Application

2015-03-31

2015-05456	Beskow, Jonas	NT-2

Information about applicant

Name: Jonas Beskow

Doctorial degree: 2003-06-11

Birthdate: 19700227

Academic title: Docent

Gender: Male

Employer: No current employer

Administrating organisation: Kungliga Tekniska högskolan

Project site: TMH, Tal, musik och hörsel

Information about application

Call name: Forskningsbidrag Stora utlysningen 2015 (Naturvetenskap och teknikvetenskap)

Type of grant: Projektbidrag

Focus: Fri

Subject area:

Project title (english): Attention and back-channelling in situated human robot interaction

Project start: 2016-01-01

Project end: 2019-12-31

Review panel applied for: NT-2

Classification code: 10208. Språkteknologi (språkvetenskaplig databehandling), 10207. Datorseende och robotik (autonoma system)

Keywords: Human robot interaction, Visual focus of attention, Gaze, Multi-party interaction, Non-verbal behaviours

Funds applied for

Year:	2016	2017	2018	2019
Amount:	1,368,000	1,458,000	1,814,000	1,949,000

Descriptive data

Project info

Project title (Swedish)*

Uppmärksamhet och återkoppling i situerad människa robot interaktion

Project title (English)*

Attention and back-channelling in situated human robot interaction

Abstract (English)*

We are witnessing a revolution in personal robotics. Once associated with heavy industrial applications and highly specialised manufacturing tasks, robots are rapidly becoming part of our everyday lives. The promise and potential of these systems is far-reaching; from co-worker robots that operate side-by-side with humans, via household and service robots that are able to carry out or assist in daily chores to robotic tutors in schools that interact with human learners in and around a shared environment. These scenarios require systems that are able to recognize and convey attention to objects in the environment and should ideally leverage channels of communication that humans understand. Human face-to-face interaction involves sophisticated mechanisms for conveying intentions, establishing common ground and acknowledging information transfer, using a combination of spoken and visual (non-verbal) signals.

The purpose of the proposed project is to enable situated conversational systems such as robots to better engage in collaborative interaction with humans. In the project we will study and implement critical non-verbal behaviours such as joint attention, mutual gaze and backchannels in situated human-robot collaborative interaction. This project aims to investigate fundamentals of situated and collaborative multi-party interaction and collect the data and knowledge required to build technical systems (e.g. social robots) that are able to handle collaborative attention and co-present interaction in a fluent way.

The research will follow an experiment-driven iterative process. By first establishing the technical platform for data collection and collaborative human-robot interaction and then utilising this platform for experimental research, we will enable cross-fertilisation between on one hand applied research on technologies for human-robot collaboration and social robotics and on the other hand research into human communicative behaviour. The project will make use of a testbed based on an existing setup that involves the Furhat robot head – an avatar/robot-hybrid where cues such as eye gaze, head movements, facial expression and speech are available and controllable with high accuracy - in combination with a multi-party spoken dialogue system and a touch table application that allows the participants and the robot to engage in collaborative problem solving using a combination of spoken language and non-verbal actions.

Through the use of motion capture and gaze tracking and other sensing technology, we will make a series of experiments/data collections, where data at each stage will be used to build statistical models of non-verbal behaviours that can be implemented in the robot at a later stage in the process. We will start with highly accurate data collection of human-human interaction and proceed to real-time mediated human-robot interaction, followed by fully autonomous interaction. In the mediated interaction stage, we will take advantage of the possibility to perform controlled experiments through a technique known as transformed social interaction (TSI), which means that some aspect of the communication is manipulated (distorted, supplemented, deleted, inverted, delayed or otherwise transformed) in order to study the effect of that particular aspect. We expect the project to result in new models, data and knowledge about effective behaviours in multiparty, co-present, collaborative embodied and situated human-human and human-robot interaction, as well as a test bed in which we can test new hypotheses and validate theories about such behaviours.

Popular scientific description (Swedish)*

Mänskligt samtal ansikte mot ansikte är vår mest grundläggande kommunikationsform. Sådana samtal innehåller ett noggrant orkestrerat samspel mellan olika talade och icke-verbala uttryck, så som huvudrörelser, blickar, ansiktsuttryck och minspel, gester samt det gemensamma rummet där samtalet tar plats. Signalerna innehåller information som är nödvändig för att ett samtal ska flyta problemfritt, speciellt om fler än två parter är inblandade:
Information om när interaktionen börjar och slutar, när en tur börjar och slutar, deltagarnas roller i samtalet (vem är talare, vem är mottagare etc), graden av delad information/förståelse, de olika parternas engagemang, vad/vem som står i centrum för uppmärksamheten etc. Dessa och flera signaler kräver att deltagarna är närvarande på samma fysiska plats vid samma tid. Effektiv flerpartsinteraktion en förutsättning för att i framtiden kunna använda humanoida robotsystem i en uppsjö verkliga tillämpningar som till exempel robotar som arbetar tillsammans med människor, robotar som hemhjälp, sällskap eller inom undervisningen i skolan.
Forskningsprojektet som presenteras här syftar till att utforska och utveckla nya metoder för att modellera viktiga beteenden i människa-robot-interaktion, med målet att samarbete mellan människor och robotar ska fungera så smidigt som möjligt. Ett sådant beteende hos en robot gäller var fokus för uppmärksamheten ska ligga vid varje tidpunkt och hur roboten den bäst visar det. I ett första steg kommer vi samla in exempel på mänskligt beteende genom inspelning av flerpartsinteraktioner med hjälp av rörelsemätningsteknik (motion capture & gaze tracking). Dessa exempel kommer sedan ligga till grund för statistisk modellering av dessa beteenden. Vi kommer även att bygga upp en experimentmiljö för människa-robot-interaktion där en persons interaktionsbeteende i detalj kopieras till en robot (med hjälp av en kombination av animering och robot-teknik) vilket gör det möjligt att utföra kontrollerade experiment i en interaktion genom att manipulera eller byta ut vissa beteenden. I en sådan miljö blir det möjligt att utvärdera enskilda beteenden och även dra slutsatser om vilka beteenden som är viktigast att modellera, och förhoppningsvis även förstå lite mer om villkoren för effektiv mänsklig kommunikation och samarbete. Slutligen kommer de resulterande modellerna att byggas in i ett helt autonomt system för social människa-robot-interaktion.

Project period

Number of project years*

4

Calculated project time*

2016-01-01 - 2019-12-31

Classifications

Select a minimum of one and a maximum of three SCB-codes in order of priority.

Select the SCB-code in three levels and then click the lower plus-button to save your selection.

