

Dept. for Speech, Music and Hearing
**Quarterly Progress and
Status Report**

**Speech research in China -
impressions from a visit**

Carlson, R. and Granström, B.

journal: STL-QPSR
volume: 21
number: 4
year: 1980
pages: 001-013



**KTH Computer Science
and Communication**

<http://www.speech.kth.se/qpsr>

I. SPEECH RESEARCH

A. SPEECH RESEARCH IN CHINA - IMPRESSIONS FROM A VISIT

R. Carlson and B. Granström

Abstract

During a three week visit to The People's Republic of China (Oct 20 - Nov 7, 1980) we had the opportunity to visit some institutes of acoustics, linguistics and physiology engaged in different aspects of speech research. We spent most of the time in Beijing but the program also included some days in Nanjing, Shanghai and Hangzhou.

We saw many examples of both advanced and successful projects. Considerable effort is spent on reorganizing the scientific work, which in most places was severely upset during the so-called cultural revolution. The lack of equipment still is felt to be an obstacle to scientific progress but the situation is rapidly improving. Presently the biggest problem is to make up for the "lost decade" in education.

Introduction

China has, since the mid sixties, appeared rather isolated from the western world. In the last few years a drastic change has been observed. This has been especially apparent in science and education. Chinese attendance at international conferences has increased and many formal agreements on scientific exchange have been signed.

Such an agreement now also exists between the Swedish Academy of Engineering Science (IVA) and the Chinese Academia Sinica. As a result of this agreement, we got the possibility to visit China during 3 weeks in October/November 1980. The objective was to study speech research in China, to inform about such activities in Sweden and also to look into the possibilities of further cooperation in this area of research.

Our visit was organized by Academia Sinica in Beijing. There we spent most of our time in the Institute of Acoustics of Academia Sinica, but we also visited the Institute of Linguistics in the Chinese Academy of Social Sciences, and we met several researchers from other institutes and universities. After Beijing we visited the Acoustic Institute in the University of Nanking and the Institute of Physiology of Academia Sinica in Shanghai.

The Chinese speech scientists are very motivated and curious about what is going on abroad. In Beijing we gave three lectures at the Acoustic Institute on the subjects "Speech Research in Sweden", "Perception and Production of Prosodic Phenomena in Speech" and "The Relation between Writing and Pronunciation in a Text-to-Speech System". All the lectures were well attended also by researchers outside the Institute. In the last lecture we presented a preliminary text-to-speech system for Chinese, developed at our Institute. It accepts the Pinjin spelling as input and produces standard Chinese as output. No group in China is presently working on speech synthesis but there is much of interest in starting such activities. Such a project, especially with Chinese character input certainly involves many fascinating problems.

Traditionally, research in China is separated from education in a way similar to the system in the Soviet Union, where research is carried out in the institutes of the academies and with universities almost entirely focussed on education. However, there appears to be an effort in China to develop a system where academies and universities share the responsibility for both education and research. This could be seen by the personal communication between the two organizations and the use of teachers from the academies at the universities. Thesis work at the academies could also be carried out by students at the universities.

The Institute of Acoustics, Academia Sinica, Beijing.

