D-1～D-2

Speech Recognition and its Realization

Speech Recognition

Practical Realization of Speech Recognition

Because telephones were important means of communication, studies on speech to send/receive via telephone started early. The study of speech recognition by computer started from the study by T. Sakai on speech typewriters using cello cross waves. Then, along with development of computers, research gathering worldwide attention were carried out—a continuous speech recognition method using DP matching by H. Sakoe, et al., a speech recognition method by S. Furui, using cepstrum and deltacepstrum, and a speech recognition method by S. Nakagawa using the hidden Markov model. In recent years, studies on the sampling of speech characters and recognition through deep learning have been carried out. The practical realization of technology then advanced. Y. Kato, et al., developed speaker independent speech recognition equipment in addition to continuous speech recognition equipment. T. Watanabe worked on the study of the practical realization of integration, compactification, and error correction of then technologies—independent speakers, large vocabulary, continuous speech recognition, and in recent years, developments toward mobile terminals using cloud services advanced.

DP: Dynamic Programming

D-3

Speech Synthesis

In Japan, studies on speech synthesis started around 1970 from the speech analysis-synthesis system based on PARCOR by F. Itakura and S. Saito. F. Itakura and N. Sugamura then proposed a speech synthesis system based on LSP characteristics which was better than PARCOR. In early studies, first input speech was analyzed, then expressed as special drops, and then through speech synthesis was dropped back to the original speech. After that, studies on speech synthesis enabled the synthesis of any sentence, and the vocal-tract model and analysis-synthesis system based on the statistical model of speech signals was proposed and advanced toward more efficient processing and higher-quality synthesized speech. High quality speech synthesis was created by storing a large amount of speech element data cut from the natural voice, and optimal elements were selected from the stock, and the signal processed. M. Akamine, et al., proposed a high-quality speech synthesis system with small data memory size through statistical learning. K. Tokuda, et al.,

proposed a speech synthesis system based on the hidden Markov model. In recent years, studies on the creation of speech synthesis through deep learning have been carried out. Studies on the synthesis of singing voices and emotional speech and studies on changing the quality of voices have also been carried out.

PARCOR: PARtial auto-CORrelation

LSP: Line Spectrum Pair