SCB-codes*

1. Naturvetenskap > 102. Data- och informationsvetenskap (Datateknik) > 10208. Språkteknologi (språkvetenskaplig databehandling)

1. Naturvetenskap > 102. Data- och informationsvetenskap (Datateknik) > 10207. Datorseende och robotik (autonoma system)

Enter a minimum of three, and up to five, short keywords that describe your project.

Keyword 1*

Human robot interaction

Keyword 2*

Visual focus of attention

Keyword 3*

Gaze

Keyword 4

Multi-party interaction

Keyword 5

Non-verbal behaviours

Research plan

Ethical considerations

Specify any ethical issues that the project (or equivalent) raises, and describe how they will be addressed in your research. Also indicate the specific considerations that might be relevant to your application.

Reporting of ethical considerations*

The proposed project will not gather or store data that is classified as sensitive and personal according to legislation as stated in 13§ and 21§ of PUL. Nonetheless, audio-visual recordings of human conversation will be made and stored for subsequent analysis and modelling, and these data will be handled with great care and sensitivity and will only be released outside the research team with prior written consent from the subjects.

The project includes handling of personal data

Yes

The project includes animal experiments

No

Account of experiments on humans

Yes

Research plan

Attention and back-channelling in situated human robot interaction

Purpose and aims.

During recent years, we have witnessed the start of a revolution in personal robotics. Once associated with heavy industrial applications and highly specialised manufacturing tasks, robots are rapidly starting to become part of our everyday lives. The promise and potential of these systems is far-reaching; from general purpose co-worker robots that operate and collaborate with humans side-by-side, via household and service robots that are able to carry out or assist in daily chores to robotic tutors in schools that interact with humans in and around a shared environment. All of these scenarios require systems that are able to recognize and convey attention to objects in the environment. Hayes and Scasselati (2011) identify *bi-directional intent recognition* as a critical skill to be addressed in human-robot interaction research, and in particular suggest that robots should *leverage channels of communication that humans understand*. Human face-to-face interaction involves sophisticated mechanisms for conveying intentions, establishing common ground and acknowledging information transfer, often based on a combination of spoken and visual (non-verbal) signals.

The premiere purpose of the proposed project is to enable situated conversational systems to better engage in collaborative interaction with humans. In the project we will study and implement critical non-verbal behaviours such as joint attention, mutual gaze and backchannels in situated human-robot collaborative interaction. The project will investigate, in a data-driven manner, what non-verbal signals are the most effective to establish common ground in a collaborative multi-party human-robot interaction scenario (one robot, one or more humans), as well as how these signals should be manifested in a social robot to best support the interaction. In addition, the proposed project will produce a data set that may serve to deepen our understanding of the intricacies of such interactions, as well as an advanced test-bed for experiments in collaborative human robot interaction.

Co-presence and multi-party interaction.

Successful multi-party interaction is known to include a wide range of phenomena that rely on the participants being physically present: ***Addressee selection*** - any given utterance can be directed at any one or more of the participants. ***Next speaker selection*** - care must be taken to learn not only when a speaker change is appropriate, but also who is an appropriate next speaker. ***Mutual gaze*** – two participants looking at each other - occurs frequently and is an important interactional social cue. ***Joint attention*** - participants directing their attention towards the same object is a fundamental grounding mechanism.

All of the above phenomena require *participants of the interaction to be co-present* - that is present in the same physical space at the same time, i.e. a typical video conferencing setup will not be able to handle most of these phenomena in a graceful way, nor will a virtual agent present on a 2D screen. This project aims to investigate fundamentals of *situated and collaborative multi-party interaction* and collect the data and knowledge required to build technical systems (e.g. social robots) that are able to handle collaborative attention and co-present interaction in a fluent way.

Physically embodied avatars.
The proposed research will be carried out using a technological premise that we will refer to as a physically embodied avatar. This represents the most recent research direction by the PI and his team. Initially sparked by the needs for flexible modelling and display of human-like signals and social cues (e.g. eye gaze) in situated human-robot-interaction, the team has developed the *Furhat* social robotic head where an animated face is projected onto a face-shaped semi transparent mask. *Furhat* has proven capable of exhibiting verbal and non-verbal cues in a manner that is in many ways superior to on-screen avatars. At the same time it has many advantages over mechatronic robot heads in terms of high facial expressivity, very low noise level and fast animation allowing e.g. for highly accurate lip-synch. The team has successfully been exploring this technology in the context of social dialogue, involving simulation and modelling of may aspects of face-to-face conversation including multi-party attention control, turn taking signals and active listening behaviours.

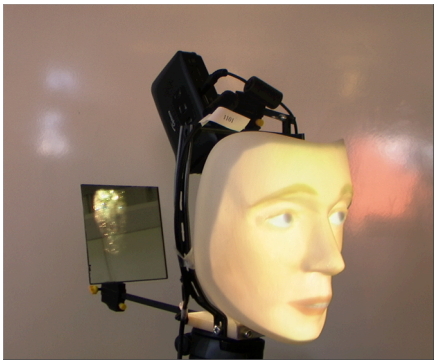

Figure 1. The "Furhat" animated avatar/robot hybrid will take the role of conversational partner in the proposed experiments.

Experiment-driven iterative research paradigm.
We propose an advanced technical collaborative test bed for situated multi-party interaction, using the physical avatar/robot-hybrid *Furhat* (see fig. 1) where cues such as the gaze, head movements, facial expression and speech are available and controllable with high accuracy. Furhat is already closely integrated with a spoken dialogue system for multi-party interaction, and the current project will make use and extend our latest experimental setup where the robot is combined with a multi-touch table and an application that allows which allows users to engage in collaborative problem solving/game playing together with the robot, using a combination of touch-based manipulation and spoken interaction (see *Preliminary work* below). In addition to this, we will employ state-of-the-art capture, sensing and display technologies, including gaze-tracking and motion capture to record and mediate human communicative behaviours. We will be able to collect real multi-party interaction data, and use the collected data to train/simulate human behaviours in a robot. This setup will offer unique possibilities of investigating the role of a multitude of non-verbal behaviours, such as attention and mutual gaze, in collaborative multi-party interaction, in a variety of settings on a scale from *human-human* to *human-autonomous robot*; we will have the ability to remote control the robot directly by the performance of a remote user, which will allow us to carry out controlled experiments by manipulating individual cues in face-to-face interaction, like replacing human behaviours with autonomous ones, for example by changing gaze targets, head motion or facial cues such as eyebrow motion, in real time. This will allow us to measure the relative importance of different cues in mediating co-presence as well as test and validate hypotheses on human interactional behaviour.

