "Dialogue management in the Waxholm system", Proc. Spoken Dialogue Systems, Vigsø
62. Bertenstam, J. Blomberg, M., Carlson, R., Elenius, K, Granström, B., Gustafson, J., Hunnicutt, S., Högberg, J., Lindell, R., Neovius, L., de Serpa-Leitao, A., Nord, L. and Ström, N. (1995) "The Waxholm system - a progress report", Proc. Spoken Dialogue Systems, Vigsø
63. Blomberg, M., Carlson, R., Elenius, K, Granström, B., Gustafson, J., Hunnicutt, S., Lindell, R., and Neovius, L. (1993) "An experimental dialogue system: WAXHOLM,", Eurospeech '93, Berlin, Sept 21-23, 1993, Vol. 3,

Other conference contributions

[1] Anderson, O., Boves, L., Dalsgaard, P., Darsinos, V., Granström, B., Gustafson, J., van den Heuvel, H., Jack, M., Kokkinakis, G., Konst, E., Logothetis, M., Mascarenhas, I., Mengel, A., Molbaek, P., Ottensen, G., Pardo, J., Pirrelli, V., Schmidt, M., Sutherland, A., Trancoso, I., Valverde, F., Viana, C., & Yvon, F.

1. Al Moubayed, S., Skantze, G., Beskow, J., Stefanov, K., & Gustafson, J. (in press). Multimodal Multiparty Social Interaction with the Furhat Head. Demopaper, to be published in Proc. of the 14th ACM International Conference on Multimodal Interaction ICMI. Santa Monica, CA, USA. (won Outstanding Demo Award)
2. Neiberg, D., & Gustafson, J. (2012). "Towards letting machines humming in the right way - prosodic analysis of six functions of short feedback tokens in English" In Fonetik 2012. Göteborg, Sweden.
3. Johnson-Roberson, M., Bohg, J., Kragic, D., Skantze, G., Gustafson, J., & Carlson, R. (2010) "Enhanced Visual Scene Understanding through Human-Robot Dialog", In Proceedings of AAAI 2010 Fall Symposium: Dialog with Robots. Arlington.
4. Beskow, J., Edlund, J., Gustafson, J., Heldner, M., Hjalmarsson, A. & House, D. (2010) "Research focus: Interactional aspects of spoken face-to-face communication", Fonetik 2010, Lund, Sweden.
5. Schötz, S., Beskow, J., Bruce, G., Granström, B., Gustafson, J. & Segerup, M. (2010) "Simulating Intonation in Regional Varieties of Swedish", Fonetik 2010, Lund, Sweden.
6. Beskow, J., Edlund, J., Gustafson, J., Heldner, M., Hjalmarsson, A., & House, D. (2010) "Modelling humanlike conversational behaviour", Short paper in proceedings of SLTC 2010. Linköping, Sweden.
7. Gustafson, J., & Neiberg, D. (2010) "Directing conversation using the prosody of mm and mhm", Short paper in proceedings of SLTC 2010. Linköping, Sweden.
8. Edlund, J., Gustafson, J., & Beskow, J. (2010) "Cocktail - a demonstration of massively multi-component audio environments for illustration and analysis", Short paper in proceedings of SLTC 2010. Linköping, Sweden.
9. Skantze, G. & Gustafson, J., "Multimodal interaction control in the NonAMI Reminder", Demo description in proceedings of DiaHolmia 09.
10. Beskow, J., & Gustafson, J., "Experiments with synthesis of Swedish dialects", Poster abstract in proceedings of Fonetik 2009.
11. Beskow, J., Edlund, J., Granström, B., Gustafson, J., Jonsson, O., & Skantze, G. (2008) "Speech technology in the European project MonAMI" In Proceedings of FONETIK 2008 (pp. 33-36). Gothenburg, Sweden.
12. Gustafson, J., & Edlund, J. (2008) "EXPROS: Tools for exploratory experimentation with prosody" In Proceedings of FONETIK 2008 (pp. 17-20). Gothenburg, Sweden. (pdf)
13. Strangert, E., & Gustafson, J. (2008) "Improving speaker skill in a resynthesis experiment" In Proceedings of FONETIK 2008. Gothenburg, Sweden
14. Lindberg, N, & Gustafson, J (2000) "Example based shallow semantic analysis in the August spoken dialogue system", STL-QPSR 1/00
15. Bertenstam, J. Blomberg, M., Carlson, R., Elenius, K, Granström, B., Gustafson, J., Hunnicutt, S., Högberg, J., Lindell, R., Neovius, L., de Serpa-Leitao, A., Nord, L. & Ström, N. (1995):" Spoken dialogue data collection in the Waxholm project", STL-QPSR 1/1995, pp. 50-73.(pdf)
16. Gustafson, J. (1994). "ONOMASTICA - Creating a multi-lingual dictionary of European names", In FONETIK ´94, Papers from the 8th Swedish Phonetics Conference, May 24-26, Lund, Sweden, pp. 66-69.1-94
17. Blomberg, M., Carlson, R., Elenius, K, Granström, B., Gustafson, J., Hunnicutt, S., Lindell, R., Neovius, L. & Nord, L. (1993):" An experimental dialogue system: WAXHOLM,", STL-QPSR 2-3/1993, pp. 15-20.

