International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 5 Issue 8 ǁ August. 2017 ǁ PP. 94-96

A Study on the Influence of wireless Communication Environment on Speech Recognition Performance

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

ABSTRACT: In this paper, we study the influence of wireless communication environments on the performance of speech recognition services We examine the issues that affect the performance of speech recognition system in adverse wireless communication environment such as background noise, codec distortion, channel errors, tandem, etc.

Keywords : Speech recognition, wireless communication environment, background noise, codec distortion, channel error

I. INTRODUCTION

The speech recognition method in wireless communication environments can be broadly classified into three types according to where the extraction and recognition of the characteristic parameters are carried out. In other words, the speech recognition system can be classified according to whether the speech recognition is performed in the terminal or in the server connected to the communication network. That is, there are a client-based approach, a server-based approach, and a client-server-based approach that combines the two approaches. The terminal-server based method is also called distributed speech recognition (DSR) [1] [2]. Terminal-based speech recognition has no difference except for the increase of background noise according to the existing speech recognition method and the calculation amount and microphone characteristic, voice acquisition performance, and mobility according to the terminal. In comparison with the server-based and terminal-server-based methods, the terminal-based method is suitable for small-capacity speech recognition such as digit recognition or name recognition. The server-based method divides into a decoded speech-based method that extracts recognition parameters from the reconstructed speech signal as shown in Fig. 1(a), and a bitstream-based method that extracts recognition parameters directly from the parameters of the speech coder as show in Fig. 1(b) [3]. When the reconstructed speech is used as it is, the recognition rate is lowered due to the degradation of the speech quality obtained in the low-rate speech coding [4]. Instead of restoring the speech signal, the bit-stream based method, which is a method for overcoming this problem, directly extracts the feature parameters for speech recognition from the parameters of the speech coder transmitted from the communication network [3] [5].

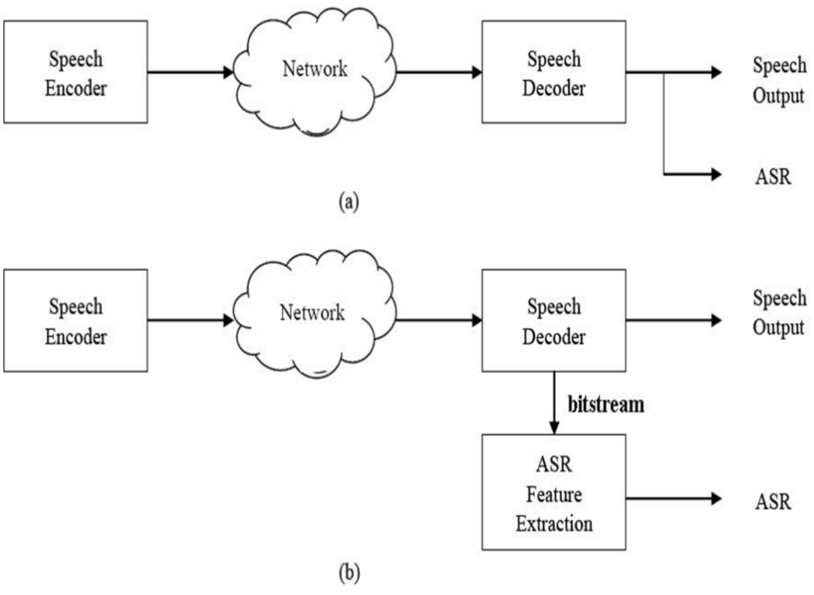

Fig. 1 Server-based speech recognition.(a) Reconstructed speech-based method, (b) Bitstream-based method [3].

On the other hand, the terminal-server type speech recognition system includes recognition parameter extraction and the terminal is built in the terminal separately. That is, it decides whether to perform voice communication or voice recognition at the beginning of the call. According to this determination, it is determined whether to compress the speech signal using the existing speech coder or to extract features for speech recognition. Speech recognition at the server uses transmitted voice feature parameters. Figure 2 shows an example of distributed speech recognition [1].

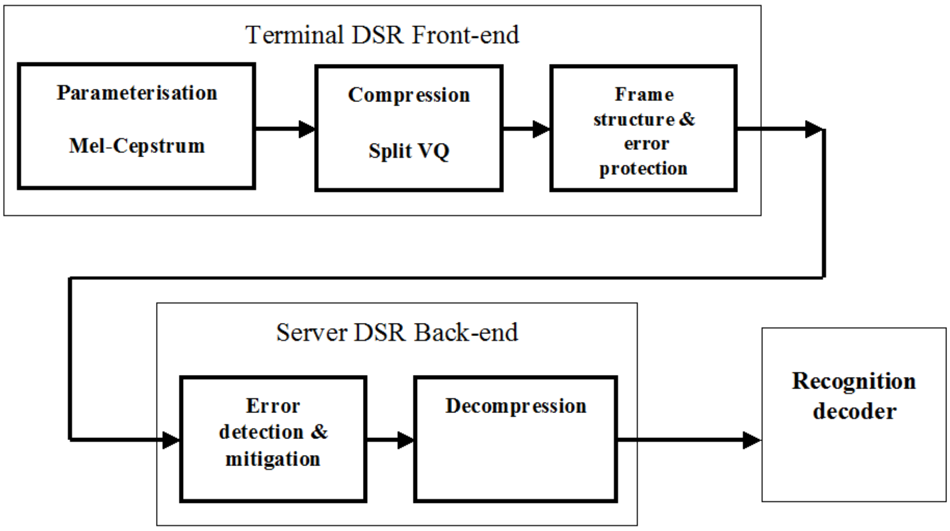

Fig. 2 Example of distributed speech recognition (DSR) [1].