Survey of the field

Performance in social situations depends fundamentally on the ability to detect and react to the multiparty and multimodal social signals that underpin human communication. These processes are complex, with coordination being achieved by means of rich and varied signals including verbal and – crucially – non-verbal information (Bavelas & Gerwing, 2011, Kendon, 1967, Duncan, 1972, Allwood, 2008).

Over the past centuries, there has been a large interest in how such human communicative behaviours may simulated and exploited in human-computer interaction through embodied conversational agents (ECAs; c.f. Cassell et al, 2000; Cassell et al 2001; Gratch et al, 2007; Ter Maat, Truong & Heylen, 2011) and more recently, in human-robot interaction. Huang and Tomaz (2011) found that a robot ensuring joint attention is perceived as having better performance and a more natural behaviour. A similar conclusion is reached by Mehlmann et al (2014), who also propose a model for how such behaviours may be implemented for single user interactions using a rule-driven approach. Mutlu et al (2013) studied the relation between speech and gaze for establishing listener/speaker roles in human robot interaction.

Many studies employ some form of mediated interaction as a tool for carrying out controlled investigations of the causal effects of different factors on interaction. Computer mediation makes it possible to manipulate interactions in systematic ways, in real time, while maintaining a high degree of ecological validity (Loomis et al., 1999). Edlund & Beskow (2009) describe this paradigm as online manipulation of human-human interaction, and it can be seen as an instance of what Bailenson et al. (2004) call transformed social interaction – a general framework describing the decoupling of actual behaviours in an interlocutor from the ones presented to other participants. In addition to our own work, examples of studies developing this paradigm include Schlangen & Fernández (2007) who manipulated the speech channel in dialogue by blurring speech segments to elicit phenomena such as clarification requests. Fernández et al. (2007) also experimented with restricting interlocutors' ability to speak whenever they want. Bente et al. (2007) investigated the effects of gaze on person perception in a sophisticated setup, again involving avatars.

Robotics-based systems for telepresence systems typically fall into one of two categories: humanoid/android systems and robotic telepresence systems based on video communication. The Geminoid HI-2 (Sakamoto et al, 2007) is a tele-operated humanoid that closely resembles a human that has been used for telepresence experiments. The smaller Telenoid (Ogawa et al, 2011) is intended to provide telepresence without conveying gender or identity. Humanoid fully mechatronic robotic systems are however associated with high cost, mechanical complexity and inevitable motor noise, which limits practical applicability. The *Furhat* system (see *physical avatar* above) that will be used in the studies proposed in this project overcomes these limitations in a simple yet effective by replacing most of the mechatronics with projected computer animation.

Project description

This research project will follow an *experiment-driven iterative process*. By first establishing the technical platform for data collection and collaborative human-robot interaction and then utilising this platform for experimental research, we will enable cross-fertilisation between on

one hand applied research on technologies for human-robot collaboration and social robotics and on the other hand research into human communicative behaviour.

Experimental setup

The technical platform builds on an existing setup that involves the *Furhat* robot head, in combination with a multi-party spoken dialogue system and a touch table application that allows the participants and the robot to engage in collaborative problem solving using a combination of spoken language and non-verbal actions (touch, gaze etc). The touch table multi-party setup allows the participants to have a meaningful discussion with each other even in cases where Furhat's understanding of the spoken interaction is limited. The symmetry of the setting allow us to compare the human behaviour towards each other with their behaviour towards the robot in order to (1) use the data as a model for Furhat's behaviour, (2) investigate to what extent they interact with the robot as if it was a human interlocutor, and (3) evaluate how human-like the robot's current behaviour is, and how it could be improved. An important feature of this setup is that it involves discussion about objects in the physical space, where the interlocutors' visual focus of attention (VFOA) must be shared between each other and the objects under discussion. All user behaviours will be captured and recorded, including speech, head & upper body motion, eye gaze and touch table events; we will make use of existing infrastructure (motion capture and head-worn gaze trackers) for this.

In addition to this, we will extend the setup to allow for full remote co-presence, i.e. establish a bi-directional link that allows for a human to control the *Furhat* robot using in an advanced low-latency capture-synthesis chain. In this case, the robot functions as a physical avatar - an embodiment of a human participant that is not present in the room. The process is analogous to video conferencing, but creates a physically embodied recreation of the remote participant rather than a video image: the remote person's actions (e.g. speech, gesture, eye-gaze) is captured and resynthesized in near-real-time in the robot, and the visual field of the robot is copied to the remote participants field of view through a wide-angle display or a virtual reality headset. This setup will allow us to not only record every relevant aspect of multi-party face-to-face human-robot interaction, but also optionally transform selected parts of it's constituent cues before re-synthesizing the behaviours on the robot/physical avatar.

This will make it possible to supplement or replace the behaviours of the mediated human by automated behaviours, opening up for a host of experiments based on transformed social interaction (TSI) for validation of the automated behaviours: if a communication channel such as eye gaze is "taken over" by the automatic system, what are the measurable effects on the interaction? – if the automatic model performs well, we would expect little impact on the interaction, while a poor model could even cause communication to break down.

Iterative process

At the start of the project, we will record collaborative task-oriented human interactions in two different physical interaction settings: a problem solving task/game where the participants collaborate around physical objects (or virtual, on a touch table), and another setting where they collaborate without any objects; only through spoken interaction. From these data, statistical models will be built to account for the observed behaviours based on events in the interaction. These statistical behavioural models will in turn be evaluated in controlled experiments which also constitute the second data collection; they are carried out in the same

settings as the initial recordings, but with one of the participants replaced by a robot, which will be either remote controlled by a human, or have some – or all - of it's behaviours controlled by automatic models. Data from the second recording will in turn feed into the second modelling phase, which in turn feeds into the next experimentation phase. As such, the research project will follow a series of sequential cycles all containing various elements of data collection, technical development, data-driven model building and evaluation. During the four years of the project we envision to go through the following phases:

Phase 1: Human data collection

The first stage of the project involves data collection of human behaviour in the specified settings. We will record three-way interactions where the participants are involved in a task-oriented dialogue. In one condition, participants will collaborate around a touch-table application in order to sort a set of items in the correct order based on some given criterion. In another condition, they will interact only verbally, but given a task of corresponding complexity to solve. All participants will be recorded using motion capture equipment, gaze tracker, audio and video.