Review articles, book chapters, books

1. Schötz, S., Bruce, B., Segerup, M., Beskow, J., Gustafson, J. and Granström, B. (2012) "Regional Varieties of Swedish: Models and Synthesis",In Niebuhr, Oliver and Pfitzinger, Hartmut (Eds.) "Prosodies. Context, Function, and Communication" festschrift in Honour of Klaus Kohler in Language, Context, and Cognition series.
2. Edlund, J. & Gustafson, J. (2010) "Ask the experts: Part I: Elicitation", in Allwood, J. & Juel Henrichsen, P (Eds.), Linguistic Theory and Raw Sound, Copenhagen Studies in Linguistics, spring 2010.
3. Edlund, J. & Gustafson, J. (2010) "Ask the experts: Part II: Analysis", in Allwood, J. & Juel Henrichsen, P (Eds.), Linguistic Theory and Raw Sound, Copenhagen Studies in Linguistics, spring 2010.
4. Beskow, J, Edlund, J, Granström, B, Gustafson, J. & House, D. (2010): "Face-to-face interaction and the KTH Cooking Show". In Esposito, Anna, Campbell, Nick, Vogel, Carl, Hussain, Amir, & Nijholt, Anton (Eds.), Development of Multimodal Interfaces: Active Listening and Synchrony . Berlin / Heidelberg: Springer, 2010.
5. Edlund, J., Heldner, M., and Gustafson, J. (2005) "Utterance segmentation and turn-taking in spoken dialogue systems". In B. Fisseni, H.-C. Schmitz, B. Schröder and P. Wagner (Eds.), Sprachtechnologie, mobile kommunikation und linguistische ressourcen (pp. 576-587). Frankfurt am Main, Germany: Peter Lang.*
6. Gustafson, J. (2002) "Developing Multimodal Spoken Dialogue Systems - Empirical Studies of Spoken Human-Computer Interactions", PhD thesis, KTH, Stockholm.
7. Gustafson, J. (1996):"A Swedish Name Pronunciation System", Licenciate thesis, TMH, KTH, 1996

3.4.1 Approved grants from Swedish research councils

Co-leader of the RJ application called "*Prosody in conversation*" (2010-2013). Mattias Heldner is the main project leader. (funding 5.150.00 SEK)

Main PI of the VR project called "*Introducing interactional phenomena in speech synthesis"* (2010-2013). (funding 2.055.000 SEK)

Co-leader of the VR.application called *"Incremental processing in multimodal conversational systems"* (funding 3.200.000 SEK). Gabriel Skantze is the main project leader.

3.4.1.1 Project funding from the Strategic Research Area ICT-The Next Generation (SRA/TNG):

Main PI for the project called "*Situated Audio-Visual Interaction with Robots*" (2010-2013). (funding 3.000.000 SEK)

3.4.2 Approved grants from EU

Co-leader of the KTH part IURO project (2010-2013). Coordinated by Technische Universität München (TUM). The consortium consists of TUM, KTH, Eidgenössische Technische Hochschule Zürich, Universität Salzburg, ACCREA Engineering. Björn is the KTH main PI. (KTH funding €626.960)

Main PI of the KTH part of the GetHomeSafe EU project (2011-2014). Coordinated by DFKI. The consortium consists of DFKI, Nuance, IBM, KTH,Daimler. (KTH funding €505.235)

KTH task leader of EIT ICT Labs KIC project Social Mobility in Action (2012). Activity leader DFKI. (KTH funding €20.000)

KTH task leader of EIT ICT Labs KIC project SAFE (2012). Activity leader DFKI. (KTH funding €42.000)

3.5 Other scientific achievements

3.5.1 Active participation in international conferences/workshops 03-10

I have since my PhD been part of the *program committee* of 37 workshops and conferences (Interspeech, SigDial, ACL, IJCAI, HLT-NAACL, SemDial, KRPD, ICMI, Coling, PIT and I*WSDS*)

This is a list of the last years PC assignments:

