

Contents

1	Introduction	1
1.1	Problem Statement	1
1.2	Scope of This Book and Contribution	3
1.3	Outline of This Book	3
2	Literature Review	5
2.1	Today's Communication Channels and Their Main Impairments	5
2.2	Channel Quality Evaluation	9
2.2.1	Subjective Speech Quality Assessment	9
2.2.2	Instrumental Speech Quality Measures	11
2.2.3	Relations Between Quality and Other Attributes of the Speech Signal	12
2.3	Human Speaker Recognition	13
2.3.1	Speech Characteristics Enabling Human Speaker Recognition	14
2.3.2	Effects of Communication Channels on Human Speaker Recognition	17
2.3.3	Literature on Human Speaker Recognition and This Book	18
2.4	Automatic Speaker Recognition	18
2.4.1	Automatic Speaker Recognition Principles and Main Systems	20
2.4.2	Effects of Phonetic Content on Automatic Speaker Recognition	25
2.4.3	Effects of Communication Channels on Automatic Speaker Recognition	28
2.4.4	NIST Speaker Recognition Evaluations	31

2.4.5	Comparison Between the Human and the Automatic Speaker Recognition Performance.	32
2.4.6	Literature on Automatic Speaker Recognition and This Book.	33
3	Human Speaker Identification Performance Under Channel Degradations	35
3.1	Experimental Environment of the Listening Tests	35
3.1.1	Database Collection for the Listening Tests	35
3.1.2	Listening Test 1	36
3.1.3	Listening Test 2	38
3.2	Effects of Codec and Bandwidth	41
3.3	Effects of Electro-Acoustic User Interface	46
3.4	Effects of Random Packet Loss	47
3.5	Target Speaker and Familiarity	48
4	Importance of Intelligible Phonemes for Human Speaker Recognition in Different Bandwidths	51
4.1	Human Speaker Recognition from Logatomes	51
4.1.1	Audio Preparation and Listening Test	52
4.1.2	Accuracies per Logatome in Different Bandwidths	53
4.2	Human Speech Intelligibility from Logatomes	57
4.3	Relation Between Speaker Recognition and Intelligibility in Narrowband and in Wideband	62
5	Automatic Speaker Verification Performance Under Channel Distortions	63
5.1	Datasets and Speech Transmissions.	64
5.2	Effects of Channel Impairments in Matched Conditions.	65
5.2.1	GMM-UBM Performance Under Channel Distortions	67
5.2.2	JFA Performance Under Bandwidth and Codec Effects	72
5.2.3	I-Vectors Performance Under Bandwidth and Codec Effects	74
5.3	Training and Testing Approaches to Reduce Possible Mismatch Effects	77
5.3.1	JFA Under Channel Mismatch	77
5.3.2	I-Vectors Under Channel Mismatch	80
6	Detecting Speaker-Discriminative Spectral Content in Wideband for Automatic Speaker Recognition.	85
6.1	Effects of the Transmission Channel on the Distribution of Speaker-Discriminative Spectral Content	86
6.1.1	Audio Material.	86

6.1.2	Spectral Sub-band Analysis	87
6.1.3	Results and Discussion	88
6.1.4	Sub-band Score-Level Fusion.	94
6.2	Different Cepstral Features for Narrowband and for Wideband Speech	95
6.2.1	Speech Filtering and Feature Extraction.	97
6.2.2	I-Vector Experiments	98
6.2.3	Results and Discussion	99
6.2.4	Score-Level Fusion of Two Frequency Ranges.	105
6.3	Relevance of Phonetic Information Under Transmission Channel Effects	106
6.3.1	Phoneme Filtering	107
6.3.2	I-Vector Experiments	108
6.3.3	Results and Discussion	109
7	Relations Among Speech Quality, Human Speaker Identification, and Automatic Speaker Verification	113
7.1	Quality and Performance Metrics for Different Channel Degradations	114
7.1.1	Instrumental Quality Measurements	114
7.1.2	Quality and Speech and Speaker Recognition Performance Metrics.	116
7.2	Predicting Human Speaker Identification Performance from Measured Speech Quality.	119
7.2.1	Model Fit with POLQA MOS as Estimator	119
7.2.2	Model Fit with DIAL Coloration as Estimator	123
7.2.3	Estimations of Human Speaker Identification Performance	129
7.3	Predicting Automatic Speaker Verification Performance from Measured Speech Quality.	131
7.3.1	New Instrumental Quality Measurements.	131
7.3.2	Model Fit with POLQA MOS as Estimator	133
7.3.3	Model Fit with DIAL Coloration as Estimator	133
7.3.4	Estimations of Automatic Speaker Verification Performance	135
7.4	Predicting Human Speaker Identification Performance from Computed Speaker Verification EERs	138
7.4.1	Model Fit with EERs as Estimators	138
7.4.2	Estimations of Human Speaker Identification Performance	139

8	Conclusions and Future Work	145
8.1	Summary of Research Findings and Contributions	145
8.2	Directions for Further Research	148
	Appendix A: Channel Transmissions	151
	References.	155



<http://www.springer.com/978-981-287-726-0>

Human and Automatic Speaker Recognition over
Telecommunication Channels

Fernández Gallardo, L.

2016, XII, 169 p., Hardcover

ISBN: 978-981-287-726-0