Phase 2: Basic models of attention and acknowledgement

In this phase the data from phase 1 is used to build statistical models of the observed behaviours. The goal is to model and predict processes of multi-party interaction based on available sensor inputs in such a way that active listener behaviours can be generated autonomously in the robot. This involves generation of feed-back and back-channels in response to other participant's speech, displaying attention (turning head and shifting the gaze) towards the current speaker etc. Visual focus of attention (VFOA) is manifested through a combination of head motion and eye gaze, and these need to be precisely timed with respect to each other. Back-channel behaviours are typically a combination of vocalisations (*u-hu*, *m-hm* etc) and or visual cues (gaze – often mutual gaze, and gestures such as head nods). The modelling will be done without taking semantics of the dialogue into account, and will only be based on directly measurable data from the other participants (voice activity level, VFOA, touch table events etc.).

Phase 3: Experimentation – mediated, semi-automatic & automatic behaviour

During this phase, the experimental platform will be established. The touch table application from phase 1 will be complemented with an autonomous robot and a multi-party spoken dialogue system (pre-existing, see preliminary results). We will also establish the mediated co-presence capture and re-synthesis setup that allows the robot to be tele-operated. The setup will allow for real-time, low latency streaming of data from the remote location hardware sensors and implement these cues in the local setting using the robot.

Using this platform we will carry out a series of trials in order to verify automatic behaviour models from phase 2. In these trials, two humans interact with the robot, which will be either fully controlled by a remote participant, or semi-automatic, where some of the behaviours of the robot (e.g. VFOA) is driven by automatic behaviours based on said models. The quality of the interaction will be monitored, and compared between the different interaction settings. We will do similar tests where the robot is in fully autonomous mode, i.e.

driven by the multi-party spoken dialogue system. These trials will serve as evaluation of the accuracy and effectiveness of the models, since they make it possible to compare various metrics of the dialogue success between human and automatic control. They will also serve as a source of interactional data for the next round of modelling.

Phase 4: Refined models

In this phase we use data from previous phase to build more refined models of the interactive behaviours. Whereas the models developed in phase 2 disregard any semantic information of the speech signal, this will now be taken into account in (e.g. mentioning objects present in the interactional space will have impact on where VFOA will be directed) In combination with this, the models should include a basic handling of joint attention, i.e. it should give special significance to objects that are receiving VFOA from two or more of the participants. Another cue that will be modelled in more detail is mutual gaze, which is known to have an important function in back-channels and turn regulation.

Phase 5: Exp phase 2 - transformed social interaction

In the last phase of the project, the full setup will be utilized and autonomous behaviour models will be employed and evaluated in online experimentation. Transformed social interaction (TSI) means that some aspect of the communication is distorted, supplemented, deleted, inverted, delayed or otherwise transformed in order to study the effect of that particular aspect. In these experiments we will replace specific behaviours of the human with automated behaviours, and study the impact on the communicative interaction.

Significance

Interfaces that leverage channels of communication that humans understand are identified as crucial for successful applications in collaborative human robot interaction. The project will result in a host of new knowledge of multiparty, co-present, collaborative embodied and situated human-human and human-robot interaction that has so far been inaccessible; a system that teaches us how co-present spoken interfaces can be used and utilized and what aspects of human likeness are worth working on in future spoken social robotics and avatar systems, as well as a test bed in which we can test new hypotheses and validate theories about effective communicative behaviours in multiparty collaborative interaction.

Preliminary results

This project builds directly and indirectly on a host of research by the PI and the interdisciplinary KTH team, which has a long history of building computational models of human conversational behavior that we evaluate in spoken dialogue systems (Edlund et al., 2008, Gustafson et al., 2008). In a previous project by the PI (VR 2010-4646) a 60 hours corpus of synchronized audio, video, and three-dimensional motion capture data in unconstrained human-human dyadic conversations was used in the study of non-verbal speaking and listening behaviour (Alexanderson et al, 2013a, Alexanderson et al, 2013b), as well as high-fidelity facial motion capture recordings and re-animation of different speaking styles (Alexanderson & Beskow, 2014).

The proposed project leverages the recent disruptive development in the area of embodied conversational agents, where the Furhat system represents the move of a virtual agent into the physical world using projected computer facial animation allowing cues such as gaze, head movements, facial expression and speech to be controllable with high accuracy (Al Moubayed et al., 2013). In combination with state-of-the-art spoken dialogue capabilities for multi-party interaction based on a flexible state-chart formalism (Skantze & Al Moubayed, 2012) this provides the framework that puts the research proposed in this project into a directly applicable use case scenario. The most recent development, that that forms the testbed of the current proposal, is based around a multi-party human robot problem solving/gameplay scenario (see figure 2). In this setup, two users interact with the robot and a touch table in order to collaboratively sort a set of items according to a specified criterion (Skantze, Johansson & Beskow, submited). In November 2014 the system was exhibited at the Stockholm Science Museum, where 800 users interacted with the system during 9 days. A video can be found at https://www.youtube.com/watch?v=5fhjuGu3d0I

Figure 2. The collaborative card sorting game for human robot interaction at the Stockholm Science Museum. Nov. 2014.

Equipment

The researchers in the programme will have access to the technical infrastructure at KTH School for Computer Science and Communication, which includes the *Performance and Interaction Lab* (PMIL) with audio, video, motion capture and gaze-tracking equipment. The *KTH Visualization Studio* provides further infrastructure for high-speed/high resolution video conferencing and Virtual Reality headsets. KTH Speech, Music and Hearing, where the programme will be carried out, already has the projection based avatar/robot-hybrid system as well as an immersive projection screen (J-Dome).

International and national collaboration

The speech group at KTH, where Jonas and his team are active, has a history of prominent research in the field of spoken and multimodal communication. At KTH, they have a close collaboration on human-robot interaction with CVAP (Danica Kragic) and on mediated communication with Media Technology and Interaction (Haibo Li). The international network includes informal cooperation and project collaboration with several prominent research groups including Institut Telecom, Paris (Catheine Pelachaud); Univ. Twente (Dirk Heylen), Univ. college London (Andrew Faulkner), University of Southern California (David Traum), Univ. Bielefeld (Stefan Kopp)., to name but a few. The Speech Group at KTH has been evaluated as part of the KTH International Research Assessment Exercise in 2008 and 2012. In RAE 2008 the evaluators commented the speech group with: *The relationship between basic research and applied research is outstanding as is the collaboration with industry* and *This is an outstanding, world leading research group - among the top and most respected (a national asset)*. In RAE2012 the unit of Applied Computer Science, of which the speech group is part, was selected as one of the most excellent at KTH: *Research output is internationally excellent in all fields, with a substantial number of units reaching the level of world-leading quality*.

