IMAI, Satoshi, "Establishment of Speech Processing Technology using Cepstrum"

1. Biography

Satoshi IMAI was born on January 12, 1937. He graduated from Tokyo Institute of Technology in 1959, completed the doctoral course at the graduate school of Tokyo Institute of Technology and received his Doctor of Engineering degree in 1964. He joined the Research Laboratory of Precision Machinery and Electronics, predecessor of the Precision and Intelligence Laboratory, at Tokyo Institute of Technology as Research Associate in 1964. He became Associate Professor and Professor at the same laboratory in 1968 and 1979, respectively. He moved to Chiba Institute of Technology as Professor in 1997 and retired in 2007. He passed away on December 14, 2014.

IMAI, Satoshi

2. Summary

In the research field of information processing of speech signals, IMAI focused on the importance of utilizing a set of parameters called "cepstrum" as representing the features of the smallest linguistic unit of speech, i.e., phoneme, and consistently conducted research of speech processing technology using cepstrum. He contributed to the development of the research by proposing novel speech processing approaches based on cepstrum, such as speech analysis techniques for extracting reliable cepstrum coefficients from speech waveforms, speech synthesis techniques for generating speech waveforms from input texts, speech recognition techniques for converting input speech waveforms to texts, and speech coding techniques for efficiently transmitting or storing speech signals. For a series of his research achievements, he received the Japan Society for Precision Engineering (JSPS) Best Paper Award in 1965 and the Acoustical Society of Japan (ASJ) Distinguished Achievement Award in 2005, and was elevated to the Institute of Electronics, Information and Communication Engineers (IEICE) Fellow in 2002.

3. Summary of Research Achievements

3.1 Speech Analysis Techniques Based on Cepstral Representation

Cepstrum, that is defined as the inverse Fourier transform of the logarithm of the spectrum, has been applied to speech processing, such as speech analysis, speech synthesis, speech recognition, and speech coding, since the early stages of speech processing research. However, the cepstral coefficients obtained by conventional analysis techniques result in biased estimates against the given log spectrum, and thus they do not always give good performance in speech processing applications. In this context, IMAI established a novel technique for obtaining unbiased estimator of log spectrum [1], in which the resultant cepstral coefficients provide higher performance than those obtained by the conventional techniques when they are used in practical speech processing applications. The proposed technique later developed into a unified analysis technique of cepstral analysis and linear predictive analysis.

3.2 Research on Natural-Sounding Speech Synthesis Systems

When synthesizing speech from cepstral coefficients, we require a speech waveform generation process that inherently accompanies with nonlinear transformation. As a result, it tended to be complicated one and the synthetic speech quality was not always high. To overcome this problem, IMAI invented a novel realization structure of speech synthesis filters, called the “log magnitude approximation (LMA)” or “mel-log spectrum approximation (MLSA)” filter, in which the filter coefficients are identical with the given cepstral coefficients [2]–[5]. He developed a rule-based Japanese text-to-speech system using cepstral parameters and MLSA filter, which was the pioneering work of the realistic text-to-speech system in Japan. In the early 2000s, the statistical parametric speech synthesis had become the mainstream approach to text-to-speech with various speakers’ voices and styles, many institutions and companies worldwide utilized MLSA filters for the waveform generation part of text-to-speech and text-to-song systems.

3.3 Research on Speech Processing Considering Auditory Characteristics

IMAI proposed to incorporate human auditory characteristics into the process obtaining cepstral coefficients, called “mel cepstrum,” that is, cepstrum obtained from log spectrum on a warped frequency scale approximating the “mel” frequency scale. Then he established speech processing techniques using mel cepstrum [5][6]. His pioneering wok revealed that the performance of speech synthesis systems, automatic speech recognition systems, and speech coding systems becomes higher by using mel cepstrum. A series of these research results have been found in books [7][8].

4. References

[1] Satoshi Imai, Chieko Furuichi, “Unbiased estimator of log spectrum and its application to speech signal processing, Proc. EUSIPICO-88, pp. 203–206, Grenoble, France (1988.9)

[2] Satoshi Imai, Tadashi Kitamura, and Hiroyuki Takeya, “A direct approximation technique of log magnitude response for digital filters,” IEEE Trans. Acoustics, Speech, and Signal Processing, Vol. ASSP-25, No. 2, pp. 127–133 (1977.4)

[3] Satoshi Imai, “Digital filter design in the quefrency domain,” IEEE Trans. Acoustics, Speech, and Signal Processing, Vol. ASSP-26 No. 3, pp.2 26–235 (1978.6)

[4] Satoshi Imai, “Log magnitude approximation (LMA) filter,” IEICE Trans. Fundamentals (Japanese Edition), Vol. J63-A, No. 12, pp. 886–893 (1980.12)

[5] Satoshi Imai, Kazuo Sumita, and Chieko Furuichi, “Mel log spectrum approximation（MLSA）filter for speech synthesis,” IEICE Trans. Fundamentals (Japanese Edition), Vol. J66-A, No. 2, pp. 122–129 (1983.2)

[6] Satoshi Imai, “Cepstral analysis synthesis on the mel frequency scale,” Proc. IEEE ICASSP ’83, pp. 93–96, Boston, USA (1983.4)

[7] Satoshi Imai, Speech Recognition, Kyoritsu Shuppan (1995) (in Japanese)

[8] Satoshi Imai, Speech Signal Processing, Morikita Publishing (1996) (in Japanese)

（Presented by KOBAYASHI, Takao on 11 09 2023）