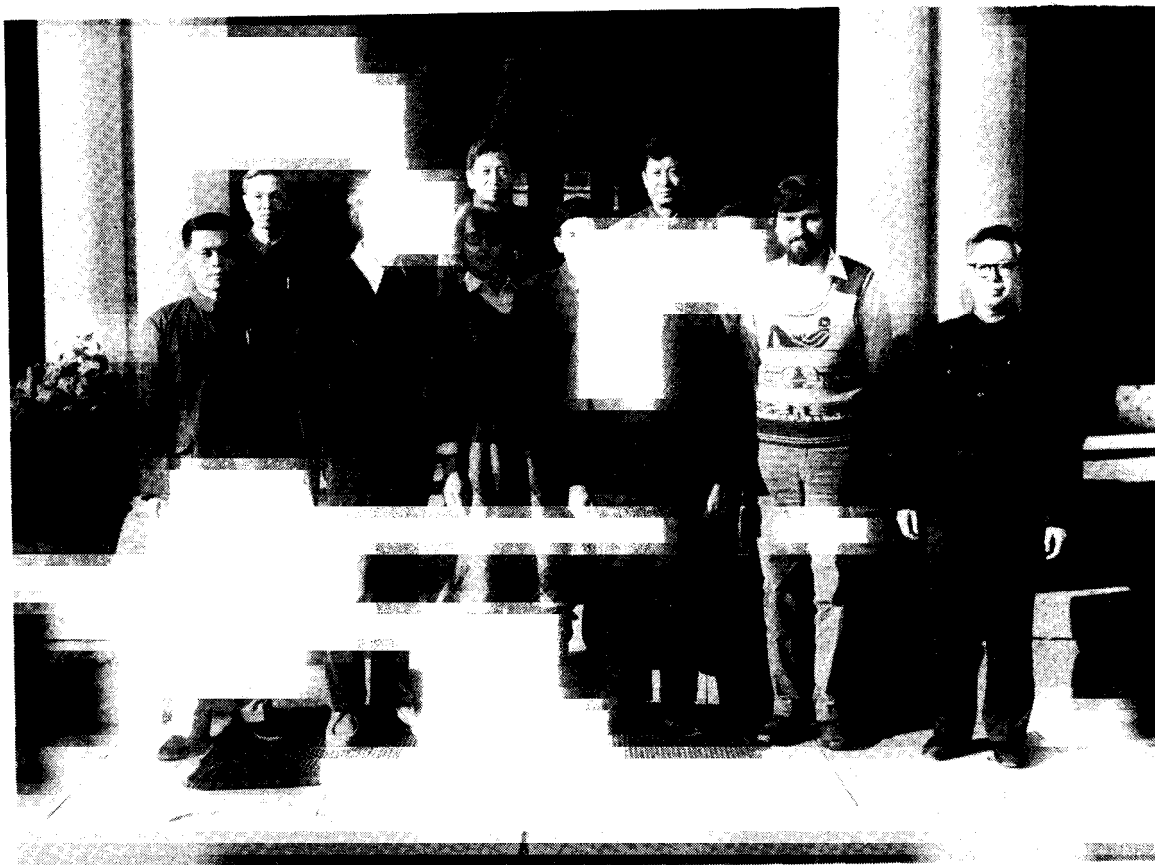
Modern acoustic research is relatively young in China, but acoustics played a role in the design of historical monuments some thousands of years ago. The echo wall in the Temple of Heaven, Beijing, is a good example of this. Another manifestation of the Chinese interest for acoustics is the brass wash "fish basin" of the Hang dynasty, that makes use of vibrations to throw water up in the air.

The research group of Acoustics, Beijing was started in 1957 and became the Institute of Acoustics in 1964. It is presently headed by Prof Wang De Zhao and has around 400 employees.

中国科学院声学研究所

INSTITUTE OF ACOUSTICS

ACADEMIA SINICA



The Institute has a central position in acoustic research in China and its main publication is *Acta Acustica* edited by Prof Maa Dah You, presently deputy director at the Institute. He has been at the Institute since it started and has worked mainly with room acoustics. He is also a former student of Leo Beranek. *Acta Acustica* is the official Journal of the Chinese Acoustical Society, and papers from researchers outside the Institute are accepted.

around 30 persons and is headed by Dr Zhang Jie. The work

The Acoustical Society of China was founded in 1964 but did not have a second meeting until May 1979. The intention now is to continue with a meeting every third year. The revival of old societies and the creation of new ones is typical for the present situation in China.

The Acoustic Institute in Beijing covers a variety of topics in acoustic research in the frequency range from 10^{-4} to 10^{+12} Hz. Even though our main interest was in speech communication we spent some time on interesting work in other areas like ultrasonics, piezoelectric hardware and SAW (Surface Acoustic Waves) devices for signal analysis. One project with immediate application concerned the localization of typhoons with the help of an infrasonic technique. Methods for visualizing sound fields with holographic methods were also demonstrated. The work was supported by a workshop with original methods for growing crystals.

