Chapter 54
Speech Enhancement Using Heterogeneous Information

Yan Xiong

Department of Computer Science, Guangdong University of Education, Guangdong, China

Fang Xu

School of Electronic and Information Engineering, South China University of Technology, Guangdong, China

Qiang Chen

Department of Computer Science, Guangdong University of Education, Guangdong, China

Jun Zhang

School of Electronic and Information Engineering, South China University of Technology, Guangdong, China

ABSTRACT

This article describes how to use heterogeneous information in speech enhancement. In most of the current speech enhancement systems, clean speeches are recovered only from the signals collected by acoustic microphones, which will be greatly affected by the acoustic noises. However, heterogeneous information from different kinds of sensors, which is usually called the "multi-stream," are seldom used in speech enhancement because the speech waveforms cannot be recovered from the signals provided by many kinds of sensors. In this article, the authors propose a new model-based multi-stream speech enhancement framework that can make use of the heterogeneous information provided by the signals from different kinds of sensors even when some of them are not directly related to the speech waveform. Then a new speech enhancement scheme using the acoustic and throat microphone recordings is also proposed based on the new speech enhancement framework. Experimental results show that the proposed scheme outperforms several single-stream speech enhancement methods in different noisy environments.

INTRODUCTION

In practical applications, speech signals are often corrupted by acoustic noises, which will degrade the quality and intelligibility of the speeches. Therefore, speech enhancement systems are required in many situations to improve the speech quality, intelligibility or the performance of speech coding and speech recognition systems (Veisi & Sameti, 2013).

Current speech enhancement techniques can be broadly divided into two categories, i.e., the single-channel and multiple-channel approaches (Low et al., 2013). The single-channel approach processes the speech signals received from a single microphone in a certain domain like the time, frequency or wavelet domain, which includes many classical speech enhancement methods such as the spectral subtraction (Boll, 1979), Winner filter (Lim & Oppenheim, 1978), MMSE (Martin, 2005), and etc. On the other hand, the multi-channel approach (Jarrett et al., 2014; Tavakoli et al., 2016) requires more than one microphone to exploit spatial information and separate the signal of interest from other interferences. The spatial filtering (or beamforming) is commonly used to form a beam towards the target signal so as to suppress the interferences from other directions. Comparing with the single-channel techniques, the multi-channel techniques can usually achieve better performance but with higher computational complexity and larger sizes. Therefore, the single microphone speech enhancement techniques are still of wide interest in many applications (Veisi & Sameti, 2013).

In most of the current speech enhancement systems, clean speeches are recovered only from the signals collected by acoustic microphones. These speech enhancement systems will be greatly affected by the acoustic noises and suffered from performance degradation in low SNR situations. The multi-stream (MS) approach (Erzin, 2012; Estellers et al., 2012; Graciarena et al., 2003; Nemala et al., 2013), which tries to use heterogeneous information from different kinds of sensors or feature extraction methods, has been successfully applied to the automatic speech recognition (ASR) for many years and proven to be an effective way to improve the recognition accuracy and robustness of the ASR systems. Based on the fact that the noises or mismatches do not affect different data streams in similar ways, the MS recognizers can usually outperform the single stream ones in various and unpredictable noisy environments by choosing and fusing complementary data streams properly. Besides the ASR, heterogeneous information is also wildly used to enhance the performances in many kinds of systems (Rychlý et al., 2015; Wu et al., 2014). However, the multi-stream approach is seldom employed in speech enhancement systems. The major difficulty of applying the multi-stream approach to speech enhancement is that the speech waveform cannot be directly recovered from many kinds of data streams, for example, the visual information of lip movements. Therefore, most of the current speech enhancement techniques only use the noisy acoustic speeches from the microphones to recover the clean speeches.

The major contributions of this paper are to propose a new multi-stream speech enhancement framework which can make use of the heterogeneous information provided by the signals from different kinds of sensors even if some of them are not directly related to the speech waveform. The proposed framework is derived from the conventional single-channel model-based speech enhancement technique, with the exception that a multi-stream classifier is employed to classify the noisy speech frames. Based on the classification results, the noisy speech is enhanced by a class dependent filter as the conventional model-based speech enhancement methods do. In this way, the heterogeneous information provided by the sensors other than the acoustic microphone does not required to be used to recover the speech waveform directly, but to improve the robustness of the speech frame classifier. Because a properly designed multi-stream classifier is usually more robust than the single-stream one, the proposed approach

Related Content

How Social Factors Influence Implicit Knowledge Construction on the Internet
Albena Antonova (2015). *Exploring Implicit Cognition: Learning, Memory, and Social Cognitive Processes (pp. 205-215).*
www.irma-international.org/chapter/how-social-factors-influence-implicit-knowledge-construction-on-the-internet/120860

Discovering News Frames: An Approach for Exploring Text, Content, and Concepts in Online News Sources
Loretta H. Cheeks, Tracy L. Stepien, Dara M. Waldand Ashraf Gaffar (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications (pp. 702-721).*
www.irma-international.org/chapter/discovering-news-frames/252053

Machine Learning Application With Avatar-Based Management Security to Reduce Cyber Threat
Vardan Mkrttchian, Leyla Gamidullaeva, Yulia Vertakovaand Svetlana Panasenko (2019). *Machine Learning and Cognitive Science Applications in Cyber Security (pp. 123-138).*
www.irma-international.org/chapter/machine-learning-application-with-avatar-based-management-security-to-reduce-cyber-threat/227579

Data Mining Approach for Predicting the Likelihood of Infertility in Nigerian Women
Peter Adebayo Idowu, Jeremiah Ademola Balogunand Olumuyiwa Bamidele Alaba (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications (pp. 1001-1027).*
www.irma-international.org/chapter/data-mining-approach-for-predicting-the-likelihood-of-infertility-in-nigerian-women/252067

Malicious Application Detection and Classification System for Android Mobiles
Sapna Malikand Kiran Khatter (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications (pp. 122-142).*
www.irma-international.org/chapter/malicious-application-detection-and-classification-system-for-android-mobiles/252023