International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 6 Issue 2 ǁ 2018 ǁ PP. 23-25

Feature Parameter Extraction Techniques for Distributed Speech Recognition

*Seung Ho Choi, **Yong Ho Suk
**Department of Electronic and IT Media Engineering, Seoul National University of Science and Technology,Seoul, Korea*
Corresponding Author:Seung Ho Choi

ABSTRACT :*In this study, we concerns feature extraction techniques for bitstream-based speech recognition and distributed speech recognition for efficient speech recognition services under wireless communication environments. In addition, we introduce the MFCC-based speech coding method for speech recognition in IP network server.*

KEYWORDS –*Speech recognition, feature extraction, bitstream-based speech recognition, distributed speech recognition, MFCC*

Date of Submission: 09-02-2018 Date of acceptance: 26-02-2018

I. INTRODUCTION

Current wireless communication systems use a low-rate encoder that compresses 8 to 16 times more than the compression rate in a wired telephone network. This compression technique not only degrades the sound quality but also the speech recognition feature parameters extracted based on the restored speech signal Furthermore, the voice data transmitted from the wireless communication network is accompanied by an error. As a result, not only the sound quality degradation but also the loss of characteristic parameters important for speech recognition result in the performance degradation of speech recognition. In addition, voice calls can go through different wireless communication methods through the roaming service. Due to the different voice encoding schemes, the voice finally reaching the speech recognition system can be greatly distorted. Therefore, much research has been carried out to overcome these constraints. In addition, some research works have been carried out for bitstream-based feature extraction techniques for efficient speech recognition and distributed speech recognition.

II. FEATURE EXTRACTION TECHNIQUES FOR BITSTREAM-BASED SPEECH RECOGNITION

In this section, we introduce a bitstream-based method [1] that performs recognition on the server side based on the parameter (bitstream) of the speech encoder transmitted from the communication network. In the bit stream based method, the feature parameters are directly extracted from the parameters of the speech encoder adopted by the system, that is, from the bitstream. The parameters of the speech coder are mainly the line spectrum pairs (LSP) or the log area ratio (LAR), which contain the envelope information of the spectrum, and the energy information and the pitch information. Recognition experiment results based on the bitstream of LPC-10e or QCELP are presented in [2] and [3], respectively. In addition, a pseudo-cepstrum scheme has been proposed to efficiently convert from LSP to cepstrum [4]. Figure 1 shows a typical procedure for extracting recognition parameters based on parameters of a speech coder.

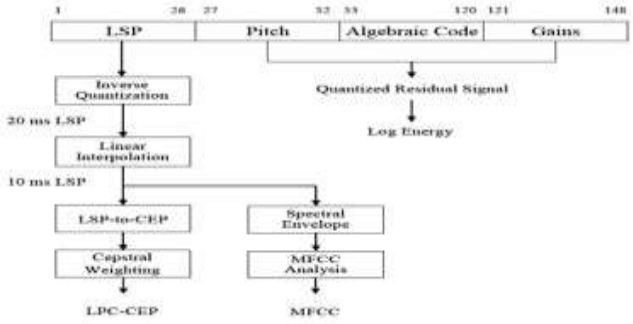

Fig. 1 Procedure for bitstream-based feature extraction using speech coder parameters [1]

III. FEATURE EXTRACTION TECHNIQUES FOR DISTRIBUTED SPEECH RECOGNITION

The Distributed Speech Recognition (DSR) method, which is carried out dispersedly on both the terminal and the server, divides the recognizer into two parts and links them to the transmission channel. Current speech recognition technologies require excessive amounts of memory and computation to implement functions such as large capacity and continuous speech recognition in terminals. In addition, a system recognized on the server side of a mobile communication network that is currently in commercial use causes a serious decrease in the recognition rate due to low-rate speech encoding and channel transmission errors. In order to overcome the limitations of the terminal and voice data loss in the channel transmission simultaneously, the DSR system only extracts the recognition parameters from the memory and the client side with limited amount of computation, And transmits through the channel of the network. On the server side, it performs recognition based on complex statistical algorithms or sufficient acoustic models and language models. As the feature parameter, the most widely used MFCC-based feature vector can be used.

[5] presented a study on quantization of cepstrum for speech recognition on the World Wide Web [6] proposed a method of extracting the 13th order MFCC vector and log energy every 10ms and transmitting the characteristic parameters at a bit rate of 4.8 kbit/s which is half of the transmission bit rate of GSM using the VQ method. [7] proposed a method of quantizing perceptual linear prediction (PLP) parameters at 400 bit/s. However, the loss of data due to errors in the digital transmission channel can be a major factor in the performance degradation of the DSR scheme. Several techniques have been proposed to compensate for these channel errors [8, 9].

IV. FEATURE EXTRACTION TECHNIQUES FOR DISTRIBUTED SPEECH RECOGNITION

In this section, we describe CELP type speech coding technology suitable for speech recognition in IP network server. Conventional speech codecs use LPC coefficients to represent spectral envelope information, which is quantized for transmission. However, if the MFCC coefficient, which is most widely used in the speech recognition system, is used as the recognition parameter instead of the LPC coefficient, recognition performance can be improved when speech recognition is performed in the server of the IP network.