Independent line of research

This project will be carried out in a research group with a long-standing interest in humanlikeness and social signals in computational and robotic systems and the visionary goal to build systems that interact like humans do. This is a collaborative effort where the Furhat robot head - which itself was a product of the research pursued by the PI and PhD student *S. Al Moubayed* - serves as an important test bed and demonstrator system where human-robot interaction technologies developed by different researchers in different projects are implemented and evaluated, which has resulted in a current demonstrator system capable of fully autonomous multi-party spoken social human-robot interaction. In particular the autonomous spoken dialogue capabilities of this system have been developed by Ass. Prof. *Gabriel Skantze* in a series of projects (e.g. VR 2011-6237), as well as the touch table interaction application that forms the foundation of the test bed in the current project. The line of research pursued by the PI focuses on core visual, multimodal and non-verbal aspects and behaviours in spoken interaction, measured and validated through motion capture, statistical modelling, animation technologies and perceptual experiments and is thus concerned less with the linguistic and semantic aspects of dialogue. That said, the true potential of this research is fully realized when the non-verbal and multimodal aspects are combined with autonomous spoken dialogue behaviours, as is the proposed project.

References

Al Moubayed, S., Skantze, G., & Beskow, J. (2013). The Furhat Back-Projected Humanoid Head - Lip reading, Gaze and Multiparty Interaction. International Journal of Humanoid Robotics, 10(1).

Alexanderson, S., House, D., & Beskow, J. (2013a). Aspects of co-occurring syllables and head nods in spontaneous dialogue. In *Proc. of 12th International Conference on Auditory-Visual Speech Processing (AVSP2013)*. Annecy, France.

Alexanderson, S., House, D., & Beskow, J. (2013b). Extracting and analyzing head movements accompanying spontaneous dialogue. In *Proc.Tilburg Gesture Research Meeting*. Tilburg University, The Netherlands.

Alexanderson, S., & Beskow, J. (2014). Animated Lombard speech: Motion capture, facial animation and visual intelligibility of speech produced in adverse conditions. Computer Speech & Language, 28(2), 607-618.

Allwood, J. (2008). Dimensions of embodied communication - towards a typology of embodied communication. In Wachsmuth, I., Lenzen, M., & Knoblich, G. (Eds.), Embodied Communication in Humans and Machines. Oxford: Oxford University Press.

Bailenson, J. N., Beall, A. C., Loomis, J., Blascovich, J., & Turk, M. (2004). Transformed social interaction: decoupling representation from behavior and form in collaborative virtual environments. Presence: Teleoperators & Virtual Environments, 13(4), 428-441.

Bavelas, J. B., & Gerwing, J. (2011). The listener as addressee in face-to-face dialogue. International Journal of Listening, 255(3), 178-198.

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (2000). Embodied conversational agents. Cambridge, MA: MIT Press.

Cassell, J., Vilhjálmsson, H. H., & Bickmore, T. (2001). BEAT: the behavior expression animation toolkit. In Proceedings of the 28th annual conference on Computer graphics and interactive techniques (pp. 477-486).

Duncan, S. (1972). Some Signals and Rules for Taking Speaking Turns in Conversations. Journal of Personality and Social Psychology, 23(2), 283-292.

Edlund, J., & Beskow, J. (2009). MushyPeek - a framework for online investigation of audiovisual dialogue phenomena. Language and Speech, 52(2-3), 351-367.

Edlund, J., Gustafson, J., Heldner, M., & Hjalmarsson, A. (2008). Towards human-like spoken dialogue systems. Speech Communication, 50(8-9), 630-645.

Edlund, J., Heldner, M., & Hirschberg, J. (2009). Pause and gap length in face-to-face interaction. In Proc. of Interspeech 2009. Brighton, UK.

Fernández, R., Schlangen, D., & Lucht, T. (2007). Push-to-talk ain't always bad! Comparing Different Interactivity Settings in Task-oriented Dialogue. In Proceedings of the 11th Workshop on the Semantics and Pragmatics of Dialogue (pp. 25-31). Trento, Italy.

Gratch, J., Wang, N., Gerten, J., Fast, E., & Duffy, R. (2007). Creating rapport with virtual agents. In Proceeding of the 7th International Conference on Intelligent Virtual Agents. Paris.

Gustafson, J., Heldner, M., & Edlund, J. (2008). Potential benefits of human-like dialogue behaviour in the call routing domain. In Proceedings of Perception and Interactive Technologies for Speech-Based Systems (PIT 2008) (pp. 240-251). Berlin/Heidelberg: Springer.

Hayes, B., & Scassellati, B. (2013). Challenges in shared-environment human-robot collaboration. learning, 8, 9.

Heylen, D., Theune, M., op den Akker, R., & Nijholt, A. (2009). Social agents: the first generations. In Proc. of the 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, 2009 (ACII 2009) (pp. 1.7). Amsterdam.

Huang, C. M., & Thomaz, A. L. (2011, July). Effects of responding to, initiating and ensuring joint attention in human-robot interaction. In *RO-MAN, 2011 IEEE*(pp. 65-71). IEEE.

Kendon, A. (1967). Some functions of gaze direction in social interaction. Acta Psychologica, 26, 22-63.

Loomis, J. M., Blascovich, J. J., & Beall, A. C. (1999). Immersive virtual environment technology as a basic research tool in psychology. Behavior Reseach Methods, Instruments and Computers, 31(4), 557-564.

Mehlmann, G., Häring, M., Janowski, K., Baur, T., Gebhard, P., & André, E. (2014, November). Exploring a Model of Gaze for Grounding in Multimodal HRI. In *Proceedings of the 16th International Conference on Multimodal Interaction*(pp. 247-254). ACM.

Mutlu, B., Terrell, A., & Huang, C. M. (2013). Coordination mechanisms in human-robot collaboration. In *Proceedings of the Workshop on Collaborative Manipulation, 8th ACM/IEEE International Conference on Human-Robot Interaction*.

Ogawa, K., Nishio, S., Koda, K., Balisteri, G., Watanabe, T. and Ishiguro, H. (2011); Exploring the natural reaction of young and aged person with Telenoid in a real world". Journal of Advanced Computational Intelligence and Intelligent Informatics, vol 15, no 5, pp 592-597.

Sakamoto, D., Kanda, T., Ono, T., Ishiguro, H and Hagita, N. (2007): Android as a telecommunication medium with a human-like presence, in Proceedings of the ACM/IEEE Conference on Human-Robot Interaction (HRI'07), pp. 193-200.

Schlangen, D., & Fernández, R. (2007). Speaking through a noisy channel: experiments on inducing clarification behaviour in human-human dialogue. In Proceedings of Interspeech 2007. Antwerp, Belgium.

Skantse, G., Johansson, M. & Beskow, J. (submitted): A collaborative human-robot card sorting game as a test-bed for modelling multi-party, situated interaction. Submited to Intelligent Virtual Agents 2015.

