Contents

1 Introduction **1**

2 High-speed Videoendoscopy **11**
2.1 Laryngoscopic Imaging 11
2.2 Practical Challenges 12

3 Image Processing **15**
3.1 Deep Learning 15
3.2 Low-Light Image Enhancement of High-Speed Video 19
3.2.1 Low-Light Enhancement 19
3.2.2 Metrics 21
3.2.3 Compared Enhancement Methods 23
3.2.4 Neural Network Architecture 25
3.2.5 Training Data 28
3.3 Results 32
3.3.1 Setup 32
3.3.2 Validation Data 33
3.3.3 Test Data 34
3.4 Discussion 35

4 Automatic Segmentation **39**
4.1 Segmentation of Endoscopic High-speed Video 39
4.2 Semantic Segmentation using Deep Learning 42

5 Vocal Fold Models and Voice Parameter Estimation **51**
5.1 Models of the Vocal Folds 51
5.1.1 The Improved Two-Mass-Model 53
5.2 Estimation of Physical Voice Parameters 58
5.2.1 Degrees of Freedom 60
5.2.2 Optimization 63

5.3 Estimation of Subglottal Pressure in ex vivo High-Speed Videos 66
5.3.1 Setup 68
5.3.2 Optimization Results 70
5.3.3 Subglottal Pressure Results 71
5.3.4 Discussion 73
5.4 Shortcomings and Limitations 77

6 Voice Parameter Estimation with a Recurrent Neural Network 81
6.1 Deep Learning in Inverse Problems 81
6.2 Estimating Subglottal Pressure with an LSTM 83
6.3 Results 89
6.3.1 Setup 89
6.3.2 Experimental Results 90
6.4 Discussion 93
6.4.1 Subglottal Pressure Estimation 93
6.4.2 Runtime 94
6.4.3 Parameter Estimation and Inverse Problems 95
6.4.4 Benefit and Applicability of the Approach 96

7 Conclusion 99
7.1 Project Status 99
7.2 Project Outlook 101
7.2.1 Technical Innovation 102
7.2.2 Experimental Validation and Extension to Different Parameters 102
7.2.3 Clinical Use Cases 103
7.3 Project Impact 103

Bibliography 105

List of Figures 132

List of Tables 137