

# Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Speaker Recognition: Introduction</b>                       | <b>1</b>  |
| 1.1      | Basic Concepts   | 2         |
| 1.2      | Development History  | 3         |
| 1.3      | System Framework   | 8         |
| 1.4      | Categories   | 8         |
| 1.4.1    | Identification, Verification, Detection, and Tracking          | 8         |
| 1.4.2    | Text-Dependent, Text-Independent and Text-Prompted             | 9         |
| 1.5      | Performance Evaluations  | 10        |
| 1.5.1    | Evaluation Metrics for Verification or Open-Set Identification | 10        |
| 1.5.2    | Evaluation Metrics for Close-Set Identification                | 12        |
|          | References   | 13        |
| <b>2</b> | <b>Environment-Related Robustness Issues</b>                   | <b>15</b> |
| 2.1      | Background Noise   | 15        |
| 2.1.1    | Speech Enhancement   | 15        |
| 2.1.2    | Feature Compensation   | 16        |
| 2.1.3    | Robust Modeling  | 16        |
| 2.1.4    | Score Normalization  | 17        |
| 2.2      | Channel Mismatch   | 17        |
| 2.2.1    | Feature Transformation   | 18        |
| 2.2.2    | Channel Compensation   | 18        |
| 2.2.3    | Score Normalization  | 20        |
| 2.3      | Multiple Speakers  | 20        |
| 2.3.1    | Robust Features  | 21        |
| 2.3.2    | Robust Speaker Models  | 21        |
| 2.3.3    | Segmentation and Clustering Algorithms                         | 22        |
| 2.4      | Discussions  | 23        |
|          | References   | 23        |

|          |   |    |
|----------|---|----|
| <b>3</b> | <b>Speaker-Related Robustness Issues</b>      | 27 |
| 3.1      | Genders                                       | 27 |
| 3.2      | Physical Conditions                           | 28 |
| 3.3      | Speaking Styles                               | 28 |
| 3.3.1    | Emotion                                       | 28 |
| 3.3.2    | Speaking Rate                                 | 29 |
| 3.3.3    | Idiom   | 30 |
| 3.4      | Cross Languages                               | 30 |
| 3.5      | Time Varying                                  | 32 |
| 3.6      | Discussion                                    | 34 |
|          | References                                    | 34 |
| <b>4</b> | <b>Application-Oriented Robustness Issues</b> | 39 |
| 4.1      | Application Scenarios                         | 39 |
| 4.1.1    | User Authentication                           | 39 |
| 4.1.2    | Public Security and Judicature                | 39 |
| 4.1.3    | Speaker Adaptation in Speech Recognition      | 40 |
| 4.1.4    | Multi-speaker Environments                    | 40 |
| 4.1.5    | Personalization                               | 40 |
| 4.2      | Short Utterance                               | 40 |
| 4.3      | Anti-spoofing                                 | 42 |
| 4.3.1    | Impersonation                                 | 42 |
| 4.3.2    | Speech Synthesis                              | 43 |
| 4.3.3    | Voice Conversion                              | 43 |
| 4.3.4    | Replay  | 44 |
| 4.3.5    | Voice Liveness Detection                      | 45 |
| 4.4      | Cross Encoding Schemes                        | 46 |
| 4.5      | Discussion                                    | 46 |
|          | References                                    | 46 |
| <b>5</b> | <b>Conclusions and Future Work</b>            | 49 |

Robustness-Related Issues in Speaker Recognition

Zheng, Th.F.; Li, L.

2017, X, 49 p. 12 illus., 7 illus. in color., Softcover

ISBN: 978-981-10-3237-0