The Underwater Acoustics Laboratory showed experiments to develop underwater transducers and also a large watertank used to evaluate different kinds of transducers. The main interest of the group was to study sound transmission under shallow water conditions (typically the Chinese Sea). A Chinese computer DJS-6, with 16k word memory, was used for signal processing. The Institute also had developed a special hardware FFT spectrum analyzer. The computer facilities were going to be expanded with a PDP 11/70 computer within a year.

In audio acoustics the Institute had good testing facilities such as an anechoic chamber (400 m^3) and a large reverberation room (450 m^3) with a maximum of 25 sec reverberation at 500 Hz.

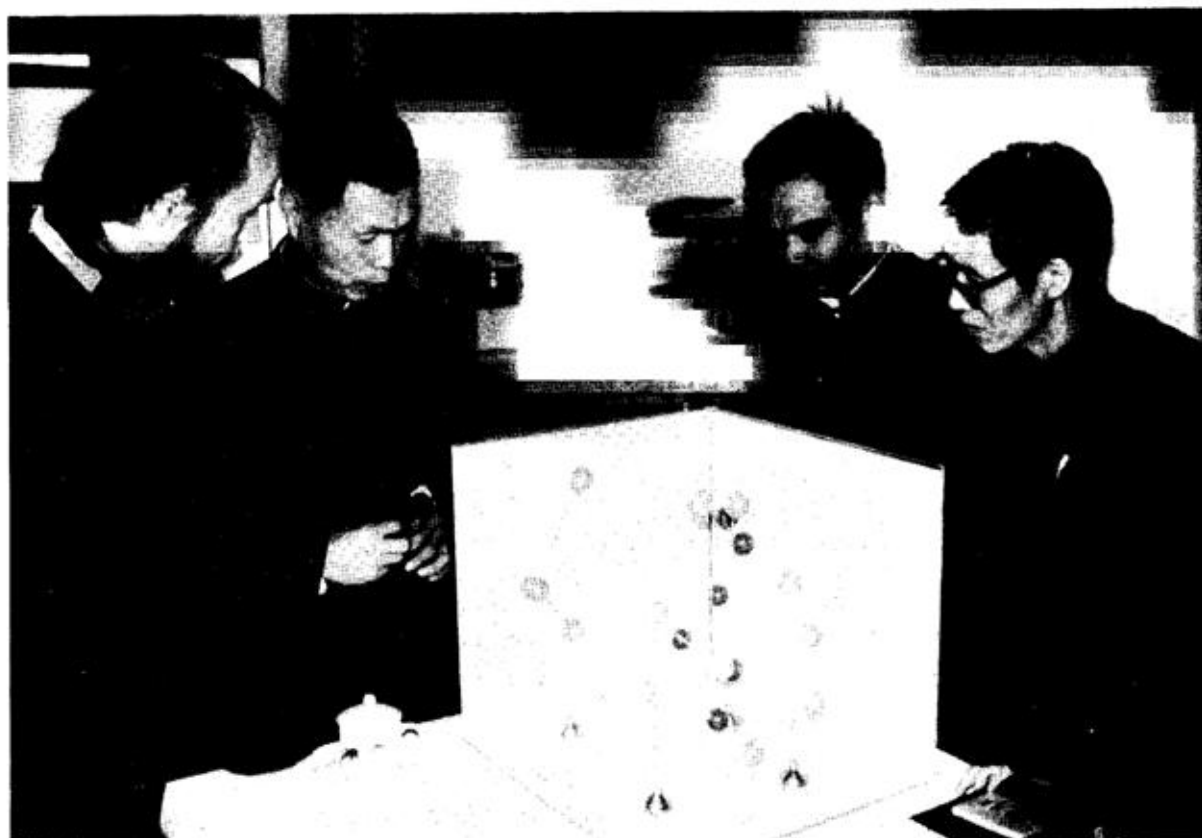
Some work was also carried out in musical acoustics with analysis of ancient instruments such as 2000-year old chime bells. This work was demonstrated to us by Prof Chen Tong.