Skantze, G., & Al Moubayed, S. (2012). IrisTK: a statechart-based toolkit for multi-party face-to-face interaction. In Proceedings of ICMI. Santa Monica, CA.

Ter Maat, M., Truong, K. P., & Heylen, D. (2011). How agents' turn-taking strategies influence impressions and response behaviors. *Presence: Teleoperators and Virtual Environments*, *20*(5), 412-430.

Interdisciplinarity

My application is interdisciplinary

☐

An interdisciplinary research project is defined in this call for proposals as a project that can not be completed without knowledge, methods, terminology, data and researchers from more than one of the Swedish Research Councils subject areas; Medicine and health, Natural and engineering sciences, Humanities and social sciences and Educational sciences. If your research project is interdisciplinary according to this definition, you indicate and explain this here.

Click here for more information

Scientific report

Scientific report/Account for scientific activities of previous project

---

Budget and research resources

Project staff

Describe the staff that will be working in the project and the salary that is applied for in the project budget. Enter the full amount, not in thousands SEK.

Participating researchers that accept an invitation to participate in the application will be displayed automatically under Dedicated time for this project. Note that it will take a few minutes before the information is updated, and that it might be necessary for the project leader to close and reopen the form.

Dedicated time for this project

	Role in the project	Name	Percent of full time
1	Applicant	Jonas Beskow	25

Salaries including social fees

	Role in the project	Name	Percent of salary	2016	2017	2018	2019	Total
1	Applicant	Jonas Beskow	25	212,000	217,000	223,000	228,000	880,000
2	Other personnel without doctoral degree	Kalin Stefanov	100	595,000	654,000			1,249,000
3	Other personnel with doctoral degree	N/N Post doc	100			856,000	942,000	1,798,000
	Total			807,000	871,000	1,079,000	1,170,000	3,927,000

Other costs

Describe the other project costs for which you apply from the Swedish Research Council. Enter the full amount, not in thousands SEK.

Premises

	Type of premises	2016	2017	2018	2019	Total
1	arbetsrum/lokal	97,000	105,000	130,000	141,000	473,000
	Total	97,000	105,000	130,000	141,000	473,000

Running Costs

	Running Cost	Description	2016	2017	2018	2019	Total
1	Travel	conferences/workshops	30,000	30,000	30,000	30,000	120,000
2	Computers		15,000		15,000		30,000
	Total		45,000	30,000	45,000	30,000	150,000

Depreciation costs

	Depreciation cost	Description	2016	2017	2018	2019

Total project cost

Below you can see a summary of the costs in your budget, which are the costs that you apply for from the Swedish Research Council. Indirect costs are entered separately into the table.

Under Other costs you can enter which costs, aside from the ones you apply for from the Swedish Research Council, that the project includes. Add the full amounts, not in thousands of SEK.

The subtotal plus indirect costs are the total per year that you apply for.

Total budget

Specified costs	2016	2017	2018	2019	Total, applied	Other costs	Total cost
Salaries including social fees	807,000	871,000	1,079,000	1,170,000	3,927,000		3,927,000
Running costs	45,000	30,000	45,000	30,000	150,000		150,000
Depreciation costs					0		0
Premises	97,000	105,000	130,000	141,000	473,000		473,000
Subtotal	949,000	1,006,000	1,254,000	1,341,000	4,550,000	0	4,550,000
Indirect costs	419,000	452,000	560,000	608,000	2,039,000		2,039,000
Total project cost	1,368,000	1,458,000	1,814,000	1,949,000	6,589,000	0	6,589,000

Explanation of the proposed budget

Briefly justify each proposed cost in the stated budget.

Explanation of the proposed budget*

Löner:
* PI 25% genomgående för effektiv styrning och aktivt deltagande i forskningen
* Doktorand, Stefanov: kommer ägna projektet sina två sista år av forskarutbildningen. Har varit mycket aktiv i framtagandet av de preliminära resultat som ligger till grund för projektet
* Post-doc. Till de två sitsta åren av projektet kommer en extern post-doc tas in för att slutföra.

Driftskostnader: resor och konferens/publiceringsavgifter 30 tkr/år. Två bärbara datorer á 15 tkr (2016 och 2018)

Other funding

Describe your other project funding for the project period (applied for or granted) aside from that which you apply for from the Swedish Research Council. Write the whole sum, not thousands of SEK.

Other funding for this project

Funder	Applicant/project leader	Type of grant	Reg no or equiv.	2016	2017	2018	2019

CV

Jonas Beskow - CV

1. Higher education degree
1995 M.Sc. in Electrical Engineering, KTH, Stockholm.

2. Doctoral degree
2003 Ph.D. in Speech communication,"Talking heads – models and applications for multimodal speech synthesis" (Supervisor: B. Granström). KTH, Stockholm.

3. Postdoctoral position
1998-1999 Pre-doc at University of California SantaCruz, Perceptual Science Lab

4. Docent qualification
2010 Docent, KTH

5. Present position
2011-03 - Associate professor (Universitetslektor), KTH Speech music and hearing (80% research, 20% teaching)

6. Previous positions
- 2010-11 – 2011-03: Researcher, KTH Centre for speech technology
- 2004-11 –2010-10: Assistant Professor (Forskarassistent), KTH Speech music and hearing
- 2003-07 – 2004-10: Researcher, KTH Centre for speech technology
- 1999-96 – 2003-06: Research engineer, KTH Speech music and hearing;
- 1998-01 – 1999-06: Guest researcher, Perceptual Science Lab, University of California, Santa Cruz
- 1995-01 – 1998-01: Research engineer, KTH Speech music and hearing.

7. Interruption in research
- 2007-02 – 2007-08: Parental leave

8. Supervision
Samer Al Moubayed (PhD 2012-12); Gunilla Svanfeldt (Tekn Lic. 2006, co-supervisor); Laura Enflo (Tekn Lic 2010, co-supervisor) - **Ongoing:** Simon Alexandersson (PhD Student since 2011-01);Kalin Stefanov (PhD student since 2012-04); Bajibabu Bollipalli (PhD student since 2012-08, co-supervisor)

9. Other information of relevance to the application
Jonas research is in the domain of analysis, modelling and synthesis of human communicative behaviour, primarily in spoken interaction, with applications in assistive technology, virtually and physically embodied conversational agents and social robotics. The basic research philosophy that underlies Jonas work is to allow the engineering perspective to co-exist with a more humanistic view grounded in curiosity on human communicative behaviour. This balance is inherent in the KTH speech group, but it was additionally fostered during his PhD years when he had the opportunity to spend 18 months at the Perceptual Science Lab at University of California Santa Cruz - one of the pioneering labs when it comes to applied engineering research in the study of human behaviour, in this particular case computer facial animation used as an investigative tool in multimodal speech perception research.

