

# Contents

<b>I</b>	<b>Introduction</b>	<b>17</b>
<b>1</b>	<b>Introduction to the Research Field</b>	<b>19</b>
1.1	Scientific goals . . . . .	20
1.2	General introduction . . . . .	23
1.2.1	Situation of the domain . . . . .	23
1.2.2	Challenges . . . . .	25
1.2.3	Application areas . . . . .	26
1.3	Approaches to language recognition . . . . .	27
1.3.1	System characteristics . . . . .	27
1.3.2	Basic human language recognition . . . . .	29
1.3.3	Automatic approaches . . . . .	30
1.4	Variability compensation . . . . .	32
1.4.1	Speaker and channel variability . . . . .	32
1.4.2	Variability between databases . . . . .	34
1.4.3	Underlying model . . . . .	35
1.5	Summary . . . . .	37
1.6	Structure of this document and our related publications . . . . .	38
<b>2</b>	<b>Fundamentals of a Language Recognition System</b>	<b>39</b>
2.1	General classification system . . . . .	40
2.2	The speech signal . . . . .	43
2.3	Feature Extraction . . . . .	43
2.3.1	Cepstral coefficients . . . . .	43
2.3.2	Speech activity detection . . . . .	45
2.3.3	Feature normalization . . . . .	46
2.4	Modelization . . . . .	46
2.4.1	Gaussian Mixture Model . . . . .	47
2.4.2	Expectation Maximization . . . . .	50
2.4.3	Universal Background Model . . . . .	53
2.4.4	Maximum A Posteriori adaptation . . . . .	54
2.5	Scoring and score processing . . . . .	56
2.5.1	Scoring . . . . .	57
2.5.2	Score normalization . . . . .	58
2.6	Evaluation . . . . .	62

4.3.3	Model compensation during testing	126
4.3.4	Feature compensation during testing	128
4.4	JFA results	128
4.4.1	Comparing JFA strategies	128
4.4.2	Variability matrix rank	130
4.4.3	JFA effect on different model sizes	131
4.5	Joint Factor Analysis conclusions	132
4.6	Similar methods	133
4.6.1	Comparison between JFA and PCA	133
4.6.2	LDA	135
<b>5</b>	<b>Support Vector Machines</b>	<b>139</b>
5.1	Linear SVM	140
5.1.1	Maximum margin SVM	141
5.1.2	Soft-margin SVM	142
5.1.3	Testing with SVMs	142
5.2	Non-linear SVM	143
5.2.1	Kernel trick	143
5.2.2	Kernel types	144
5.3	SVM structure	145
5.3.1	Results	146
5.4	JFA-based SVMs	147
5.4.1	Results	147
5.5	SVM conclusion	149
<b>6</b>	<b>Variability Between Databases</b>	<b>151</b>
6.1	NIST LRE 2009 particularity	153
6.1.1	8 language task	154
6.1.2	23 language task	155
6.2	Baseline MAP results	155
6.2.1	Evaluation on 8 languages	156
6.2.2	Evaluation on 23 languages	157
6.3	Pure single category JFA systems	157
6.3.1	Evaluation on 8 languages	158
6.3.2	Evaluation on 23 languages	158
6.4	Data level fusion: Global approach with pooled data	160
6.4.1	Evaluation on 8 languages	160
6.4.2	Evaluation on 23 languages	161
6.5	Model level fusion: Merged channel compensation matrices	162
6.5.1	Evaluation on 8 languages	163
6.5.2	Evaluation on 23 languages	164
6.6	Score level fusion: Fully separate category modeling	166
6.6.1	Evaluation on 8 languages	167
6.6.2	Evaluation on 23 languages	167
6.7	Channel category conclusion	168

2.6.1	Data sources . . . . .	62
2.6.2	Performance metric . . . . .	66
2.6.3	Protocols . . . . .	70
2.7	System implementation . . . . .	71
<b>3</b>	<b>State of the Art</b>	<b>75</b>
3.1	Chronology of approaches to speech and speaker recognition . . . .	76
3.2	Early times of language recognition . . . . .	79
3.2.1	The manual or supervised era . . . . .	79
3.2.2	First automatic approaches . . . . .	80
3.3	Feature extraction . . . . .	82
3.3.1	Types of features . . . . .	82
3.3.2	Shifted Delta Cepstra . . . . .	84
3.3.3	Comparison of parameterizations . . . . .	87
3.3.4	Speech activity detection . . . . .	88
3.3.5	Feature Extraction conclusion . . . . .	90
3.4	Modelization . . . . .	91
3.4.1	Maximum Mutual Information . . . . .	91
3.4.2	GMM-SVM pushback . . . . .	93
3.4.3	Modelizations on other levels . . . . .	95
3.4.4	Conclusion on modelization . . . . .	99
3.5	Score processing . . . . .	99
3.5.1	Ratio to the Universal Background Model . . . . .	99
3.5.2	Back-end score normalizations . . . . .	100
3.5.3	Gaussian Back-End . . . . .	102
3.5.4	Back-end LDA . . . . .	104
3.5.5	Score Support Vector Machine . . . . .	106
3.5.6	Logistic Regression . . . . .	106
3.5.7	Score processing conclusion . . . . .	107
3.6	Performance metric . . . . .	108
3.6.1	Log-likelihood ratio cost . . . . .	108
<b>II</b>	<b>Novelties</b>	<b>111</b>
<b>4</b>	<b>Joint Factor Analysis</b>	<b>113</b>
4.1	The Factor Analysis model . . . . .	116
4.1.1	Complete Joint Factor Analysis . . . . .	120
4.2	Algorithm for JFA estimation . . . . .	121
4.2.1	General statistics . . . . .	122
4.2.2	Latent variables estimation . . . . .	122
4.2.3	Inter-speaker/channel matrix estimation . . . . .	123
4.2.4	Algorithm summary . . . . .	124
4.3	JFA variability compensation . . . . .	125
4.3.1	Model compensation during training . . . . .	125
4.3.2	Feature compensation during training . . . . .	126