Speech Acoustics and Communication.

Most of our time at the Institute we spent with the group working in speech acoustics and communication. This group has around 30 persons and is headed by Dr Zhang Jia-Lu. The work

covers different aspects of speech communication such as phonetic analysis of speech and the development of speech transmission systems (vocoders). Much effort has been put into a statistical and physical description of the Chinese standard dialect (Putong-hua). Speech recognition has had a high priority during the last ten years.

An analysis of one million words from running text has been undertaken by Prof Xii. The work is comparable to the collection of the Brown corpus for English and the Allén corpus for Swedish. The result strongly spoke against the misconception of Chinese as a language of basically monosyllabic words. The majority of the words in the corpus (69%) consisted of two syllables.



Multidimensional model of Chinese consonants.

Analysis of Chinese by Prof Zhang and colleagues in terms of articulation index called for some modifications of the frequency bands with equal contribution to intelligibility. Mid-frequency bands were relatively more important in Chinese compared to

English. Confusions between consonants also showed some language specific peculiarities. The phoneme /l/ was frequently confused not only with /n/ but also with /m/. It was mentioned that the /l/ and /m/ distinction was actually neutralized in some dialects. Chinese does not have consonant clusters within syllables and phonetic dependencies between consonants and vowels are well developed. These facts could partially explain the interesting finding that Chinese syllables have considerably higher intelligibility compared to European languages under similar transmission conditions. In the total communication situation this advantage for Chinese is counteracted by the many words with the same pronunciation. This results from the small number of possible Chinese syllables (about 1200 including the distinctive tone). The four tones (five including neutral tone) were almost never confused in an intelligibility test.

Speech transmission techniques were studied especially aiming for an improved vocoder. Microcomputers of several kinds were used both in simulating systems and in the final design. It was clear that there is a big effort in China to start production of chips and even microcomputers. In the designs, we found that about half of the chips were domestically made.

Since Chinese is a tone language, it is essential for the understanding of vocoder speech to get a good fundamental frequency analysis. The research was very much concentrated on this subject. A method with recursive digital filters and associative decision logic to trace the frequency contour was demonstrated in a digital 2400 bit/sec channel vocoder. The fundamental frequency tracking was remarkably good. The over-all quality of the system appeared somewhat muffled. Perhaps this was due to the selections of channels which were chosen to conform to the bands of equal contribution found in the previously mentioned experiments. The comparatively wide low frequency bands may be negative for the overall quality. Intelligibility tests showed a result of around 65 % recognition of consonants in nonsense syllables for both male and female speakers. Research was also conducted on the excitation function used in the synthesis circuits of the vocoder. In an experimental system the pulse shape could be changed depending on fundamental frequency and amplitude.

Speech recognition has been a project at the Institute for almost ten years and is headed by Prof Xü Huan-zhang. Although the machinery used seems to be very simple in the eyes of the Westener, an excellent system for recognizing isolated words was demonstrated by Dr Yu Tie-chen. The system was speaker adaptive. A Brüel and Kjaer 1/3 octave band frequency analyser was used after some modifications to shorten the integration time. It was connected to a Varian minicomputer with 8k words (16 bit) 1.8 usec memory cycle. Each word (utterance) was stored as 15 binary spectra. The sampling points were selected so that the cumulative spectral derivatives between samples were approximately equal. The comparison between the input and the reference was done with help of an "exclusive-or-method". The biggest vocabulary that has been tried was 400 words, which were recognized in about real time with 97.7 % accuracy.



Yu Tie-chen at the speech recognition system.

