

Contents

1	Introduction	1
1.1	Motivation	2
1.2	Overview	4
2	Fundamentals	5
2.1	Speech Production	5
2.2	Front-End	8
2.3	Speaker Change Detection	16
2.3.1	Motivation	16
2.3.2	Bayesian Information Criterion	18
2.4	Speaker Identification	21
2.4.1	Motivation and Overview	22
2.4.2	Gaussian Mixture Models	25
2.4.3	Detection of Unknown Speakers	29
2.5	Speech Recognition	30
2.5.1	Motivation	30
2.5.2	Hidden Markov Models	31
2.5.3	Implementation of an Automated Speech Recognizer	35
2.6	Speaker Adaptation	37
2.6.1	Motivation	38
2.6.2	Applications for Speaker Adaptation in Speaker Identification and Speech Recognition	39
2.6.3	Maximum A Posteriori	41
2.6.4	Maximum Likelihood Linear Regression	44
2.6.5	Eigenvoices	45
2.6.6	Extended Maximum A Posteriori	51
2.7	Feature Vector Normalization	53
2.7.1	Motivation	53

2.7.2	Stereo-based Piecewise Linear Compensation for Environments	54
2.7.3	Eigen-Environment	56
2.8	Summary	57
3	Combining Self-Learning Speaker Identification and Speech Recognition	59
3.1	Audio Signal Segmentation	59
3.2	Multi-Stage Speaker Identification and Speech Recognition	63
3.3	Phoneme Based Speaker Identification	65
3.4	First Conclusion	69
4	Combined Speaker Adaptation	71
4.1	Motivation	71
4.2	Algorithm	72
4.3	Evaluation	76
4.3.1	Database	76
4.3.2	Evaluation Setup	77
4.3.3	Results	79
4.4	Summary	82
5	Unsupervised Speech Controlled System with Long-Term Adaptation	83
5.1	Motivation	83
5.2	Joint Speaker Identification and Speech Recognition	85
5.2.1	Speaker Specific Speech Recognition	86
5.2.2	Speaker Identification	91
5.2.3	System Architecture	93
5.3	Reference Implementation	95
5.3.1	Speaker Identification	95
5.3.2	System Architecture	96
5.4	Evaluation	97
5.4.1	Evaluation of Joint Speaker Identification and Speech Recognition	98
5.4.2	Evaluation of the Reference Implementation	105
5.5	Summary and Discussion	111
6	Evolution of an Adaptive Unsupervised Speech Controlled System	115
6.1	Motivation	115
6.2	Posterior Probability Depending on the Training Level	116
6.2.1	Statistical Modeling of the Likelihood Evolution	117
6.2.2	Posterior Probability Computation at Run-Time	122
6.3	Closed-Set Speaker Tracking	125
6.4	Open-Set Speaker Tracking	128

6.5	System Architecture	132
6.6	Evaluation.....	134
6.6.1	Closed-Set Speaker Identification	134
6.6.2	Open-Set Speaker Identification	139
6.7	Summary.....	143
7	Summary and Conclusion.....	145
8	Outlook	149
A	Appendix.....	153
A.1	Expectation Maximization Algorithm	153
A.2	Bayesian Adaptation.....	155
A.3	Evaluation Measures	158
	References.....	161
	Index.....	171