

Contents

About the author	4
1 Introduction	5
1.1 Motivation, scope and structure	5
1.2 Acknowledgments	6
2 BUT on SR research and NIST SRE time-line	7
2.1 Prehistory: NIST-NSA SRE 1998	7
2.2 Beginning of 2000s – Historical context	7
2.3 End of relevance MAP adaptation era	8
2.4 JFA era	8
2.5 i-vector era	9
2.6 Neural network Era	11
2.6.1 Hybrid i-vector/Neural network era	11
2.6.2 Era of neural network embedding and end-to-end systems	12
3 Speaker modeling techniques	13
3.1 Gaussian mixture models	13
3.2 Relevance-MAP	13
3.3 GMM eigen-channel compensation	14
3.4 Joint factor analysis	14
3.5 i-vectors	14
3.6 Neural network embedding: x-vector topology	15
3.7 Backend—PLDA	16
3.8 Score normalization	16
3.9 Data augmentation	16
4 Data and evaluation metric	16
4.1 Training data	16
4.1.1 Switchboard	17
4.1.2 NIST speaker recognition evaluations	17
4.1.3 VoxCeleb	17
4.2 Evaluation data	18
4.3 Evaluation metrics	18
5 Longitudinal analysis	19
5.1 Redoing NIST SREs with all major techniques	19
5.2 Impact of training data size on performance	21
5.3 Increasing difficulty of NIST SREs	23
5.4 Looking forward—wideband data	24
6 Conclusions	25
References	26
Abstrakt	33