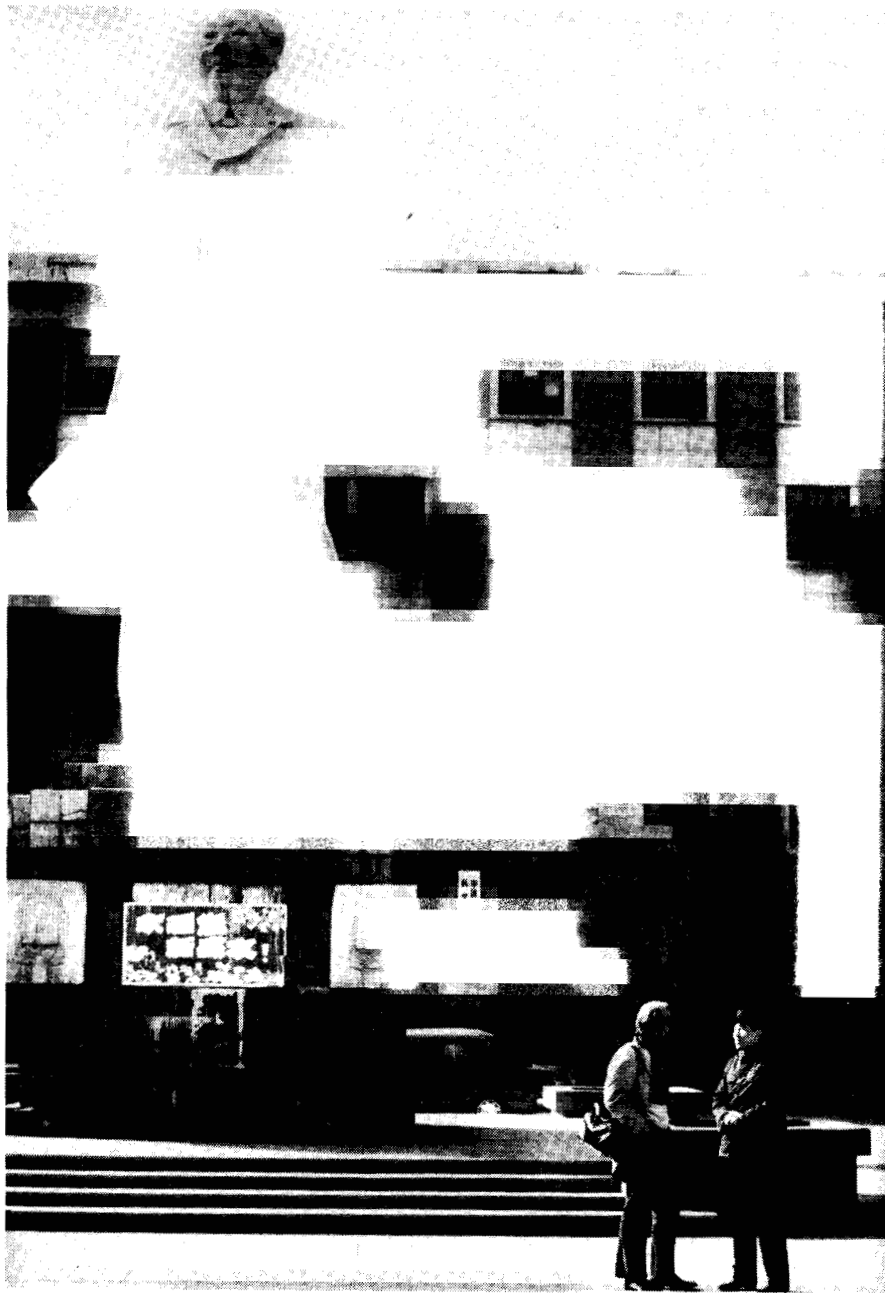
Yu Tie-chen had just been to an international computer conference in Hong-Kong. The system received appropriate appreciation and he received several offers to make a commercial implementation of the speech recognition system. He could also, with a big smile, report on a speech recognition system from the province of Taiwan. They had used LPC analysis as input and obtained around 70 % correct on a 20-word vocabulary.

One application that is under development is a parcel sorting system for the post office in Hangzhou. The vocabulary consisted of the names of the provinces and around 100 city names. The system is planned to be ready in a couple of years.

Institute of Linguistics, Academy of Social Sciences, Beijing.