With the advent of new paradigms in human-computer interaction based on human face-to-face interaction, as well as an increased interest in mediated human-human communication for more efficient distance collaboration, the importance of being able to analyse, model, process and synthesize both verbal and non-verbal communicative cues has become evident. Jonas' team is actively pursuing different aspects of synthesis and modelling of human non-verbal communication for a variety of applications. Facial animation and visual speech is an important part of this, but the scope has been extended in several directions. The VR-funded project "Large-scale massively multimodal modelling of non-verbal behaviour in spontaneous dialogue" (VR 2010-4646, 2011-2013) used the rich *Spontal* data set, involving motion capture, audio- and video data for 60 hours of human face-to-face conversations, in order to analyse and model different phenomena in human-human dialogue. Jonas is also recently finished the Tivoli-project (PTS 2011-2013) focusing on sign-language synthesis (signing avatars) and recognition of sign language, applied to language training for children, in a game-like environment.

Physically embodied avatars represent the most recent research direction, which is also reflected in the current research project proposal. Initially sparked by the needs for flexible modelling and display of human-like signals and social cues (e.g. eye gaze) in situated human-robot-interaction, the team has developed the "Furhat" social robotic head where an animated face is projected onto a face-shaped semi transparent mask. Furhat has proven capable of exhibiting verbal and non-verbal cues in a manner that is in many ways superior to his virtual on-screen counterparts. The team has successfully been exploring this technology in the context of social dialogue, involving simulation and modelling of may aspects of face-to-face conversation including multi-party attention control, turn taking signals and active listening behaviours. The current research project proposal suggests to take this ground-breaking research to a new level by building a world unique setup for collaborative multiparty interaction, and to use this platform both for experimental basic research on human face-to-face interaction behaviour, as well as for applied research into new solutions for human-robot co-presence and collaboration.

Jonas is an active member of the international academic community, and is frequently engaged in conference and workshop programme committees and as peer reviewer for journals. Some specific notable engagements: Member of the editorial board, Journal on Multimodal User Interfaces, Member of the executive board of International Speech Communication Association Special Iterest Group on Audio-Visual Speech Communication (ISCA-SIG AVSP), Member of the Organizing Commitée and Workshop Chair of 13:th International Conference on Intelligent Virtual Agents (IVA 2013), Edinburgh, UK, Organizer and conference chair of the 11:th International Conference on Auditory-Visual Speech Processing (AVSP 2011), Volterra, Italy.

He has received the following awards and prizes: Fulbright Scholarship Award, 1998 (Fulbright Commision); Chester Carlssons Forskningspris, 2006 (Ingenjörsvetenskapsakademien & Stiftelsen Xerox svenska fond för forskning i informationsvetenskap); Outstanding Demo Award, ICMI 2012 Santa Monica - *Multimodal Multiparty Social Interaction with the Furhat Head* (with S. Al Moubayed, J. Gustafsson, G. Skantze and K. Stafanov); Robotdalen Innovation Award 2013 2:nd prize - *Furhat the social robot* (with G. Skantze & S. Al Moubayed)

Publication list

Jonas Beskow – Bibliography

The number of citations was obtained using *Google Scholar*, which lists **149 publications** for Jonas Beskow with a total of **2691 citations** and an **H-index of 27**.
https://scholar.google.se/citations?user=G33Tyb0AAAAJ&hl=sv&oi=sra

1. Peer-reviewed articles

* Alexanderson, S & Beskow, J (2014). Animated Lombard Speech: Motion capture, facial animation and visual intelligibility of speech produced in adverse conditions. Computer Speech & Language. *(Number of citations: 1)*

Al Moubayed, S., Skantze, G., & Beskow, J. (2013). The Furhat Back-Projected Humanoid Head - Lip reading, Gaze and Multiparty Interaction. International Journal of Humanoid Robotics. 10(1). *(Number of citations: 14)*

Mirnig, N., Weiss, A., Skantze, G., Al Moubayed, S., Gustafson, J., Beskow, J., Granström, B., & Tscheligi, M. (2013). Face-to-Face with a Robot: What do we actually talk about?. International Journal of Humanoid Robotics, 10(1). *(Number of citations: 3)*

* Al Moubayed, S., Edlund, J., & Beskow, J. (2012). Taming Mona Lisa: communicating gaze faithfully in 2D and 3D facial projections. ACM Transactions on Interactive Intelligent Systems, 1(2), 25. *(Number of citations: 23)*

* Al Moubayed, S., Beskow, J., & Granström, B. (2010). Auditory-Visual Prominence: From Intelligibility to Behavior. Journal on Multimodal User Interfaces, 3(4), 299-311. *(Number of citations: 13)*

* Salvi, G., Beskow, J., Al Moubayed, S., & Granström, B. (2009). SynFace—Speech-Driven Facial Animation for Virtual Speech-Reading Support. EURASIP Journal on Audio, Speech, and Music Processing, 2009. *(Number of citations: 17)*

* Edlund, J., & Beskow, J. (2009). MushyPeek - a framework for online investigation of audiovisual dialogue phenomena. Language and Speech, 52(2-3), 351-367. *(Number of citations: 20)*

Beskow, J., Engwall, O., Granström, B., Nordqvist, P., & Wik, P. (2008). Visualization of speech and audio for hearing-impaired persons. Technology and Disability, 20(2), 97-107. *(Number of citations: 3)*

2. Peer-reviewed conference contributions

Al Moubayed, S., Beskow, J., & Skantze, G. (2014). Spontaneous spoken dialogues with the Furhat human-like robot head. In HRI'14. Bielefeld, Germany.

Alexanderson, S., House, D., & Beskow, J. (2013). Aspects of co-occurring syllables and head nods in spontaneous dialogue. In Proc. of 12th International Conference on Auditory-Visual Speech Processing (AVSP2013). Annecy, France.

Alexanderson, S., House, D., & Beskow, J. (2013). Extracting and analyzing head movements accompanying spontaneous dialogue. In Proc.Tilburg Gesture Research Meeting. Tilburg University, The Netherlands. Al Moubayed, S., Beskow, J., & Skantze, G. (2013). The Furhat Social Companion Talking Head. In Interspeech 2013 - Show and Tell. Lyon, France.

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. Proc of LREC Workshop on Multimodal Corpora. Istanbul, Turkey. (Number of citations: 6)

Edlund, J., Alexandersson, S., Beskow, J., Gustavsson, L., Heldner, M., Hjalmarsson, A., Kallionen, P., & Marklund, E. (2012). 3rd party observer gaze as a continuous measure of dialogue flow. In proc. of LREC 2012. Istanbul, Turkey.