Recently, a study on the MFCC-based speech coding scheme using the speech recognition parameters as the spectrum information necessary for speech coding is underway [6]. As a result, it is possible not only to perform high-performance speech recognition on an IP network server, but also to restore a voice signal.

II. THE INFLUENCE OF WIRELESS COMMUNICATION ENVIRONMENT ON SPEECH RECOGNITION

This section defines the factors affecting speech recognition in wireless communication environment and examines how each affects actual speech recognition.Since the call in the existing telephone network is mainly performed in a house or an office, the noise added during a call is relatively small compared to the size of the voice, and since noise characteristics are mostly predictable. However, under the wireless communication environment, not only the kinds of background noise are very diverse, but also the size and characteristics of the noise are inputted in a state in which the noise cannot be predicted according to the movement of the user. Also, as the terminal is miniaturized, the distance between the microphone of the terminal and the mouth of the user is increased, so that noise can be easily applied even if a directional microphone is used. This is collectively referred to as background noise, and speech recognition under wireless communication is burdened with a function of processing the noise.

Next, distortion due to speech coding exists. In other words, voice coding compresses speech assuming that speech can be modeled as a convolution of an impulse response and an excitation signal containing spectral envelope information. On the other hand, the feature parameters used for speech recognition mainly model the spectral envelope of speech in a relatively detailed manner. Most of the currently used codes The speech compression method of the code-excited linear prediction (CELP) class allocates only 1/4 of the total transmission amount to the spectral envelope used in speech recognition. There is considerable lack of information for discrimination. The bit loss in the channel appears as a loss of the speech transmission frame in the speech decoder. In other words, a frame erasure occurs. In this case, the speech decoder generates the speech of the current frame lost from the immediately preceding transmission frame. Therefore, when the frame erasure occurs in the section where the voiced sound transits from the unvoiced section or the unvoiced section, a serious distortion may be applied to the produced voice.

Finally, when a voice call is realized by roaming between different communication systems (tandem), voice encoding and decoding must be performed in each communication system, which causes distortion in sound quality. And further deteriorates the performance of the speech recognition system. The amount of computation in the mobile terminal is burdensome because all calculations, such as feature extraction and recognition, must be performed by the processor of the terminal in the case of the terminal-based system. The computational burden is small because the terminal-server method requires only the feature extraction process to be performed below the computational complexity of the speech encoder (about 17 WMOPS). Network protocols do not require a server-based method, and terminal-based methods are not required for systems that are currently being commercialized, but may be needed depending on the application service area. In the terminal-server method, a network protocol is essential by quantization of characteristic parameters and channel coding. The amount of recognition vocabulary can be applied to small capacity or medium capacity application by hardware limit of terminal-based method at present. Server-based and terminal- server-based methods can also be large capacity. Lastly, the performance degradation factors of each scheme are limited computational load in the terminal-based scheme, distortion of the voice signal by the transmission channel and the voice encoder in the server-based scheme, and parameter quantization and transmission channel distortion in the terminal-server scheme.

III. CONCLUSION

In this paper, we have studied the research trend of distributed speech recognition technology for efficient speech recognition service under wireless communication environment. First, we summarize the issues affecting the performance of speech recognition system in wireless communication environment and examined the efforts to solve them.

IV. ACKNOWLEDGEMENTS

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

REFERENCES

[1]. H. G. Hirsch and D. Pearce, "The AURORA experimental framework for the performance evaluations of speech recognition systems under noisy conditions", in Proc. ISCA ITRW ASR2000, Paris, France, Sept. 2000.

[2]. ETSI ES 201 108, Speech Processing, Transmission and Quality Aspects (STQ); Distributed Speech Recognition; Frond-end Feature Extraction Algorithm; Compression Algorithms, Feb. 2000.

[3]. 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., 9(5), July 2001, 558-568.

[4]. R. A. Sukkar, R. Chengalvarayan, and J. J. Jacob, "Unified speech recognition for the landline and wireless environments," in Proc. ICASSP, Orlando, FL, May 2002, 293-296.

[5]. H. K. Kim, R. V. Cox, and R. C. Rose, "Performance improvement of a bitstream-based front-end for wireless speech recognition in adverse environments," IEEE Trans. Speech Audio Process., 10(8), Nov. 2002, 591-604.

[6]. J. S. Yoon, G. H. Lee, and H. K. Kim, "A MFCC-based CELP speech coder for server-based speech recognition in networkenvironments."IEICE Trans. Electronics, Communications and Computer Sciences, E90-A(3), 2007, 626-632.

*Seung Ho Choi. "A Study on the Influence of wireless Communication Environment on Speech Recognition Performance." A Study on the Influence of wireless Communication Environment on Speech Recognition Performance, vol. 05, no. 08, 2017, pp. 94–96.