The Institute of Linguistics consists of many different research groups, several with quite technically-oriented projects. We were met by Prof Wu Zong-ji and Dr Lin Mao-tsan. Prof Wu Zong-ji has been engaged in linguistic research during his whole life and is still very active in the development of standard Chinese (Putonghua). Part of this work is devoted to an acoustic analysis of the Chinese syllable and has resulted in a monograph that will be in press in 1981. At the Ninth International Congress of Phonetic Sciences in Copenhagen 1979 Prof Wu presented a feature system describing the Chinese phoneme structure. Prof Lin Mao-tsan is studying Chinese tone sandi rules for phrases of varying length (i.e. context dependent tone changes). The knowledge of these rules is of great importance for the development of e.g. a text-to-speech system. Results of their linguistic research are presented in one of the journals from the Institute, Zhongguo Yuwen.

The Institute is the center of linguistic research and has the main responsibility for defining standard Chinese (Putonghua). Some spelling reforms have been carried out and resulted in simplifications of some Chinese characters. Pinjin, the official way of transcribing Chinese with the Latin alphabet, has not enjoyed any large-scale acceptance. There does not seem to be any political intention to promote Pinjin at the expense of Chinese character writing. In primary school, Pinjin is, however, starting to get used in parallel with the traditional writing. In books for the blind, Braille coded Pinjin is used, but the problem to specify the different tones has not been solved. The necessary loss of information when a Chinese character is transformed to Pinjin spelling is a large problem. The Institute of

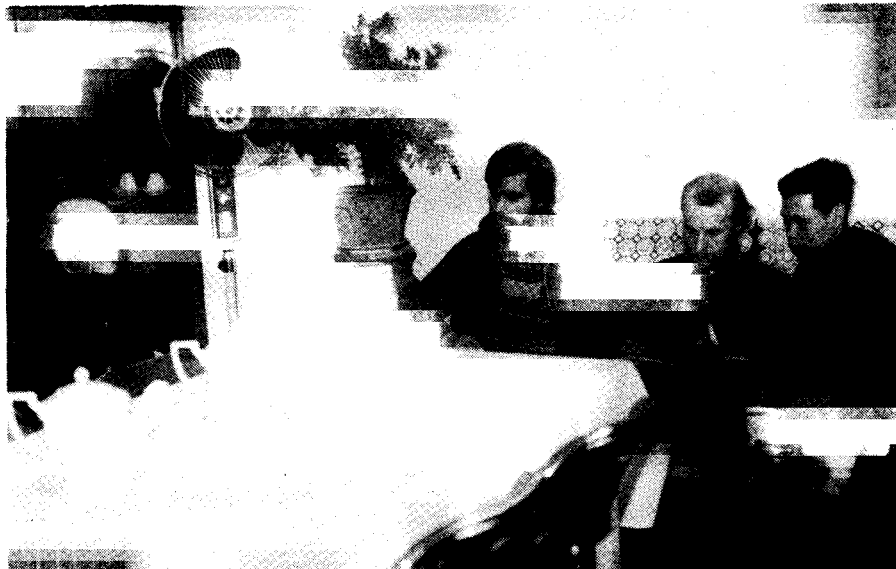


Outside the Institute of Linguistics, Beijing.

Linguistics has recently published a dictionary of the basic Chinese words (800) organized according to the Pinjin spelling.

Working with Chinese characters in information systems call for new techniques, and the Institute is heavily engaged in this. Methods for input/output and storing of Chinese text are still far from standardized. There is even a newly-founded society

devoted to these problems, "The Chinese Society of Chinese Information Processing". Optical character recognition (OCR) techniques play an important role in this work to find methods to input characters into a computer.



Discussions with Wu Zong-ji and Zhang Jia-lu.

Two parallel projects on machine translation were demonstrated to us by Prof Liu Yang-quan and Prof Liu Zhou from the machine translation research section. The source language used to be Russian but, since 1960, has been English. Both projects produce Chinese character output. The current ambition is to handle titles of scientific papers.