- **IWSDS 2009** *(International Workshop Series On Spoken Dialogue Systems Technology)*
- **ICMI-MLMI 2009** *(International Conference on Multimodal Interfaces / Machine Learning for Multimodal Interaction)*
- **KRPD 2009** *(Knowledge and Reasoning in Practical Dialogue Systems)*
- **ICMI-MLMI 2010**
- **ACL2010** *(Association for Computational Linguistics)*
- **SEMDIAL 2010** *(Workshop Series on the Semantics and Pragmatics of Dialogue)*
- **IWSDS 2010**
- **Interspeech 2010**
- **IWSDS 2011** *(Workshop on Paralinguistic Information and its Integration in Spoken Dialogue Systems)*
- **NODALIDA 2011** *(Nordic Conference on Computational Linguistics)*
- **SIGDIAL 2011** *(Annual Meeting of the Special Interest Group on Discourse and Dialogue)*
- **Interspeech 2011**
- **ICMI 2011** (International Conference on Multimodal Interaction)
- **KRPDS 2011**
- **ICRA 2012** (IEEE International Conference on Robotics and Automation),
- **The Listening Talker 2012** (an interdisciplinary workshop on natural and synthetic modification of speech in response to listening conditions)
- **ICASSP 2012** (International Conference on Acoustics, Speech, and Signal Processing)
- **ICRA 2012** (IEEE International Conference on Robotics and Automation.)
- **Interspeech 2012**
- **SIGDIAL 2012**
- **EMNLP 2012** (Conference on Empirical Methods in Natural Language Processing)
- **IWSDS 2012** (Workshop on Spoken Dialogue Systems: Towards a Natural Interaction with Robots, Knowbots and Smartphones)
- **IVA 2012 (**Workshop on Real-time Conversations with Virtual Agents)
- **ICMI 2012 (**ACM International Conference on Multimodal Interaction, Demo and Exhibition session)

I have been part of the *organizing committee* of:

- *The international conference on multimodal systems (ICMI08)* at Crete, October 2008.
- *The international SEMDIAL workshop (DiaHolmia)* in Stockholm, June 2009.
- *A special session at Interspeech'09 called "Active Listening & Synchrony",* in Brighton, September 2009.

3.5.4 Editorial/advisory board in international journals

I am member of the Editorial Board of the journal Speech Communication.

3.5.5 Reviewing assignments for journals

Speech Communication, regularly since 2004 (most recently 2012)
IEEE transactions on Multimedia, first time 2008
Computer Speech and Language, first time 2008
ACM Transactions on Speech and Language Processing, first time 2010
Cognitive Computation, first time 2012

3.5.6 Assignments as public examiner/opponent

Member of graduating committee for thesis defenses:
Yingying Huang, MDI/CSC/KTH, Supervisor Yngve Sundblad, June 10, 2010,
Andrzej Pronobis, CVAP/CSC/KTH, Supervisor Patric Jensfelt, June 10, 2011,
Alper Aydemir, CVAP/CSC/KTH, Supervisor Patric Jensfelt, June 10, 2011,

3.5.7 Assignments as outside expert

2012: Referee to the position as Assistant Professor (Lektor) in Informatik (Ref IDE 11/11) at Högskolan I Halmstad.

2012: In my capacity as scientific board in *the Centre for Next Generation Localisation* (CNGL) proposal sent to Science Foundation Ireland, I was part of the Mock Review Panel for CNGL Dress Rehearsal in Dublin 2/7 2012.

3.5.9 Exhibitions

The August exhibition at Stockholms akademiska forum 1998 to 1999

The FrittFram exhibition at the Science Museum in Stockholm 2000-2001

The TänkOm exhibition at the Telecom museum in Stockholm 2002-2004

The Robotville exhibition at the Science museum in London. 2011

Lecture and demonstration about the FurHat robotic head at Tällberg Forum. 2012

4.0 PEDAGOGICAL ACHIEVEMENTS

4.3 Own teaching

1996-2000

- Development of the spoken dialog system laboratory toolkit (Gulan), documentation and exercises.
- Responsible for the laboratory course parts of language technology courses at TMH, Linköping university, NADA and ELSNET summer schools
- Lecturer at speech technology courses at TMH, NADA and the ELSNET summer schools

2007-2012

- In charge of DT2112 (Speech technology), undergraduate course with Student from KTH, Stockholm University and Linköping University. My current task is to re-design the course so that it can become part of Master Programmes in Language Technology at Stockholm University, Uppsala University and Göteborgs University.
- In charge of the KTH part of a joint undergraduate course with Uppsala university (Automatisk analys och syntes av tal) for language technology students at Uppsala university. Next year this will be replaced with our DT2112.
- In charge of the Speech Technology PhD course in GSLT (Graduate School of Language Technology)
- Supervisor in a course on project management (EH1010 Elektroprojekt), undergraduate course.
- Lecturer in the undergraduate courses DH2413 (Avancerad grafik och interaction), DH2418 (Språkteknologi) and DT1174 (Ljud som informationsbärare) at KTH and the undergraduate course Språkteknologi/Natural Language Processing at Uppsala University.
- Lecturer and laboratory exercise responsible in the VISPP'08 summer school in Kuressaare, PhD course.
- Lecturer in the course "*Development of Multimodal Interfaces*", arranged by COST 2102 at Trinity College.

