FURUI Sadaoki, "Establishment of Automatic Speech Recognition".

1. Biography

1945-9 Born in Tokyo. B.E. Department of Mathematical Engineering, University of Tokyo, 1968-3. M.E. Graduate School of Engineering, University of Tokyo, 1970-3. Ph.D., Engineering, 1983-12. In 1986, Director, NTT Basic Research Laboratories Research Laboratory IV. In 1989, Director, Speech Information Research Department, Human Interface Research Laboratories, NTT Corporation. In 1991 Director, Furui Research Laboratory, NTT Research Laboratories. From 1994, Visiting Professor, Department of Computational Engineering, Graduate School of Information Science and Technology, Tokyo Institute of Technology. From 1997, Professor, Tokyo Institute of Technology. From 2007, Dean, Graduate School of Information Science and Technology, from 2009, Director, University Library, from 2011, Professor Emeritus, from 2017, Professor Emeritus, all in Tokyo Institute of Technology. In 2013, Provost, Toyota Technological Institute of Chicago, and in 2019, its chairman. From 2019, Director-General of the National Institute of Informatic. Died 2022-7 (age 76). Director-General of the Science and Technology Agency Award, Commendation from the Minister of Education, Culture, Sports, Science and Technology, IEEE James L. Flanagan Speech and Audio Processing Award, NHK Broadcasting Culture Award, Okawa Award, Medal with Purple Ribbon, Persons of Cultural Merit.

2. Summary

Dr. Sadahiro Furui has long been engaged in research and education on speech information processing, especially automatic recognition of spoken speech by computer, and has proposed basic technologies for automatic speech recognition and understanding, contributing to the advancement of these technologies and training many researchers. These researchers are now widely active as researchers and engineers of speech technology in Japan and abroad.

3. Summary of research achievements

3.1 Extraction of Dynamic Features of Speech

Focusing on the fact that human hearing is sensitive to the dynamic features of sound, he proposed a speech recognition method that obtains regression coefficients (dynamic cepstrum) from the cepstrum time series of speech sounds, and combines the instantaneous and dynamic cepstrums as features of speech sounds. Conventional speech recognition methods use only the instantaneous features of speech, which lacks important auditory information and is prone to confusion with different speech sounds. The method he proposed reduced the number of speech recognition errors by almost half. This method was used in most of the speech recognition systems (products) in the world at that time.

3.2 The Beginning of Spoken Language Engineering

As the project leader of the "Corpus of Spontaneous Japanese" project funded by the Special Coordination Funds for Promoting Science and Technology from FY1999 to FY2003, he promoted the research and constructed the world's largest and most accurate spoken language database (corpus) consisting of 7 million words. He has constructed the world's largest and most accurate database (corpus) of spoken language consisting of 7 million words. By developing phoneme models, language models, and speech recognition methods based on this database, he contributed to a significant improvement in the recognition accuracy of spoken speech, which had been extremely

difficult to achieve. Compared to the previous method, which used a phoneme model based on speech read out loud of written sentences and a language model based on written words, the recognition errors for speech such as lecture speech were reduced by almost half. For example, this technology is useful as a basic technology for automatic transcription of speech for NHK subtitling broadcasts.

3.3 Development of Speaker Adaptation Technology

In addition to focusing on the fact that voice variation due to speaker differences and background noise is one of the important causes of poor speech recognition performance, he experimentally confirmed that human hearing is extremely adaptable, and have long and vigorously pursued research to realize this capability on a computer. As part of this research, he has advanced research on technologies for adapting phoneme models and proposed various creative methods based on hierarchical modeling of phoneme models, which have greatly contributed to the advancement of technology.

3.4 Promotion of Fusion of Literature and Science

As the leader of the 21st Century COE Program "Systematization of Large-Scale Knowledge Resources and Construction of Utilization Infrastructure" from 2003 to 2007, he contributed to the promotion of fusion research between the humanities and sciences to construct technologies that enable electronic large-scale knowledge resources (content) to be interconnected and easily used.

3.5 Development of Speech Recognition Technology for Real Environment

From FY2006 to FY2008, as a leader of the "Development of Technology for Utilization of Information Appliance Sensors and Interface Devices (Development of Fundamental Speech Recognition Technology)" project of the Ministry of Economy, Trade and Industry, in collaboration with a major Japanese electronics manufacturer, he developed highly accurate speech recognition technology that can withstand real-world use and easy-to-use speech.

4. Special note

The textbook "Digital Speech Processing," written in 1985, has been translated into English and read around the world as a standard textbook on speech engineering. He is a board member, vice president, and president of the Acoustical Society of Japan, a board member of the Institute of Electronics, Information and Communication Engineers, president of the International Speech Communication Association (ISCA), a board member of the IEEE Signal Processing Society, a member of the Journal of Speech Processing Society. He has made significant contributions to the development of the academic community by serving in a number of important positions, including Director of the IEEE Signal Processing Society, Editor-in-Chief of the Journal of Speech Communication, and President of the Asia Pacific Signal Processing Association (APSIPA).

5. References

[1] Sadaoki Furui, "Digital Speech Processing," Tokai University Press, 1985-9.
[2] Sadaoki Furui, "New Acoustic and Speech Engineering," Kindai Kagaku-sha, 2006-9.
[3] Sadaoki Furui, "University and Society in the Age of AI," Maruzen Planet, 2021-5.
[4]"Oral-History: Sadaoki Furui", https://ethw.org/Oral-History:Sadaoki_Furui, Engineering and Technology History Wiki, 2021-12.

6. Representative Papers

- Sadaoki Furui, "Cepstral Analysis Technique for Automatic Speaker Verification," IEEE Transactions, Vol. ASSP-29, No. 2, pp. 254-272, 1981
- Sadaoki Furui, "Speaker Independent Isolated Word Recognition Using Dynamic Features of Speech Spectrum", IEEE Transactions on Acoustic, Speech and Signal Processing, Vol. ASSP-34, No. 1, pp. 52-59, 1986

(Presented by Koichi Shinoda on 26 December, 2022)