Edlund, J., Beskow, J., Elenius, K., Hellmer, K., Strömbergsson, S., & House, D. (2010). Spontal: a Swedish spontaneous dialogue corpus of audio, video and motion capture. In Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odjik, J., Piperidis, S., Rosner, M., & Tapias, D. (Eds.), Proc. of the Seventh conference on International Language Resources and Evaluation (LREC'10) (pp. 2992 - 2995). Valetta, Malta. (Number of citations: 21)

Al Moubayed, S., & Beskow, J. (2010). Prominence Detection in Swedish Using Syllable Correlates. In Interspeech'10. Makuhari, Japan.

Al Moubayed, S., & Beskow, J. (2009). Effects of Visual Prominence Cues on Speech Intelligibility. In Proceedings of Auditory-Visual Speech Processing AVSP'09. Norwich, England.

Al Moubayed, S., Beskow, J., Öster, A., Salvi, G., Granström, B., van Son, N., & Ormel, E. (2009). Virtual Speech Reading Support for Hard of Hearing in a Domestic Multi-media Setting. In Proceedings of Interspeech 2009.

Beskow, J., Salvi, G., & Al Moubayed, S. (2009). SynFace - Verbal and Non-verbal Face Animation from Audio. In Proceedings of The International Conference on Auditory-Visual Speech Processing AVSP'09. Norwich, England.

Beskow, J., Bruce, G., Enflo, L., Granström, B., & Schötz, S. (2008). Recognizing and Modelling Regional Varieties of Swedish. In Proceedings of Interspeech 2008. (number of citations: 1)

Beskow, J., Granström, B., Nordqvist, P., Al Moubayed, S., Salvi, G., Herzke, T., & Schulz, A. (2008). Hearing at Home – Communication support in home environments for hearing impaired persons. In Proceedings of Interspeech 2008. Brisbane, Australia. (number of citations: 4)

Edlund, J., & Beskow, J. (2007). Pushy versus meek – using avatars to influence turn-taking behaviour. In Proceedings of Interspeech 2007. Antwerp, Belgium. (number of citations: 9)

5. Books and book chapters

Al Moubayed, S., Beskow, J., Bollepalli, B., Hussen-Abdelaziz, A., Johansson, M., Koutsombogera, M., Lopes, J., Novikova, J., Oertel, C., Skantze, G., Stefanov, K., & Varol, G. (2014). Tutoring Robots: Multiparty multimodal social dialogue with an embodied tutor. In Proceedings of eNTERFACE2013. Springer.

Edlund, J., Al Moubayed, S., & Beskow, J. (2013). Co-present or not? Embodiement, situatedness and the Mona Lisa gaze effect. In Nakano, Y., Conati, C., & Bader, T. (Eds.), Eye Gaze in Intelligent User Interfaces - Gaze-based Analyses, Models, and Applications. Springer.

Al Moubayed, S., Beskow, J., Skantze, G., & Granström, B. (2013). Furhat: A Back-projected Human-like Robot Head for Multiparty Human-Machine Interaction. In Esposito, A., Esposito, A., Vinciarelli, A., Hoffmann, R., & C. Müller, V. (Eds.), Cognitive Behavioural Systems. Lecture Notes in Computer Science. Springer.

Beskow, J., Carlson, R., Edlund, J., Granström, B., Heldner, M., Hjalmarsson, A., & Skantze, G. (2009). Multimodal Interaction Control. In Waibel, A., & Stiefelhagen, R. (Eds.), Computers in the Human Interaction Loop. Berlin/Heidelberg: Springer. (cited by 4)

Beskow, J., & Cerrato, L. (2008). Evaluation of the Expressivity of a Swedish Talking Head in the Context of Human-Machine Interaction. In Magno Caldognetto, E., Cavicchio, E., & Cosi, P. (Eds.), Comunicazione Parlata e manifestazione delle emozioni. (number of citations: 5)

Beskow, J., Granström, B., & House, D. (2007). Analysis and synthesis of multimodal verbal and non-verbal interaction for animated interface agents. In Esposito, A., Faundez-Zanuy, M., Keller, E., & Marinaro, M. (Eds.), Verbal and Nonverbal Communication Behaviours (pp. 250-263). Berlin: Springer-Verlag.

6. Patents

Massaro, Dominic W., Michael M. Cohen, and Jonas Beskow. Visual display methods for in computer-animated speech production models. U.S. Patent No. 7,225,129. 29 May 2007.

7. Open access computer programs

WaveSurfer (Beskow & Sjölander) is an Open Source tool for sound recording, visualization, annotation and manipulation. http://sourceforge.net/projects/wavesurfer/ (35000 downloads last year)

CV

Name: Jonas Beskow	**Doctorial degree:** 2003-06-11
Birthdate: 19700227	**Academic title:** Docent
Gender: Male	**Employer:** No current employer

Research education

Dissertation title (swe)

Dissertation title (en)

Talking Heads - models and applications for multimodal speech synthesis

Organisation	Unit	Supervisor
Kungliga Tekniska Högskolan, Sweden Sweden - Higher education Institutes	TMH, Tal, musik och hörsel	Björn Granström

Subject doctors degree	ISSN/ISBN-number	Date doctoral exam
10208. Språkteknologi (språkvetenskaplig databehandling)	91-7283-536-2	2003-06-11

Publications

Name: Jonas Beskow	**Doctorial degree:** 2003-06-11
Birthdate: 19700227	**Academic title:** Docent
Gender: Male	**Employer:** No current employer

Beskow, Jonas has not added any publications to the application.

Register

Terms and conditions

The application must be signed by the applicant as well as the authorised representative of the administrating organisation. The representative is normally the department head of the institution where the research is to be conducted, but may in some instances be e.g. the vice-chancellor. This is specified in the call for proposals.

The signature*from the applicant*confirms that:

- the information in the application is correct and according to the instructions form the Swedish Research Council
- any additional professional activities or commercial ties have been reported to the administrating organisation, and that no conflicts have arisen that would conflict with good research practice
- that the necessary permits and approvals are in place at the start of the project e.g. regarding ethical review.

The signature *from the administrating organisation* confirms that:

- the research, employment and equipment indicated will be accommodated in the institution during the time, and to the extent, described in the application
- the institution approves the cost-estimate in the application
- the research is conducted according to Swedish legislation.

The above-mentioned points must have been discussed between the parties before the representative of the administrating organisation approves and signs the application.

Project out lines are not signed by the administrating organisation. The administrating organisation only sign the application if the project outline is accepted for step two.

Applications with an organisation as applicant is automatically signed when the application is registered.