4.6 Pedagogical courses (18.5 hp)

- Grundläggande kommunikation och undervisning (GKU, 9L4001), 3hp KTH, 2006.
- Forskarhandledning (LH207V), 3 hp, 2007.
- Lärande och undervisning (LH201V) 7,5 hp, 2009-2010
- Föreläsningsretorik och presentationsutformning för högskolelärare, (LH211V), 5 hp, 2011

4.8 Popular scientific presentations

- A number of public demonstrations of the August system at Kulturhuset (1998-1999).
- A large number of internal presentations and demonstrations of speech technology at Telia's visions center for internal and external groups. (2001-2007)
- Presentation about speech technology for the staff at the Stockholm Telecom Museum (2003)
- Presentation about speech technology for children (Unga Spekulerar) at the Stockholm Technical Museum, (2004)
- Presentation of MonAmi at ID-dagarna, Swedish trade show on assistive technology, 1600 participants (2008)
- Interview about spoken dialogue systems in the TV program Hjärnkontoret (2009)
- Interview about the future of speech synthesis published at Skånes Taltidning web site (2011-06-10)
- Presentation at introduction day for all new students at the KTH media program at Tekniska Museet (2011)
- Presentation at Tällberg Forum about giving robots a face and voice (2012)

4.9 Supervision

4.9.1 Master Thesis students

Björn Söderström (IcePeak) Teleping - Automatic Evaluation of Spoken Dialogue Systems, 20 points 2008.
Kristian Ronge (VoiceProvide) Automatic Methods for Dialogue Classification and Prediction, 20 points 2008.
Miray Merkes (TMH) Methods for Eliciting spontaneous Dialogue Phenomena during recording of Speech Synthesis Voices, 20 points 2008-09.
Martin Johansson (TMH / IURO project) Managing computer-directed route directions, 20 points 2010-2011
Carl Flodeen (Language technology program Uppsala University) 20 points 2012.

4.9.2 PhD students that have examined

Anna Hjalmarsson, assistant supervisor, defended her thesis in September 2010
Jens Edlund, main supervisor, defended his thesis in March 2011
Daniel Neiberg, main supervisor, defended his thesis in September 2012

4.9.3 Current PhD students

Raveesh Meena, main supervisor
Catherina Oertel, main supervisor
Bajubabu Bollepalli, main supervisor
Martin Johansson, assistant supervisor

4.10 Pedagogical merits gained from outside the academic sphere.

As an industrial researcher at Telia I often held presentations about speech technology, for customers as well as colleagues. As part of the development of new speech applications, I usually had to educate many levels of management in TeliaSonera in order get it funded. During projects with Tekniska Museet I held several public lectures at the museum about speech technology, either for children or for the museum staff. As the project leader of the Swedish part of the EU project NICE I had a close collaboration with the programmers and graphical designers at the computer game company Liquid Media. This was a two-way pedagogical process, where I taught them about speech technology and dialogue design, and they taught me about computer game design and development.

5.0 OTHER ASSIGNMENTS

5.1 Administrative assignments

5.1.1 Experience from unit leadership; indicate size of unit and duration.

I am deputy department head (ställföreträdande avdelningsföreståndare) for the department of Speech, Music and Hearing since 2011-01-01. The group has about 35 employees. I was elected as deputy department head for one term of office, after which the plan is for me to take over as Department Head for another term of office. In my capacity of deputy Head of Department I assisted Head of Department, Prof Askenfelt, in two interviews with two candidates when we hired our current department secretary in 2011.

Together with Prof Granström and Ass Prof Beskow, I have authored the ads for PhD students on three occasions (September 2010, October 2011 and July 2012), where we got 90, 140 and 120 applications that we jointly reviewed to select candidates for interview. In 2010 Prof Granström participated in the interviews, but since then it is my and Ass Prof Beskow's responsibility to interview the PhD applicants. So far we have been involved in the hiring 5 new PhD students, and we are currently interviewing 6-9 candidates for 2 new PhD positions.

5.1.2 Membership of university boards or councils in the last 5 years.

I am part of the CSC strategic Council.

I am part of the KTH International Advisory Group (IAG)

I was the speech group representative in the coordinating board of the CSC *Human Communication* platform

5.1.3 Other professional assignments of an administrative nature.

I am responsible for the fire detection at TMH (Brandskyddsansvarig).
I am responsible for the introduction of new PhD students at the department(Iintroduktör)
Part of the reference group for the CSC communication plan