Figure 2 shows the encoding process of the MFCC-based speech coder. As shown in the figure, the coding process of the existing CELP type speech coder is modified. First, the MFCC is extracted from the speech signal, quantized, and transmitted. And converting the quantized MFCC to an LPC coefficient as in the speech decoder. In order to recover the speech signal, the speech decoder needs to convert the transmitted MFCC coefficients into LPC coefficients as shown in Fig. First, the MFCC coefficients are transformed into frequency domain samples by an inverse discrete cosine transform (IDCT) and an inverse logarithm process. Then, a power spectrum is obtained from a spectrum subjected to a spectral interpolation process, and an autocorrelation coefficient is calculated by performing inverse fast Fourier transform (IFFT) on the power spectrum. Then, a lag window is applied to the autocorrelation coefficient, and the LPC coefficient is finally obtained using the Levinson-Durbin algorithm.

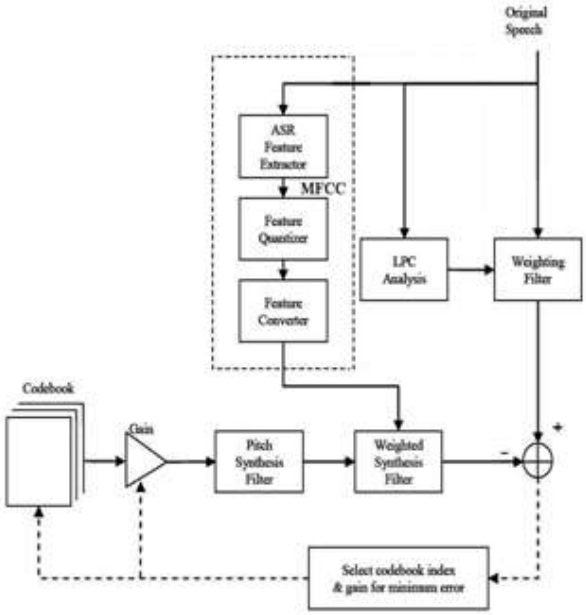

Fig. 2Procedure for MFCC-based speech coding [1]

V. CONCLUSION

In this study, we have studied the research trend of feature extraction technology for bitstream based speech recognition and distributed speech recognition, and introduced the MFCC based speech coding method for speech recognition in IP network server. Each recognition method in the wireless communication and network environment has advantages and disadvantages in its own way, and it is necessary to make an appropriate selection according to the complexity in the implementation of the speech recognition system, the resources of the terminal, and existence of the transmission protocol between the terminal and the server.

Acknowledgements

This study was supported by the Research Program funded by the SeoulTech(Seoul National University of Science and Technology).

REFERENCES

[1]. [1] H. K. Kim and R. V. Cox, ˝A bitstream-based front-end for wireless speech recognition on IS-136 communications system," IEEE Trans. Speech Audio Process., vol. 9, no. 5, pp. 558-568, July 2001.
[2]. [2] A. T. Yu and H. C. Wang, ˝A study on the recognition of low bit-rate encoded speech," in Proc. ICSLP, Sydney, Australia, pp. 1523-1526, Nov. 1998.
[3]. [3] S. H. Choi, H. K. Kim, H. S. Lee, and R. M. Gray, ˝Speech recognition method using quantized LSP parameters in CELP-type coders," Electronics Letters, vol. 34, no. 2, pp. 156-157, Jan. 1998.
[4]. [4] H. K. Kim, S. H. Choi, and H. S. Lee, ˝On approximation line spectral frequencies to LPC cepstral coefficients," IEEE Trans. Speech Audio Process., vol. 8, no. 2, pp. 195-199, Mar. 2000.
[5]. [5] V. Digalakis, L. Neumeyer, and M. Perakakis, ˝Quantization of cepstral parameters for speech recognition over the World Wide Web," IEEE J. Select. Areas Commun., vol. 17, no. 1, pp. 82-90, Jan. 1999.
[6]. [6] D. Pearce, ˝Enabling new speech driven services for mobile devices: an overview of the proposed ETSI standard for a distributed speech recognition front-end," in Proc. IEE Colloquium on Interactive Spoken Dialogue Systems for Telephony Applications, London, UK, pp. 5/1-5/6, Nov. 1999.
[7]. [7] W. Gunawan and J. M. Hasegawa, ˝PLP coefficients can be quantized at 400 bps," in Proc. ICASSP, Salt Lake City, UT, pp. 77-80, May 2001.
[8]. [8] A. M. Peinado, V. Sanchez, J. C. Segura, and J. L. Perez-Cordoba, ˝MMSE-based channel error mitigation for distributed speech recognition," in Proc. EUROSPEECH, Aalborg, Denmark, pp. 2707-2710, Sept. 2001.
[9]. [9] V. Sanchez, A. M. Peinado, and J. L. Perez-Cordoba, ˝Low complexity channel error mitigation for distributed speech recognition over wireless channels," in Proc. ICC, Anchorage, AK, pp. 3619-3623, May 2003.
[10]. [10] H. K. Kim, "Speech recognition over IP networks," Chapter 3 in Automatic Speech Recognition on Mobile Devices and over Communication Networks, Z.-H. Tan and B. Lindberg (Eds.), Springer-Verlag, London, Mar. 2008.

Seung Ho Choi "Feature Parameter Extraction Techniques for Distributed Speech Recognition"
International Journal of Research in Engineering and Science (IJRES), vol. 06, no. 02, 2018, pp. 23–25.