A look at 2001

Anders Askenfelt
Chairman

The year 2001 was in many respects a good year for the Department of Speech, Music and Hearing. Our research was as active as ever with numerous conference presentations, journal articles and exchange of guests. The number of research students increased to 42. Our teaching in undergraduate courses at KTH is steadily increasing, albeit from a low volume. *Pattern Recognition*, *Speech Technology*, and *Information Theory and Source Coding* are examples of courses that reflect central areas of the research activity at the department, attracting more and more students. The responsibility for teaching the basics of sound and speech communication in the new Media Technology program at KTH has exposed the activities within the department to new groups of students.

The financial situation has improved considerably. After a couple of years with weak finances the department could look back at the closing of the accounts for 2001 with satisfaction. A positive net result of 2.5 MSEK was reached with a turnover of 36 MSEK.

Funding of research is a complicated affair with certain streaks of fortune-like components. The competition at national and international research agencies and councils is razor-sharp, and the outcome may differ between years, despite similar levels of effort as regards the quality and number of applications. The outcome for 2001 was more successful than for the preceding years. The department now has a balanced portfolio of research funding, including Swedish research councils, industry, and EU projects. The high number of projects funded by the European Commission at the department (11 projects running during 2001, and 5 new approved for 2002), is a gratifying indicator that research at our department well meets the international quality marks.

The positive financial net at the end of 2001 is not only due to increased income. It is also the result of a number of actions to

cut the costs at the department without a detrimental influence on the research and teaching activities.

During 2001, four departments at KTH working in areas related to mobile telecommunications organized their activities in *Wireless@KTH*, a research and education center in wireless applications supported by a group of prominent industrial partners, including Ericsson, Nokia, Microsoft and Telia. The speech-processing group at our department has joined this center. The number of centers at KTH, bridging the traditional organization in departments, is steadily increasing. This trend reflects the need for expertise in more than one area to reach strategic research goals, and meet the demands from industry. In what way the increased activities at cross-departmental centers will influence the future organization of KTH remains to be seen.

The expectations for the growth in telecommunications, in particular via more advanced use of mobile Internet including pictures and animations, have been reduced during 2001. The general trend in the forecasts now seems to be that the dominating use of mobile phones during the next years still will be voice communication and SMS-messages. This is partly due to technical and financial constraints, leading to delays in making new technology accessible to the consumer. Awaiting the widespread use of mobile Internet and other wireless applications, we can take the opportunity to reflect on the strengths of spoken communication. These strengths seem to be so pronounced that the dominance of unsupported speech in telecommunications may well prevail longer than expected.

Our ability to communicate with other people using our speech exceeds by far our other senses, like touch and vision. We can easily produce speech sounds at rate of 10 sounds/s in normal speech, a rate which is well adapted to our hearing. In passing we may note that fast-speaking people like sport referees may exceed the normal speaking rate by as much as a factor of five. The redundancy in normal speech is high, around 80%, giving information of about 1 bit per speech sound on the average. In all, speech gives us access to a comfortable and robust information exchange person-to-person at a typical rate of 10 b/s. This is faster than pressing buttons, for most people much faster. Due to the high redundancy in speech, disturbances like noisy environments, negligent articulation habits, and more than one speaker at the same time, do not seriously reduce the information in the conveyed message.

A dialogue between human beings is enhanced by vision. Facial expression and body language may be important supplements to the spoken message. This does, however, not detract from the basic quality and communicative power of voice communication alone. The importance of vision in supporting the spoken message is particularly evident when talking to a computer. An animated agent that takes the role of your speaking partner facilitates the communication with the machine by adding a human component to the, so far, rather straightforward dialogues. In contrast, the unexpected twists and turns that usually appear in dialogues between humans may make the spoken-only conversation over the telephone an enduring experience. Lacking the support from vision, the message "between the lines" – communicated via prosody – becomes a most essential component of the dialogue.

Spoken man-machine dialogues, which have been one main research topic at our department during the last ten years, can consequently be foreseen to attract continued interest. In everyday life we need to control machines such as computers, cars, vacuumcleaners, refrigerators, audio equipment and the like. In many cases we would like to do that by speaking to them instead of pressing buttons, the function of which often becomes evident after trying only. Voice control will be particularly apt when addressing a multi-purpose machine like a robot. We may also like to be able to use the whole palette of human expression, including emotional behaviour. For example, we may want to stress the importance of a request, "Turn off the water!", or speed up an initiated process, for example bringing more coffee. Having addressed the machine we would like to get a sensible answer, if needed, which opens up for a dialogue. A speaking car, which fails to start at the beginning of a busy day, would provide an interesting test situation for such dialogue systems. Whatever the

future applications may be, the fact is that our voice is our fastest output organ, and our natural way of communicating is by speaking.

The reflections above on speech in human communication illustrate why speech and its applications and processing will continue to be core areas in the future research at our department. The speech technology and speech processing groups work in fast-developing areas with a high commercial potential, involving many actors from academia as well as industry.

As stressed in earlier annual reports, the breadth of the research within the department makes it kind of unique. The hearing technology group studies the capacity of the normal and impaired ear for detecting and discriminating between sounds in general, and speech sounds in particular. The music acoustics group studies musical communication, which shares many of its general features with speech. In all, the common interests and interactions between the groups make us a true Department of Speech, Music, and Hearing. Read more about the groups on the following pages, and meet us also for current news, scientific articles, and upcoming events at http://www.speech.kth.se.

Filename:	Alookat.doc
Directory:	I:\annual\Annual2001
Template:	J:\tmh.kth.se\tmh\proj\common\template\tmh\Normal.dot
Title:	A look at 2001
Subject:	
Author:	andersa
Keywords:	
Comments:	
Creation Date:	2002-05-22 16:01
Change Number:	2
Last Saved On:	2002-05-22 16:01
Last Saved By:	Personal
Total Editing Time:	3 Minutes
Last Printed On:	2002-06-12 10:35

As of Last Complete Printing

Number of Pages:	4
Number of Words:	1 113 (approx.)
Number of Characters:	6 346 (approx.)