学　位　論　文　内　容　の　要　旨

博士の専攻分野の名称　　博士（情報科学）　　氏名　孫 喜浩

学　位　論　文　題　名

A Study on Efficient Robust Speech Recognition with Stochastic Dynamic Time Warping
（確率的 DTW を用いた高効率ロバスト音声認識に関する研究）

In recent years, great progress has been made in automatic speech recognition (ASR) system.The hidden Markov model (HMM) and dynamic time warping (DTW) are the two main algorithms which have been widely applied to ASR system. Although, HMM technique achieves higher recognition accuracy in clear speech environment and noisy environment. It needs large-set of words and realizes the algorithm more complexly. Thus, more and more researchers have focused on DTW-based ASR system.

DTW is based on template matching, it can accomplish time alignment of reference and test speech features by dynamic programming. Conventional DTW is fast and less complexity, however its recognition accuracy is limited. Therefore, Conventional DTW has mostly been used for speech recognition in clear environment. Recently, a DTW with multireferences (mDTW) algorithm has also been developed to improve the recognition accuracy in comparison to HMM algorithm under noisy conditions. However the mDTW algorithm increases the calculation cost and requires more memory resources which reduce the system practicability.

It is possible to reconstruct the multireferences. The new method should be require less memory resources and reduce the calculation cost. Therefore, this study proposes a reconstruction method which add a training part to the DTW-based ASR system. The proposed reconstruction of references is aimed at making the DTW algorithm more effective. According to the DTW algorithm, the optimal warping path implies a minimum error between any two given sequences. The algorithm that we have proposed will give us a way to build a new reference to replace the original two. This process will be done in three stages; First, for each reference word, speech utterances will be divided into two subsets. Second, for each pair of subsets, the optimal path will be computed and the new reference will replace the pair of subsets. Finally, the new references will be input to the DTW-based ASR system to get the recognition accuracy.

The feasibility of the proposed technique was examined using computer simulations. The results demonstrated the effectiveness of the proposed technique. The simulation results show that our approach yields 96.94% accuracy compared with the 97.54% accuracy of mDTW in 20 dB white noise and 84.4% accuracy compared with 86.44% accuracy of mDTW in 10 dB white noise. Our approach yields 94.12% accuracy compared with 94.14% accuracy of mDTW in 20 dB babble noise and 80.82% accuracy compared with 81.64% accuracy of in 10 dB babble noise. Comparing our proposed technique, the calculation cost of mDTW has been reduced 41.6%.

This thesis presents the principle, implementation, and the results to show the feasibility of the proposed method. The main body of the thsis is organized as follows

In chapter 1, the background of automatic speech recognition (ASR) systems has been introduced. Current ASR recognizes ether the small set of words and phrases or the large vocabulary of speech sentences. For each task, a suitable ASR has been developed and improved recently. In this doctor thesis, Dynamic Time Warping (DTW) has been explored and modified suitable for an efficient robust speech recognition system.

Chapter 2 introduces the basic technologies used into ASR. The speech features are extracted by speech analysis methods and they are used for speech recognition. Normally speech features are disturbed by various noises and thus its noise components should be reduced by using noise robust technologies. After that, noise robust speech features are estimated and used for speech recognition. As commonly used speech clustering technologies, DTW and Hidden Markov Model (HMM) have been already developed. In this chapter, the overview of these technologies have been explained.

In chapter 3, the importance of automatic voice activity detection (VAD) has been discussed. In particular, under noise circumstances, it has been quite difficult to design the automatic voice activity detection with a speech recognition system. The basic concept about VAD and its current techniques have been discussed in this chapter.

Chapter 4 introduces current noise reduction technologies used into speech processing. Among them, CMS, and RSF/DRA are explained in this chapter.

Chapter 5 introduces conventional DTW methods. Some DTW methods have been developed and applied into several real applications. However, they have somewhat weak against speaker independent mechanism and various noises. Some of issues in the conventional DTW have been discussed in this chapter.

Chapter 6 has proposed new techniques using DTW, VAD, CMS and RSF/DRA. It can realizes noise robust mechanism, robust automatic VAD and high speech recognition accuracy. In addition, the proposed method can reduce the total calculation cost drastically compared with other methods whose recognition accuracy is almost the same.

Chapter 7 presents the summary of this doctor thesis.