

Contents

- 1 Introduction** 1
 - 1.1 Motivation 1
 - 1.2 Outline 3
 - References. 3
- 2 Speech Synthesis** 5
 - 2.1 Setup of a Speech Synthesizer 5
 - 2.1.1 Natural Language Processing (NLP) 6
 - 2.1.2 Prosody Generation. 7
 - 2.1.3 Concatenation and Generation of Speech-Signal Parameters. 7
 - 2.1.4 Speech Signal Generation 7
 - 2.2 The Mary Text-to-Speech System (MaryTTS) 15
 - References. 17
- 3 Auditory and Instrumental Quality Evaluation Metrics**. 19
 - 3.1 What Is Perceptual Quality? 19
 - 3.2 Taxonomy for the Quality Assessment of Synthetic Speech. 21
 - 3.2.1 Glass Box Versus Black Box 22
 - 3.2.2 Laboratory Versus Field Studies 22
 - 3.2.3 Linguistic Versus Acoustic 23
 - 3.2.4 Auditory Versus Instrumental 23
 - 3.3 Auditory Quality Evaluation Metrics 23
 - 3.3.1 Functional Tests 24
 - 3.3.2 Judgment Tests 26
 - 3.4 Instrumental Quality Evaluation Metrics. 29
 - 3.4.1 Reference-Based Measures 30
 - 3.4.2 Reference-Free Measures 32
 - References. 34

4	Perceptual Quality Dimensions	37
4.1	State-of-the-Art Perceptual Quality Dimensions	38
4.1.1	Study: Kraft and Portele	38
4.1.2	Study: Mayo et al. I	38
4.1.3	Study: Viswanathan and Viswanathan	39
4.1.4	Study: Seget	39
4.1.5	Study: Hinterleitner	39
4.1.6	Study: Mayo et al. II	40
4.1.7	Restrictions of Discussed Studies	40
4.2	Semantic Differential and Factor Analysis	43
4.2.1	Experimental Setup	43
4.2.2	Statistical Analysis	46
4.3	Sorting Task and Multidimensional Scaling	51
4.3.1	Experimental Setup	52
4.3.2	Statistical Analysis	54
4.4	Summary of the SD/FA and ST/MDS Studies	59
4.5	Universal Perceptual Quality Dimensions	61
4.5.1	Naturalness of Voice	63
4.5.2	Prosodic Quality	63
4.5.3	Fluency and Intelligibility	63
4.5.4	Absence of Disturbances	63
4.5.5	Calmness	64
4.5.6	Instructions for TTS Quality Assessment	64
4.6	Summary	66
	References	66
5	Influencing Factors on Perceptual Quality	69
5.1	Influence of the Application	70
5.1.1	Pretest	70
5.1.2	Main Test	78
5.1.3	Conclusions	81
5.2	Influence of a Speakers Voice	82
5.2.1	Experimental Setup	82
5.2.2	Statistical Analysis	84
5.2.3	Conclusions	89
5.3	Influence of Corpus Size and Utterance	90
5.3.1	Experimental Setup	90
5.3.2	Statistical Analysis	93
5.3.3	Conclusions	98
5.4	Summary	98
	References	99

6 Instrumental Quality Assessment	101
6.1 Reference-Based Measures	102
6.1.1 State of the Art	102
6.1.2 Quality Prediction	102
6.1.3 Conclusions	106
6.2 Reference-Free Measures	106
6.2.1 State of the Art	107
6.2.2 Linear Regression Models	111
6.2.3 Conclusions	121
6.3 Summary	122
References	122
7 Requirements for the Integration of an Instrumental Quality Measure into a Concatenative TTS System	125
7.1 Regular Perception Model (RPM)	126
7.1.1 Model Training	126
7.1.2 Results	126
7.2 Unit Selection Voice Creation in MaryTTS	128
7.2.1 Test Database	128
7.2.2 Generation of Alternative Versions	129
7.3 Experimental Setup	131
7.3.1 Pairwise Comparison (PC)	131
7.3.2 Semantic Differential (SD)	131
7.3.3 Test Procedure	132
7.4 Statistical Analysis	132
7.4.1 PC Data	132
7.4.2 SD Data	133
7.4.3 Discussion	133
7.5 Quality Prediction	133
7.5.1 RPM1 Performance	134
7.5.2 RPM2 Training	134
7.5.3 RPM2 Performance	134
7.6 Automatic Selection of Alternative Versions	136
7.7 Potential Improvements	137
7.8 Summary	137
References	138
8 Conclusions and Future Work	139
8.1 Summary	139
8.2 Conclusions	140
8.3 Future Work	144
8.3.1 Perceptual Quality Dimensions	144
8.3.2 Influencing Factors	144
8.3.3 Instrumental Quality Measurement	145
Appendix: Statistical Analysis of Chap. 7	147

Quality of Synthetic Speech
Perceptual Dimensions, Influencing Factors, and
Instrumental Assessment

Hinterleitner, F.

2017, XVI, 157 p. 29 illus., Hardcover

ISBN: 978-981-10-3733-7