PhD position in Articulatory measurements and modeling
at the Centre for Speech Technology, KTH, Sweden

Background
The work will be carried out within ASPI – audiovisual-to-articulatory speech inversion, a project funded by the European Union’s 6th Framework Program with partners in LORIA, Nancy, France; ENST, Paris, France; ULB, Brussels, Belgium and ICCS-NTU, Athens, Greece.

Audiovisual-to-articulatory inversion consists in recovering the dynamics of the vocal tract shape from the acoustical speech signal and image analysis of the speaker's face. Being able to recover this information automatically would be a major break-through in speech technology, e.g. for pronunciation training in computer-assisted language learning.

Objectives
The work aims at providing a flexible acquisition technique and associated processing methods in order to train and test articulatory inversion. The set-up will combine various imaging and sensor techniques that bring complementary information on the articulation either on a temporal or a spacial point of view.

The first objective is to design a measurement set-up consisting of ultrasound, electromagnetic tracking and stereovision in collaboration with LORIA, Nancy, France, where an identical measurement set-up will be used. The second is to successfully use the data collected with the acquisition set-up to improve a dynamic three-dimensional model of speech production.

Suitable background of applicants
Previous experiences in engineering, computing, phonetics, medical imaging, image analysis or articulatory measurements, as well as an excellent level of English are desirable.

Duration of employment
3 years with possibility for prolongation.

Salary
The salary will follow KTH PhD salaries, i.e. a monthly base salary of 20200 SEK (approx. 2200 EUR) increasing to 24600 SEK (approx. 2650 EUR) before graduation.

Starting date
Flexible, but preferably as soon as the application process and the successful applicant allow for.

Applications
Applications should be sent electronically to Olov Engwall olov@speech.kth.se and Björn Granström bjorn@speech.kth.se as soon as possible. We will begin reviewing applications by November 15th 2005, but accept applications until the position is filled.

Applications should include a CV, a detailed resume and name and contact information of two references (e.g. Master thesis supervisor, head of department).

Questions and information
Olov Engwall, olov@speech.kth.se, +468-790 75 65
Centre for Speech Technology, KTH
Lindstedtstr. 24
SE-100 44 Stockholm, SWEDEN

A description of the work can be found on http://www.speech.kth.se/vacancies
**Description of work**

**Defining a multimodal acquisition set-up**

In the ASPI project, Magnetic Resonance Imaging (MRI) and ultrasound will be the most important imaging modalities to study the internal articulators. MRI has proved to be efficient for tongue modeling of sustained articulations. However, the static MRI data need to be complemented with dynamic data, in order to generate a model representative of dynamic speech. Ultrasound is currently the most attractive modality for such measurements.

Unfortunately, the ultrasound technique can not measure the tip of the tongue whenever there is an air cavity between the transducer and a raised tongue body. Electromagnetic sensors are hence needed in order to complement the dynamic information brought by the ultrasound data. These sensors have small dimensions and can be fixed on the tongue tip, allowing the motion of the apex to be recovered with a good accuracy.

Finally, a stereovision system with two video cameras will be used to image the external articulators and for visual calibration.

The first task of the PhD student will be to define, set up and test the multimodal acquisition equipment, in collaboration with the supervisors and colleagues at LORIA, Nancy, France. The evaluations will focus on sensor accuracy and the ability to image the entire vocal tract, including e.g. the tongue tip and the palate.

**Synchronization and calibration of data streams**

The various imaging and sensor modalities in the acquisition set-up need to be spatially and temporally aligned. As the various image/sensor devices are not referenced the same way, an important task is to combine the image/sensor modalities in a common reference frame. In addition, as sounds must often be uttered several times for the acquisition, temporal speech alignment must be performed to compensate for speaking rate differences and to ensure that data corresponding to the same speech segment are collected during repetitions.

The second task of the PhD student is to plan and perform calibration and synchronization experiments with the different modalities.

**Recovering the articulatory model**

Once all the acoustic, optical and ultrasound data are registered, the main objective is to build a dynamic articulatory model based on our present wire-frame talking head (refer to the PhD theses by Beskow, 2003 and Engwall, 2002 available from http://www.speech.kth.se/ctt/publications).

The PhD student will work on curve tracking in ultrasound images, defining a tracking method to follow tongue movements in data from ultrasound images and magnetic sensors and combine the articulatory movements with three-dimensional MRI data in a model in order to create a articulatorily correct speech production model.

This three-dimensional model and the collected data will be used for articulatory inversion experiments and evaluation within the ASPI project.