We also heard about a language understanding project that has recently started. The intention is to use an ATN grammar (augmented transition network grammar) and case theory as the framework for sentence analysis. Free word order is common in Chinese and has to be taken into account even in a first approach to sentence analysis. (10)

On another occasion, we met Li Jia-zhi from the Institute of Psychology, Academia Sinica. He is working on the same problems and has also a total question-answering system running. It is operating on a small data base containing information about animals.

The Institute of Acoustics, University of Nanjing.

The Institute of Acoustics, University of Nanjing was founded in 1955 and is headed by Prof Wei Rong-jue. He is also the chairman of the Department of Physics. It is unusual for institutes to be within universities in China, but this was an exception. During the cultural revolution it had been difficult to continue work, and the Institute was more or less closed during that time. Now when the interaction between education and research is encouraged, the situation is quite favourable. The Institute is, as the Institute of Acoustics in Beijing, engaged in a variety of projects in audio acoustics, surface acoustics, ultra acoustics and molecular acoustics. The Institute had good measuring facilities with different test chambers and recording studios, e.g., the best anechoic chamber in China (1000 m³).

Many of the audio acoustic projects were focused on noise. Several noise suppression schemas had been simulated on computers. In room acoustics, noise absorbers were tested and developed, including active absorbers for low frequency noise reduction. Holographic techniques were used to test and correct loudspeakers.

In the more specifically speech-related areas, studies had been carried out dealing with statistical analysis of Chinese and intelligibility tests. Also, under reverberant conditions Chinese syllables appeared more intelligible than what is reported for English.

One project in progress is to classify and describe different voices, both subjectively and objectively. Both professional radio announcers and singers had been under study. The instrumental techniques used were long term average spectra, fundamental frequency analysis and standard spectrographic analysis.

Institute of Physiology, Academia Sinica, Shanghai.

The Institute of Physiology, Shanghai was split off from an institute dealing with bio-chemistry in 1958. It has today about 350 persons employed. A new split is planned for 1981 when a separate institute for brain research will be created. Shanghai appeared to be the centre for physiologically-oriented research in China.

Our visit was concerned only with the group working on the auditory functions headed by Prof Liang Zhi-an. This group was very interesting in that it combined basic physiological studies with development of clinical methods and the group was also engaged in psychoacoustical research.

The analyzing properties of the auditory system had been investigated with psychophysical methods. Measures such as difference limen could be duplicated by measuring evoked responses from the cortex of animals.



Liang Zhi-an at a chinese-built computer for running electro-physiological experiments.

Acknowledgement

In the fall 1979 the Department of Speech Communication (KTH) was visited by three scientists from The People's Republic of China, Prof Wu Zong-ji, Prof Zhang Jia-lu and Dr Lin Mao-tsan. At that time we had both inspiring and valuable discussions about phonetic, linguistic and telecommunication matters. It was felt that this could be a start of future cooperation between China and Sweden. Our trip could be seen as a manifestation of this idea and we hope that the interchange of ideas and personnel should increase.

We would like to express our sincere thanks to Prof Zhang Jia-lu who spent much of his valuable time together with us during the whole visit. Valuable discussions at his laboratory and his great knowledge of scientific work in China is very much appreciated. We would also like to thank all his research colleagues we met in China, for all the effort they spent in making our stay so rewarding.

We would like to express our sincere thanks to the Academia Sinica and the Swedish Academy of Engineering Science (IVA) for making this trip possible. Special thanks go to Mr Zheng Er-li who organized an excellent program including not only scientific matters but also visits of general nature to introduce us to the Chinese Culture.

In this connection we also want to thank Göran Malmqvist, professor of the Chinese language at the University of Stockholm. Without his kind and patient guidance we would never have managed to construct the Pinjin-to-Putonghua Speech Synthesis System. Our thanks are also due to Ola Svensson, technical attaché at the Swedish embassy in Beijing. He shared with us his extensive experience with Chinese society, organization of research in particular, and contributed much to make our stay in Beijing very pleasant and interesting.