Contents

中文摘要 i
Abstract ii
Acknowledgments iii
Contents iv
List of Figures v
List of Tables vii
Chapter 1. Introduction 1
Chapter 2. Related Work 12
2.1. HMM-based Phonetic Segmentation 12
2.2. Boundary Refinement 13
Chapter 3. Preprocessing of Corpus-based TTS/SVS 17
3.1. Corpus Design Principle 17
3.2. Phonetic Transcription 18
3.3. Pitch Estimation/Marking 20
3.3.1. Pitch Estimation 20
3.3.2. Pitch Marking 20
Chapter 4. Initial Phonetic Segmentation via HMM and DTW 30
4.1. Speech/Singing Voice Corpora 31
4.2. HMM-based Alignment with MFCCs 32
4.3. DTW-based Alignment with Pitch Contours 35
Chapter 5. Boundary Refinement Based on Hybrid Approach 41
5.1. Phonetic Transition Categories in Mandarin 41
5.2. Feature Definition 42
5.2.1. Entropy 42
5.2.2. Bisector Frequency 43
5.2.3. Acoustic Feature Vector 44
5.3. Candidate Boundaries for Training 45
5.4. Statistics-based Method 48
5.5. Performance Evaluation of Statistics-based Method 50
5.6. Heuristic Method 52
5.7. Performance Evaluation of Heuristic Method 55
Chapter 6. Boundary Refinement Based on a Score Predictive Model 57
6.1. Score Function 58
6.2. Candidate Boundaries for Training 59
6.3. Regression Model by Using Support Vector Machine 62
6.4. Boundary Refinement by Using SPM 66
6.5. Performance Evaluation of SPM 69
6.6. Performance Comparison Using Three Regression Approaches 71
6.7. Performance Comparison Using Different Boundary Refinement Methods 74
6.8. Two Attempts Regarding Performance Improvement 76
Chapter 7. Conclusions and Future Work 83
Bibliography 85
List of Publications 91

List of Figures

Fig. 1.1. The preprocessing tasks of corpus-based TTS/SVS.....2
Fig. 1.2. The flowchart of automatic phonetic segmentation on speech corpora.....8
Fig. 1.3. The flowchart of automatic phonetic segmentation on singing voices.....11
Fig. 2.1. According to [24], the representation of the super vector for a boundary.....15
Fig. 2.2. Block diagram of an MLP-based phone boundary refining system.....16
Fig. 3.1. The processing of the proposed two-phase algorithm for detecting pitch marks.22
Fig. 3.2. An example of the simplest pitch marking method.....23
Fig. 3.3. The result from peak searching based pitch marking.....25
Fig. 3.4. The result from valley searching based pitch marking. 25
Fig. 3.5. The result of pitch marking based on dynamic programming. 29
Fig. 4.1. The manual phonetic segmentation results of a Mandarin sentence, “請把這籃兔子送走” (“ging2-ba3-zhe4-lan2-tu4-z5-song4-zou3”). 30
Fig. 4.2. The local constraint of DTW alignment.....36
Fig. 4.3. The global constraint of DTW alignment. 37
Fig. 4.4. An example of DTW-based alignment.....38
Fig. 4.5. The key transposition for getting the best amount of pitch shift.....39
Fig. 5.1. The spectral entropy and the bisector frequency of the sentence, “我明年將離開彰化去日本” (“uo3-ming2-nian2-jiang-li2-kai-xang-hua4-ju4-r4-ben3”).....44
Fig. 5.2. Training data of 5 correct boundaries and 6 wrong boundaries around the true boundary labeled by humans. The content of this speech waveform was “將離” (“jiang-li2”).....46
Fig. 5.3. Training data of 9 correct boundaries and 10 wrong boundaries around the true boundary labeled by humans. The content of this singing voice waveform was “寧靜” (“ning2-jing4”).....48
Fig. 5.4. Performance comparison between HMM-based alignment and SKL boundary refinement. Top: closed test. Bottom: open test. (Evaluted data: TTS-455).....51
Fig. 5.5. Performance comparison among HMM-based alignment, DTW-based alignment and SKL boundary refinement. Top: closed test. Bottom: open test. (Evaluted data: SVS-1384) 52
Fig. 5.6. Formants comparison between two singing voice data of different pitch (the same pronunciation, “ㄚ” (“a”)).....56
Fig. 6.1. The score function of the first phonetic transition category.....59
Fig. 6.2. A typical example of the 27 candidate boundaries around a human-labeled true boundary. The content of this speech waveform was “將離” (“jiang-li2”). 60
Fig. 6.3. A typical example of the 41 candidate boundaries around a human-labeled true boundary. The content of this singing voice waveform was “寧靜” (“ning2-jing4”).....62
Fig. 6.4. The regression function of ε-SVR is represented by a tube with radius ε and slack variables ξ_i. The data points outside the ε-insensitive zone are referred to as support vectors (black dots).....64
Fig. 6.5. The construction of 54 SPMs.....66
Fig. 6.6. Boundary refinement using the proposed SPM.....68
Fig. 6.7. The performance of the proposed SPM approach. Top: closed test. Bottom: open

test. (Evaluted data: TTS-455) 70
Fig. 6.8. The performance of the proposed SPM approach. Top: closed test. Bottom: open test. (Evaluted data: SVS-1384) 71
Fig. 6.9. Performance comparison using different regression approaches. Top: closed test. Bottom: open test. (Evaluted data: TTS-455) 73
Fig. 6.10. Performance comparison using different regression approaches. Top: closed test. Bottom: open test. (Evaluted data: SVS-1384) 74
Fig. 6.11. Performance comparison between two boundary refinement approaches. Top: closed test. Bottom: open test. (Evaluted data: TTS-455) 75
Fig. 6.12. Performance comparison between two boundary refinement approaches. Top: closed test. Bottom: open test. (Evaluted data: SVS-1384) 76
Fig. 6.13. Performance comparison between two kinds of acoustic features. Top: closed test. Bottom: open test. (Evaluted data: TTS-455) 78
Fig. 6.14. Performance comparison between two kinds of acoustic features. Top: closed test. Bottom: open test. (Evaluted data: SVS-1384) 79
Fig. 6.15. Performance comparison of SPM and SPM combined with Lee's method. Top: closed test. Bottom: open test. (Evaluted data: TTS-455) 81
Fig. 6.16. Performance comparison of SPM and SPM combined with Lee's method. Top: closed test. Bottom: open test. (Evaluted data: SVS-1384) 82

List of Tables

TABLE 4.1 SEGMENTATION ACCURACY RESULTS W.R.T. DIFFERENT TRAINING MODELS ... 35
TABLE 4.2 THE SEGMENTATION RESULTS OF DTW-BASED ALIGNMENT. ... 39
TABLE 4.3 THE SEGMENTATION RESULTS OF THREE DIFFERENT MANNERS ... 40
TABLE 5.1 SIX TYPES OF INITIAL ... 42
TABLE 5.2 NINE TYPES OF FINAL ... 42
TABLE 5.3 A FORMANT-BASED DATA SET. ... 54