Contents

Chinese Abstract..... i

English Abstract..... ii

Acknowledgement..... iii

Contents..... iv

Index of Figures..... vi

Index of Tables..... viii

Chapter 1 Introduction..... 1

1.1 Motivation..... 1

1.2 Overview..... 2

Chapter 2 Speech Signal Pre-Processing and Parameterization..... 3

2.1 Pre-emphasis..... 3

2.2 Frame Blocking with Hamming Window..... 4

2.3 Cepstral Analysis..... 6

2.4 Mel-Frequency Cepstrum Coefficient..... 8

Chapter 3 Speech Modeling and Recognition.....13

3.1 Definition of Hidden Markov Model..... 14

3.2 HMM in Speech Recognition..... 17

3.3 The Three Basic Problems for HMM..... 20

3.3.1 Solution to Problem 1-The Forward/Backward Algorithm..... 21

3.3.2 Solution to Problem 2-The Viterbi Algorithm..... 22

3.3.3 Problem 3-Parameter Estimation(HMM training)..... 23

3.4 Recognition Procedures..... 29

Chapter 4 Speech Recognition with Additive Noise..... 31

4.1 The Parallel Model Combination Method..... 31

4.2 Robust Speech Recognition Using the Pre-Trained Noisy Models..... 36

4.3 Experiments Result..... 39

4.3.1 Experiments A..... 39

4.3.2 Experiments B..... 48

Chapter 5 Conclusion and Future work..... 51

Reference..... 53

Index of Figures

Fig. 2.1 Procedures of producing MFCCs……...…... 3

Fig. 2.2 The diagrams of the frame, overlap and frame duration……..... 4

Fig. 2.3 Magnitude response of (a) Hamming window and (b) rectangular window.. 6

Fig. 2.4 Computation of cepstral coefficients……..... 8

Fig. 2.5 Motivation behind the real cepstrum and some of the accompanying vocabulary…….....8

Fig. 2.6 Frequency warping according to Mel scale (a) linear frequency scale (b) logarithm frequency scale…….....9

Fig. 2.7 The Mel filter banks for (a) 8 KHz (b) 16 KHz……..... 11

Fig. 3.1 An example for the observed Markov model.....15

Fig. 3.2 An example for the hidden Markov model.....16

Fig. 3.3 Scheme of an ASR system functionality..... 18

Fig. 3.4 A 3 state left-to-right HMM..... 19

Fig. 3.5 Boundary information: the red line indicates the boundary information..... 24

Fig. 3.6 (a) transcription with boundary information. (b) transcription without boundary information.....24

Fig. 3.7 Block diagram of getting an initial HMM model with boundary information..... 25

Fig. 3.8 Training procedures of the HMM..... 28

Fig. 3.9 The connection of the HMM of “one” and “two”..... 30

Fig. 4.1 The scheme of PMC..... 32

Fig. 4.2 The illustration of data-driven PMC..... 35

Fig. 4.3 The transition matrix of model “four” in (a) clean speech model (b) noisy model with 0dB (c) noisy model with 5dB (d) noisy model with 10dB..... 36

Fig. 4.4 The scheme of the method using the pre-trained noisy models to do the robust speech recognition..... 38
Fig. 4.5 The recognition results of the noisy speech using the clean speech models under different SNR corrupted by (a) babble noise (b) f16 noise (c) factory noise (d) white noise..... 41
Fig. 4.6 Normalized mean spectrums of (a) f16 (b) factory (c) babble (d) white (e) drilling machine..... 49

Index of Tables

Table 4.1 The statistics of (a) training speech data (b) testing speech data..... 40

Table 4.2 The number of state of each digit. (sil means the silence)..... 40

Table 4.3 The recognition results of the noisy speech under different SNR using the clean speech models with different mixtures of GMM. The noisy speech is corrupted by (a) f16 noise (b) babble noise (c) factory noise (d) white noise..... 42

Table 4.4 The recognition results of the noisy speech under different SNR using the clean speech models (the values in the brackets) and matched noisy models with different mixtures of GMM. The noisy speech is corrupted by (a) f16 noise (b) factory noise (c) white noise (d) babble noise..... 44

Table 4.5 The recognition results of the noisy speech under 15dB SNR using the same kind of noisy models in different SNR with 8 mixtures of GMM. The highlighted values are the results using the matched noisy models..... 45

Table 4.6 The recognition results of the noisy speech under different SNR using the clean speech models with different mixtures of GMM. The noisy speech is corrupted by (a) f16 noise (b) babble noise (c) factory noise (d) white noise..... 46

Table 4.7 The recognition results of the noisy speech under different SNR using the clean speech models (the values in the brackets) and matched noisy models with different mixtures of GMM. The noisy speech is corrupted by (a) f16 noise (b) factory noise (c) white noise (d) babble noise..... 48

Table 4.8 The recognition results of the drilling noisy speech under different SNR using the clean speech models (the values in the brackets) And babble noisy models with different mixtures of GMM..... 50

Table 5.1 The recognition results of using babble noisy model and f16 noisy model..... 52