4.3.3	Model compensation during testing . . . . .	126
4.3.4	Feature compensation during testing . . . . .	128
4.4	JFA results . . . . .	128
4.4.1	Comparing JFA strategies . . . . .	128
4.4.2	Variability matrix rank . . . . .	130
4.4.3	JFA effect on different model sizes . . . . .	131
4.5	Joint Factor Analysis conclusions . . . . .	132
4.6	Similar methods . . . . .	133
4.6.1	Comparison between JFA and PCA . . . . .	133
4.6.2	LDA . . . . .	135
<b>5</b>	<b>Support Vector Machines</b>	<b>139</b>
5.1	Linear SVM . . . . .	140
5.1.1	Maximum margin SVM . . . . .	141
5.1.2	Soft-margin SVM . . . . .	142
5.1.3	Testing with SVMs . . . . .	142
5.2	Non-linear SVM . . . . .	143
5.2.1	Kernel trick . . . . .	143
5.2.2	Kernel types . . . . .	144
5.3	SVM structure . . . . .	145
5.3.1	Results . . . . .	146
5.4	JFA-based SVMs . . . . .	147
5.4.1	Results . . . . .	147
5.5	SVM conclusion . . . . .	149
<b>6</b>	<b>Variability Between Databases</b>	<b>151</b>
6.1	NIST LRE 2009 particularity . . . . .	153
6.1.1	8 language task . . . . .	154
6.1.2	23 language task . . . . .	155
6.2	Baseline MAP results . . . . .	155
6.2.1	Evaluation on 8 languages . . . . .	156
6.2.2	Evaluation on 23 languages . . . . .	157
6.3	Pure single category JFA systems . . . . .	157
6.3.1	Evaluation on 8 languages . . . . .	158
6.3.2	Evaluation on 23 languages . . . . .	158
6.4	Data level fusion: Global approach with pooled data . . . . .	160
6.4.1	Evaluation on 8 languages . . . . .	160
6.4.2	Evaluation on 23 languages . . . . .	161
6.5	Model level fusion: Merged channel compensation matrices . . . . .	162
6.5.1	Evaluation on 8 languages . . . . .	163
6.5.2	Evaluation on 23 languages . . . . .	164
6.6	Score level fusion: Fully separate category modeling . . . . .	166
6.6.1	Evaluation on 8 languages . . . . .	167
6.6.2	Evaluation on 23 languages . . . . .	167
6.7	Channel category conclusion . . . . .	168

<b>7</b>	<b>Channel-Detectors for System Selection</b>	<b>171</b>
7.1	Designing channel category detectors . . . . .	172
7.1.1	Simple-sum . . . . .	172
7.1.2	Feature-based MAP . . . . .	173
7.1.3	SVM on channel variability . . . . .	173
7.1.4	MAP on channel variability . . . . .	173
7.1.5	Oracle . . . . .	173
7.2	Novel score normalizations . . . . .	174
7.2.1	llkMax0 . . . . .	174
7.2.2	llkInt01 . . . . .	175
7.3	Evaluation on 8 common languages . . . . .	176
7.3.1	Pure systems . . . . .	177
7.3.2	Systems with merged-Umatrix . . . . .	177
7.4	Evaluation on all 23 languages . . . . .	178
7.4.1	Pure systems . . . . .	178
7.4.2	Systems with merged-Umatrix . . . . .	179
7.5	Discussion . . . . .	179
<b>8</b>	<b>Perspectives</b>	<b>181</b>
8.1	Development set . . . . .	181
8.1.1	Chunked training files . . . . .	182
8.2	Higher level approaches . . . . .	183
8.2.1	Signal dynamics on Gaussian indices . . . . .	183
8.2.2	Pair-wise SVMs . . . . .	187
8.3	More technical directions . . . . .	188
8.4	Conclusion on perspectives . . . . .	189
<b>9</b>	<b>Conclusion</b>	<b>191</b>
9.1	Language recognition and variability . . . . .	191
9.1.1	Challenges and solutions . . . . .	191
9.1.2	Troublesome variability . . . . .	192
9.1.3	Acoustic language recognition . . . . .	193
9.2	Novelties . . . . .	193
9.2.1	Joint factor analysis . . . . .	194
9.2.2	Support Vector Machines . . . . .	196
9.2.3	Channel category variability . . . . .	196
9.2.4	Channel category detectors . . . . .	197
9.3	Wrap-up . . . . .	198
	<b>Acknowledgments</b>	<b>199</b>
	<b>List of Symbols</b>	<b>201</b>
	<b>Acronyms</b>	<b>203</b>
	<b>Glossary</b>	<b>205</b>
	<b>Bibliography</b>	<b>221</b>