## 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
6 Instrumental Quality Assessment

6.1 Reference-Based Measures
   6.1.1 State of the Art
   6.1.2 Quality Prediction
   6.1.3 Conclusions

6.2 Reference-Free Measures
   6.2.1 State of the Art
   6.2.2 Linear Regression Models
   6.2.3 Conclusions

6.3 Summary

References

7 Requirements for the Integration of an Instrumental Quality Measure into a Concatenative TTS System

7.1 Regular Perception Model (RPM)
   7.1.1 Model Training
   7.1.2 Results

7.2 Unit Selection Voice Creation in MaryTTS
   7.2.1 Test Database
   7.2.2 Generation of Alternative Versions

7.3 Experimental Setup
   7.3.1 Pairwise Comparison (PC)
   7.3.2 Semantic Differential (SD)
   7.3.3 Test Procedure

7.4 Statistical Analysis
   7.4.1 PC Data
   7.4.2 SD Data
   7.4.3 Discussion

7.5 Quality Prediction
   7.5.1 RPM1 Performance
   7.5.2 RPM2 Training
   7.5.3 RPM2 Performance

7.6 Automatic Selection of Alternative Versions

7.7 Potential Improvements

7.8 Summary

References

8 Conclusions and Future Work

8.1 Summary

8.2 Conclusions

8.3 Future Work
   8.3.1 Perceptual Quality Dimensions
   8.3.2 Influencing Factors
   8.3.3 Instrumental Quality Measurement

Appendix: Statistical Analysis of Chap. 7
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